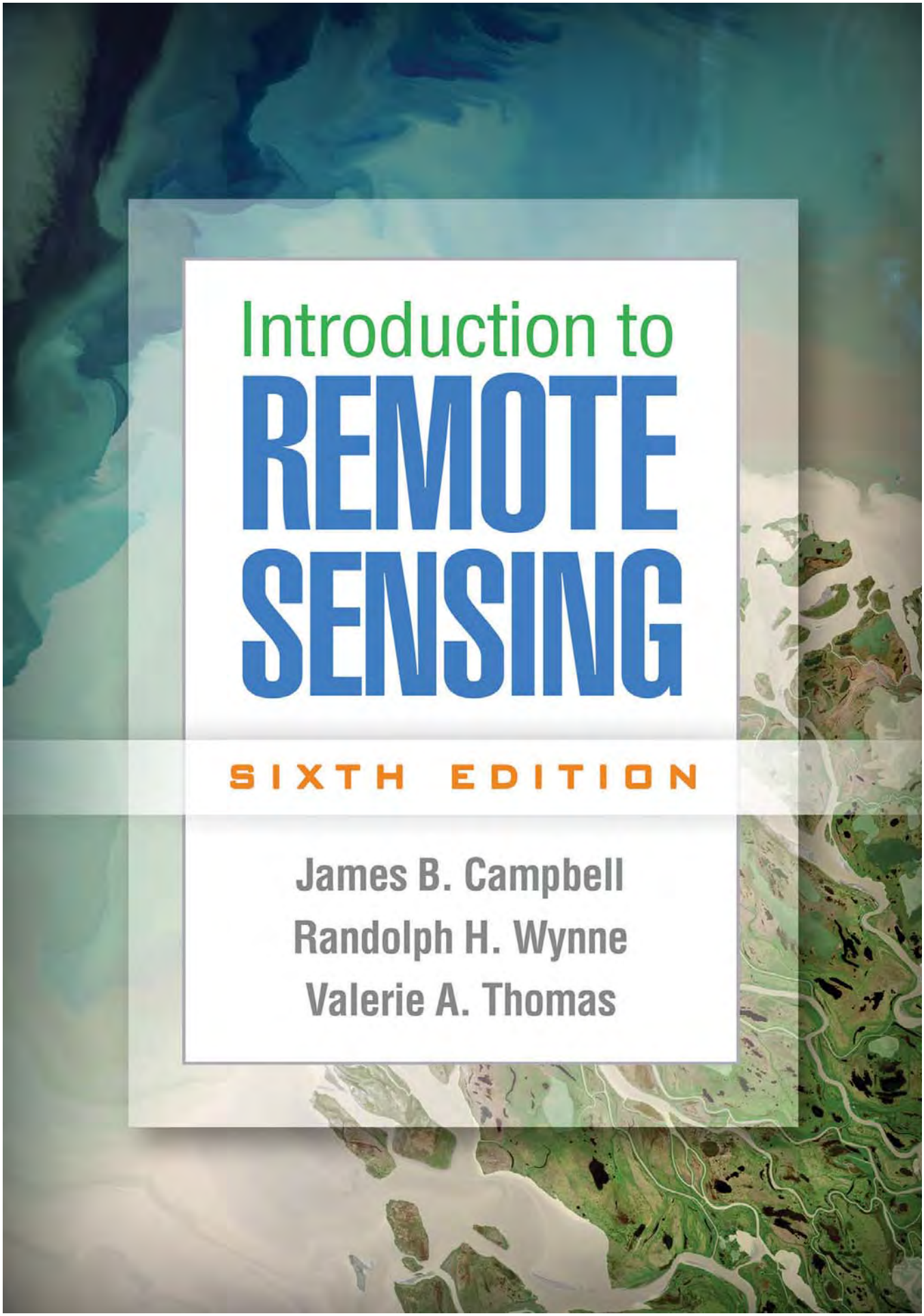

Introduction to
REMOTE
SENSING
SIXTH EDITION
James B. Campbell
Randolph H. Wynne
Valerie A. Thomas

GUILFORD PRESS
e-book

INTRODUCTION TO REMOTE SENSING

Introduction to REMOTE SENSING

SIXTH EDITION

James B. Campbell
Randolph H. Wynne
Valerie A. Thomas

THE GUILFORD PRESS
New York London

A Division of Guilford Publications, Inc.
370 Seventh Avenue, Suite 1200, New York, NY 10001
www.guilford.com

Printed in the United States of America

This book is printed on acid-free paper.

Last digit is print number: 9 8 7 6 5 4 3 2 1

Library of Congress Cataloging-in-Publication Data is available from the publisher.

ISBN 978-1-4625-4940-5 (hardcover)

Contents

Preface

This preface presents the sixth edition of our remote sensing text. During the interval that has elapsed since our fifth edition, the pace of remote sensing has increased in several respects, expanding the scope of our text and presenting new content. The advanced capabilities of satellite systems, thermal imagery, and drones have made essential contributions to our efforts to address the significant challenges of our time, including wildfires, floods, and coastal erosion. Few could have anticipated the development of analytical tools and techniques that are now available for analysis of remotely sensed data, the explosion of new sensor systems, or the multiplicity of remote sensing's applications throughout society. Such developments alone present challenges for any text on this subject.

Our sixth edition benefits from the addition of a new author, Dr. Valerie Thomas, an experienced faculty member at Virginia Tech, who now joins Dr. Randolph Wynne and Dr. James Campbell. Dr. Thomas brings new knowledge and perspectives to the text that will benefit users at all levels.

Changes in our field have also been reflected in the new four-color design throughout, with hundreds of new photos and figures, including original drawings by Susmita Sen. In addition, there are three new chapters on remote sensing platforms, agriculture, and forestry. Technological advances and cutting-edge applications are presented throughout this volume, including (1) discussions of Landsat 8 and Sentinel-2, (2) the growth of

unmanned aerial systems, (3) mobile data collection, (4) current directions in climate change detection, (5) fire monitoring, (6) disaster response, and many other timely topics.

With the use of varied local and global examples and case studies, this sixth edition will provide readers with an understanding of the latest tools and principles of collecting remote images, analyzing and interpreting the images, and applying them to land and water use. It shows how remote sensing data are used in multiple fields, including plant sciences, agriculture, forestry, earth sciences, hydrology, and land-use analysis.

WHO THIS BOOK IS FOR

We wrote this text as a two-in-one book that will provide students with an accessible introduction to remote sensing, which can also serve as a foundational reference book. For students who intend to specialize in remote sensing, this text forms not only an introduction but also a framework for subjects to be studied in greater detail. Students who do plan specialization in remote sensing should consult their instructors to plan a comprehensive course of study based on work in several disciplines, as discussed in Chapter 1. This approach is presented in the text itself, introducing students to the principal topics of significance for remote sensing, but acknowledging that students will require additional depth in their chosen fields of specialization.

For those students who do not intend to pursue remote sensing beyond the introductory level, our text serves as an overview and introduction, so that they can understand remote sensing, its applications in varied disciplines, and its significance in today's world. For many, the primary emphasis will likely be on study of the chapters and methods of greatest significance in the student's major field of study.

SPECIAL FEATURES TO AID READERS

The chapters now open with a list of the chapter's major topics, and, as mentioned earlier, new case examples, such as Washington State's Oso River debris flow, illustrate each chapter's concepts. Chapters conclude with end-of-chapter review questions on the chapter's content. And, for many chapters, we have added a short list of teaching and learning resources—principally a selection of online tutorials or short videos, such as those found on YouTube; these videos provide depth or breadth to the content presented in the chapter or simply illustrate content. They have been selected for their brevity (most are less than 3–4 minutes or so in length) and for their effectiveness in explaining or illustrating content relative to the chapter in question. For the most part, we have excluded videos that focus on promotional content. Those videos that do serve a promotional purpose have been selected for their effectiveness in presenting technical content rather than as an endorsement of a particular product or service.

ORGANIZATION

We have retained the popular short-chapter format used in previous editions that can be taught in any order to meet the specific needs of each instructor. Numbered sections within chapters form smaller units that instructors can select and combine with other

content as preferred. Our content provides organization at several levels to encourage instructors to select specific structures for their courses. At the broadest level, the rough division into four units offers a progression in the knowledge presented, with occasional concessions to practicality (such as placing the "Image Interpretation" chapter in Part II under "Image Acquisition" rather than in its logical position in Part III, "Analysis"). Here, we present each division as consisting of three or more chapters organized as follows:

NEW TO THIS EDITION

Now in full color with over 400 figures, hundreds of which are new to this edition, this sixth edition includes the latest technological advances and cutting-edge applications throughout. Three new chapters have been added, and they cover remote sensing platforms, agriculture, and forestry. Additional updates by chapter include:

Chapter 1. Introducing Remote Sensing Basics: A new section introduces both workers who contributed to the basics of aviation and those who have advanced

worldwide achievements. This chapter also includes updated coverage of Open Landsat data policy, unmanned aerial vehicles (UAVs), and thermal infrared sensors (TIRS) thermal imaging.

Chapter 3. Remote Sensing Platforms: This new chapter introduces readers to fundamental aspects of remote sensing, including aerial cameras, Landsat imagery, unmanned aerial systems (UASs), tethered balloons used in the monitoring of wildlife, satellite systems (including geofencing), and mobile collection of field data (such as cell phones).

Chapter 7. Land Observation Satellites: This chapter describes the most recent Landsat system, Sentinel-2A and 2B, from the European Space Agency, and SPOT 7 (a French satellite), which provides information for land management, disaster response, and security programs.

Chapter 9. Lidar: Lidar (*light detection and ranging*) systems record the intensity and timing of returns from a pulsed laser. Airborne laser scanning systems often use near-infrared lasers to map the elevation of land surfaces in fine detail. New coverage includes lidar pulse densities, the normalized point cloud, lidar profiles, point clouds derived using digital aerial photogrammetry, heights extracted from waveforms, and lidar data collected from the International Space Station (ISS).

Chapter 10. Thermal Imagery: This chapter provides new examples of aerial images depicting thermal features, such as the heating of urban landscapes or forest fires (such as those within the Shasta-Trinity Forest); thermal images of residential structures; urban heat islands; and daytime/nighttime satellite imagery.

Chapter 11. Statistics and Preprocessing: In the context of digital analysis of remotely sensed data, preprocessing refers to those operations that are preliminary to the principal analysis. Thus, preprocessing forms a preparatory phase that, in principle, improves image quality as the basis for later analyses that will extract information from the image. New content includes expanded coverage of principal components analysis (PCA), image statistics, and conversion to top- or bottom-of-atmosphere reflectance.

Chapter 12. Image Classification: This chapter has been almost completely rewritten to include modern machine learning approaches that have largely supplanted Bayesian maximum likelihood. These include *k*-nearest neighbor, classification trees, and random forests.

Chapter 13. Accuracy Assessment: This chapter now reflects current best practices, including guiding the user step-by-step through probabilistic sampling design, computing and reporting area proportions in the area matrix, and using the area proportions in summary statistics like the overall accuracy, user's accuracy, and producer's accuracy.

Chapter 14. Hyperspectral Remote Sensing: This chapter now includes common processing protocols like the spectral hourglass and their component steps (e.g., locating and identifying endmembers and spectral unmixing).

Chapter 15. Change Detection: This chapter has been updated to include a typology of multitemporal change detection techniques and details of common algorithms such as exponentially weighted moving average change detection and continuous change detection and classification. Change attribution is also explicitly addressed.

Chapter 16. Plant Science Fundamentals: Expanded discussion of leaf or canopy water content has been added, along with indices appropriate for their estimation. The land surface phenology section has been updated and expanded. Methods for the remote estimation of chlorophyll and solar-induced fluorescence have been detailed.

Chapter 17. Agricultural Remote Sensing: This new chapter shows how remote sensing's aerial perspective provides significant insights into agricultural landscapes. It includes coverage of analytical strategies for agricultural analysis of satellite imagery, irrigated agriculture, crop calendars, storm-damaged crops, tillage status, agricultural management, and the USDA cropland data layer.

Chapter 18. Forestry: Forestry's wide-ranging management objectives—including timber, forest products, water quality, carbon, sequestration, biodiversity, and wildlife conservation—are explored in this new chapter. Other topics include the assessment of competing vegetation, species identification, forest photogrammetry, airborne laser scanning, fire fuel loading, and forest measurements and monitoring.

Chapter 19. Earth Sciences: This chapter examines terrain, physiography, and geomorphic systems in an Earth systems context, with new coverage of major landslides (Oso, Washington, and Grand Mesa, Colorado); examples of stream diversion (northwestern Virginia); soil mapping; and soil scientists.

Chapter 20. Coastal Processes and Landforms: New topics include multispectral bathymetry, wave generation, swash and backwash, beach profiles (Virginia's barrier islands), coastal classification (Texas Point), storm damage (Mantoloking, New Jersey), and renewal (Miami Beach).

ACKNOWLEDGMENTS

We gratefully acknowledge the contributions of those who assisted in the work on this book, especially Dr. Susmita Sen, who devoted special attention to create 81 new valuable graphics for our sixth edition. We also acknowledge the support of Kathryn Hollandsworth (Virginia Tech), who reviewed, corrected, and edited text that benefited from her careful attention. We note also the contributions of Michelle Klopfer and Stacy Kuhar, both of whom supported our project. We also recognize the assistance of several colleagues, who provided graphics, insight, and information that were sometimes acquired at remote locations. Among these colleagues were Dr. Baojuan Zheng, Dr. Tammy Parece, Dr. Hoa Tran, Dr. Iris Fynn, and Dr. Jie Ren, who all provided useful and insightful content that contributed significantly to our project over long intervals. Other students who contributed include Robert Severynse, Austin Hays, Michael Graham, Jessica Dorr, and Eric Guenther.

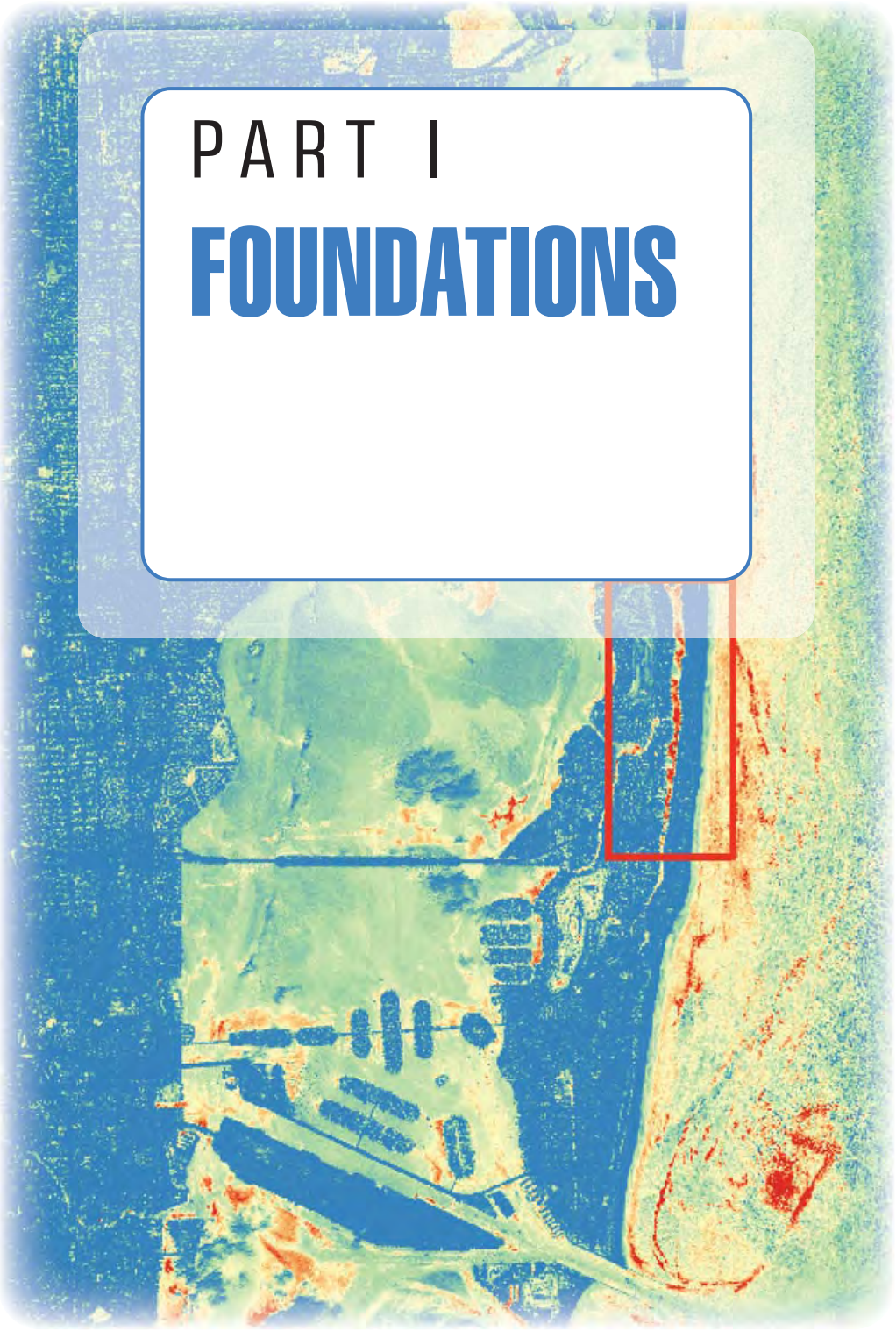
We would also like to thank past instructors and students who made suggestions for earlier editions of this book and for the sixth edition, including eight initially anonymous Guilford reviewers whose identities have recently been revealed to us: Nate Currit, Texas State University; Mary Henry, Miami University, Oxford; Anthony Filippi, Texas A & M University; Robert A. Washington-Allen, University of Tennessee, Knoxville; Jeffrey Chambers, University of California–Berkeley; Greg Gaston, University of North Alabama; Andrew Klein, Texas A & M University; and Eric S. Kasischke, University of Maryland.

And at The Guilford Press, Seymour Weingarten, Editor-in-Chief, has continued his support of our project over our six editions, with the support of Guilford and its many members. C. Deborah Laughton, Publisher in Methodology and Senior Editor in Geography, has been an essential partner in producing this project and supporting us through challenging conditions. We specifically recognize the role of Katherine Sommer for her strong editorial support, organizing the final manuscript, tracking down figures and their permissions, and finding alternatives when needed. Robert Sebastiano, also at Guilford, was available to assist with a variety of permissions questions and tasks.

List of Tables

List of Figures

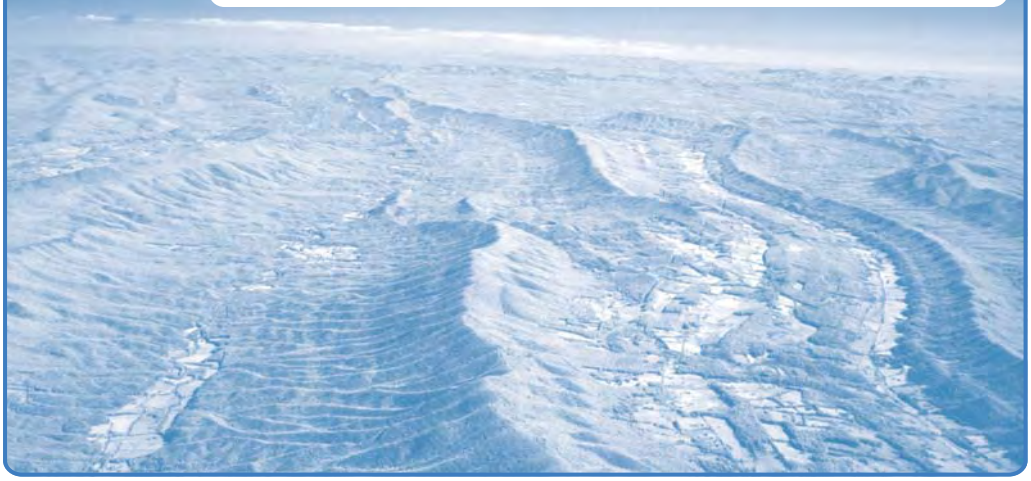
81 original color and black-and-white figures created by Susmita Sen.



PART I

FOUNDATIONS

1 Introducing Remote Sensing Basics



MAJOR TOPICS TO UNDERSTAND

- Definitions
- Milestones in the History of Remote Sensing

1.1 INTRODUCTION

The field of remote sensing is devoted to analysis of a special class of images that employ an overhead perspective (e.g., maps, aerial photographs, and similar images), including many that are based on radiation not visible to the human eye. These images have special properties that offer unique advantages for the study of the Earth's surface: we can see patterns instead of isolated points and see relationships between features that otherwise seem independent. They are especially powerful because they permit us to monitor changes over time; to measure sizes, areas, depths, and heights; and, in general, to acquire information that is very difficult to acquire by other means. However, our ability to extract this kind of information is not innate; we must work hard to develop the knowledge and skills that allow us to use images.

Foundations of remote sensing rely on the partnership of aerial photography with the airplane. Photography began in the early 1800s; the airplane was used for aerial photography in 1908. Aerial photography began at an early date (using balloons and kites), requiring many years of experimentation before the camera became a practical means for aerial photography. Use of the aircraft for photography requires a feasible match between the elements of remote sensing, including some of its many practical applications. Here in Chapter 1 we consider a few topics to outline remote sensing's content, origins, and scope as a foundation for the more specialized chapters that follow.

1.2 DEFINITIONS

The field of remote sensing has been defined many times (Table 1.1). Examination of common elements in these varied definitions permits identification of the topics' most important themes. From a cursory look at these definitions, it is easy to identify a central concept: the gathering of information at a distance. This excessively broad definition, however, must be refined if it is to guide us in studying a body of knowledge that can be approached in a single course of study.

The practice of remote sensing is devoted to observation of the Earth's land and water surfaces by means of reflected or emitted electromagnetic energy. This more focused definition excludes applications that could be reasonably included in broader definitions, such as sensing the Earth's magnetic field or atmosphere or the temperature of the human body. For our purposes, the definition can be a modification of concepts given in Table 1.1:

> Remote sensing is the practice of deriving information about the Earth's land and water surfaces using images acquired from an overhead perspective, using electromagnetic radiation in one or more regions of the electromagnetic spectrum, reflected or emitted from the Earth's surface.

This definition serves as a concise expression of the scope of this volume. It is not, however, universally applicable, and it is not intended to be so because practical constraints limit the scope of this volume. So, although this book must omit many interesting topics (e.g., meteorological or extraterrestrial remote sensing), it can review knowledge and perspectives necessary for pursuit of topics that cannot be covered in full here.

1.3 MILESTONES IN THE HISTORY OF REMOTE SENSING

We begin by outlining some of the noteworthy events and developments that introduce the field of remote sensing. By necessity, we abbreviate our account to present concise highlights of some key developments. More complete accounts are given by Fischer (1975), Simonett (1983), and others. For some of our initial discussion, our thread follows the trajectory mainly as recorded by U.S., British, and French records. Readers should, however, be aware that aviation, aerial photography, and photointerpretation all have rich histories in Germany, Switzerland, Italy, Canada, Australia, Russia, and many other nations. This chapter traces the evolution of this field to outline significant trends and key innovations.

TABLE 1.1 Remote Sensing: Some Definitions

Remote sensing has been variously defined but basically it is the art or science of telling something about an object without touching it. (Fischer, Hemphill, and Kover, 1976, p. 34)
Remote sensing is the acquisition of physical data of an object without touch or contact. (Lintz and Simonett, 1976, p. 1)
Imagery is acquired with a sensor other than (or in addition to) a conventional camera through which a scene is recorded, such as by electronic scanning, using radiations outside the normal visual range of the film and camera—microwave, radar, thermal, infrared, ultraviolet, as well as multispectral, special techniques are applied to process and interpret remote sensing imagery for the purpose of producing conventional maps, thematic maps, resources surveys, etc., in the fields of agriculture, archaeology, forestry, geography, geology, and others. (American Society of Photogrammetry)
Remote sensing is the observation of a target by a device separated from it by some distance. (Barrett and Curtis, 1976, p. 3)
The term *remote sensing* in its broadest sense merely means "reconnaissance at a distance." (Colwell, 1966, p. 71)
Remote sensing, though not precisely defined, includes all methods of obtaining pictures or other forms of electromagnetic records of the Earth's surface from a distance, and the treatment and processing of the picture data. . . . Remote sensing then in the widest sense is concerned with detecting and recording electromagnetic radiation from the target areas in the field of view of the sensor instrument. This radiation may have originated directly from separate components of the target area; it may be solar energy reflected from them; or it may be reflections of energy transmitted to the target area from the sensor itself. (White, 1977, pp. 1–2)
"Remote sensing" is the term currently used by a number of scientists for the study of remote objects (earth, lunar, and planetary surfaces and atmospheres, stellar and galactic phenomena, etc.) from great distances. Broadly defined . . . , remote sensing denotes the joint effects of employing modern sensors, data-processing equipment, information theory and processing methodology, communications theory and devices, space and airborne vehicles, and large-systems theory and practice for the purposes of carrying out aerial or space surveys of the earth's surface. (National Academy of Sciences, 1970, p. 1)
Remote sensing is the science of deriving information about an object from measurements made at a distance from the object, i.e., without actually coming in contact with it. The quantity most frequently measured in present-day remote sensing systems is the electromagnetic energy emanating from objects of interest, and although there are other possibilities (e.g., seismic waves, sonic waves, and gravitational force), our attention . . . is focused upon systems which measure electromagnetic energy. (D. A. Landgrebe, quoted in Swain and Davis, 1978, p. 1)

Early Photography (1839–1909)

Because the practice of remote sensing focuses on the examination of images of the Earth's surface, remote origins lie in the beginnings of the practice of photography. Early attempts to form photographic images date from the early 1800s, when a number of scientists, now largely forgotten, conducted experiments with photosensitive chemicals.

In 1839, Louis Daguerre (1789–1851), well recognized as one of the early photographers, publicly reported results of his experiments with photographic chemicals. This date forms a convenient, though arbitrary, milestone for the birth of photography. History has generally credited acquisition of the first aerial photograph to Gaspard-Félix Tournachon (1829–1910) (known also by his pseudonym, Nadar). In 1858, he acquired an aerial photo using a tethered balloon. Nadar's aerial photographs have been lost, although other early

balloon photographs survive. In succeeding years, numerous improvements were made in photographic technologies and methods of acquiring photographs of the Earth from balloons and kites. Early photography employed slow exposures, so the subjects in photos of the day often appear stiff and contrived because any movement would create blurred images. Photography from hills and mountainous terrain were of interest because of their scenic character and distant vistas. Thus, even before there were aerial cameras, there were aerial vistas from high buildings and mountainous terrain that engaged popular interest. Aerial images of the Earth are among the first to fit our definition of remote sensing, although many of these images must be regarded as curiosities rather than as foundations for the field of remote sensing.

The Camera and the Airplane

The use of powered aircraft as platforms for aerial photography formed the next milestone. In 1909, Wilbur Wright piloted the plane that acquired motion pictures of the Italian landscape near the town of Centocelli; these are said to be the first aerial photographs taken from an airplane. The maneuverability of the airplane provided the capability of controlling the speed, altitude, and direction required for systematic use of the airborne camera. Although there were many attempts to blend the camera with the airplane, the initial efforts were clearly not effective for systematic use with each other (Figure 1.1).

FIGURE 1.1 Early aerial photography by the U.S. Navy, 1914. This photograph illustrates the difficulties encountered in early efforts to match the camera with the airplane. At that time, neither device was well suited for use with the other. From U.S. Navy, National Archives and Records Administration, ARC 295605.

World War I (1914–1918)

World War I (1914–1918) marked the beginning of the acquisition of aerial photography on a systematic basis. The earliest efforts for aerial reconnaissance relied upon notes and sketches recorded by aerial observers in the rear cockpit. Soon, pilots and observers were using their own cameras, which were subsequently replaced by early aerial cameras specifically designed for aerial observation. Such cameras, typically handheld, required manual transfer of glass or metal plates. (Although, by that time, film cameras had been in use for many years, they were not routinely used for military applications until late in the conflict; **Figure 1.2**.)

Although cameras used for aerial photography during World War I were intended for use with the airplane, the match between the two instruments was still rudimentary (**Figure 1.3**). The value of aerial photography for military reconnaissance and surveillance became increasingly clear as the war continued, and aerial photography became increasingly sophisticated (see Campbell, 2008). By the conclusion of the conflict, aerial photography's role in military operations was valued, especially among pilots, observers, and aviation leadership, although senior military leadership remained skeptical. (In due course, the value of military aviation, including aerial observation, was recognized by the formation of the U.S. Army Air Corps in 1926, and eventually, in 1947, of the U.S. Air Force.)

FIGURE 1.2 Early aerial photography, World War I (ca. 1917–1918). By the time of World War I, attempts to match the camera and the airplane had progressed only to a modest extent, as illustrated by this example. This photographer, using a Graflex camera (here, likely posed for a publicity photograph), designed by the U.S. Army Signal Corps, aims the camera over the edge of the fuselage. The wooden camera body, with leather jacket, and focal plane shutter, required use of glass plates (likely 4 × 5 in.). Photographers operated within the cold slip-stream, manually changing plates in flight. From U.S. Army.

FIGURE 1.3 Aircraft showing later designs (relative to those in **Figures 1.1 and 1.2**) for both the biplane and the aerial camera. Here, the aerial camera has a lateral port (the square shape near the photographer, designed for oblique photography to allow the photographer to aim the camera through the port from within the fuselage, to avoid the disadvantages of leaning over edges of the cockpit. The photographer wears a chest-mounted microphone for communication with the pilot, shown here holding a supply of extra plates for the camera. From U.S. National Archives and Records Administration, Still Pictures, E-4156.

Interwar Years (1919–1939)

Numerous improvements followed from these beginnings. Camera designs were improved and tailored specifically for use in aircraft. The science of *photogrammetry*—the practice of making accurate measurements from photographs—applied specifically to aerial photography, with the development of instruments specifically designed for analysis of aerial images. Military applications of aerial photography drove rapid innovations of photographic technologies. During the postwar years, military budgets were tight, so much of the aviation effort was devoted to refining civil applications of aerial imagery. After the conclusion of the war, cameras and aircraft formed the basics for photogrammetric engineering. Although the fundamentals of photogrammetry had been previously defined, the field developed toward its modern form in the 1920s, with the development of specialized photogrammetric instruments.

During this period, the well-illustrated volume by Willis T. Lee (1922), *The Face of the Earth as Seen from the Air,* surveyed a range of civil applications of aerial photography. About 12 years after the end of World War I, Lee presented examples (**Figure 1.4**) of practical civil applications for aerial photography and of advanced aerial cameras. Many of his examples introduced applications beyond those developed during the war, including vertical aerial photographs and applications of panchromatic film. Lee used panchromatic films, developed by Kodak in 1913 for color imagery and, later (1918), for maintaining balanced tones to capture shades of gray recorded by the imagery. (Panchromatic

films provide black-and-white photographic emulsions sensitive to wavelengths of the visible spectrum. These films also provide balanced shades of gray that convey the tones observed in nature.)

Photogrammetric technologies, initially devised and perfected largely in Switzerland and Germany, provided the basis for broad-scale aerial mapping. These technologies were brought to the United States, initially by the U.S. Geological Survey and the Tennessee Valley Authority, then later, more generally by other organizations. From these origins, more or less routine applications of aerial mapping became significant for government programs and private industry, initially for topographic mapping, but later for soil survey, geologic mapping, forest surveys, highways, dams, and agricultural statistics.

Many innovations during this era were led by visionary pioneers who established successful niches in private industry to develop civil applications of aerial mapping. Sherman Fairchild (1896–1971) founded numerous companies, including Fairchild Aerial Surveys and Fairchild Camera and Instruments, which became leaders in aviation and in aerial camera design. Talbert Abrams (1895–1990) led many innovations in aerial survey, aviation, camera design, training, and worldwide commercial operations.

Although applications that Lee envisioned matured at a slow pace, expression of governmental interest ensured continuity in the scientific development of aerial photography and the training of many in uses of aerial photography. Nonetheless, the acceptance of aerial photography in governmental and scientific activities developed slowly because of resistance among traditionalists, imperfections in equipment and technique, and genuine uncertainties regarding the proper role of aerial photography in scientific inquiry and practical applications.

The worldwide economic depression of 1929–1939 was not only an economic and financial crisis but also, for many nations, an environmental crisis. National concerns about the social and economic impacts of rural economic development, widespread soil erosion, reliability of water supplies, and similar issues led to some of the early gov-

FIGURE 1.4 Aerial image acquired by the K-2 Eastman mapping camera, illustrating post–World War I aerial cameras with fuselage-mounted aerial cameras designed for vertical aerial photography. Here, the white plate positioned at the top of the image frame provides real-time records of altitude, date, time of day, and orientation of the lens. From Lee (1922).

ernmental applications of aerial surveys to record and monitor rural economic development. In the United States, the U.S. Department of Agriculture and the Tennessee Valley Authority led efforts to apply aerial photography to guide environmental planning and economic development. Such efforts made an important contribution to the institutionalization of the use of aerial photography in government and to the creation of a body of practical experience in applications of aerial photography (Figure 1.5).

World War II (1939–1945)

World War II (1939–1945) marks the next milestone in our history. During the war years, use of the electromagnetic spectrum extended from almost exclusive emphasis on the visible spectrum to other regions, most notably the infrared and microwave regions (far beyond the range of human vision). Knowledge of these regions of the spectrum had been developed in both basic and applied sciences during the preceding 150 years. However, during the war years, application and further development of this knowledge accelerated, as did dissemination of the means to apply it. Although research scientists had long understood the potential of the nonvisible spectrum, the equipment, materials, and

FIGURE 1.5 Progress in applications of aerial photography, 1919–1939. During the interval between World War I and World War II (1919–1939), integration of the camera and the airplane progressed, as did institutionalization of aerial photography in government and industry. By June 1943 (the date of this photograph), progress on both fronts was obvious. Here, an employee of the U.S. Geological Survey (USGS) uses a specialized instrument, the *Oblique Sketchmaster,* to match detail on an aerial photograph to an accurate map. By the time of this photograph, aerial photography formed an integral component of USGS operations. From USGS and U.S. Library of Congress, digital ID fsa.8d38549.

experience necessary to apply it to practical problems were not at hand. Wartime research and operational experience provided both the theoretical and the practical knowledge required for everyday use of the nonvisible spectrum in remote sensing.

Furthermore, the wartime training and experience of large numbers of pilots, camera operators, and photointerpreters created a large pool of experienced personnel who were able to transfer their skills and experience into civilian occupations after the war. Many of these people assumed leadership positions in the efforts of business, scientific, and governmental programs to apply aerial photography and remote sensing to a broad range of problems. Whereas photointerpreters of the World War I era focused on identification and examination of military equipment and fortifications, their counterparts in World War II also examined topography, vegetation, trafficability, and other terrain features, thereby expanding the scope and knowledge base and the practice of photointerpretation.

The Cold War (1946–1989)

The successes of strategic photointerpretation during World War II set the stage for continued interest in aerial surveillance during the cold war era. Initially, technological trends established during the war were continued and improved. However, as the nature of the cold war became more clearly defined, strategic photointerpretation formed one of the few means of acquiring reliable information from within the closed societies (**Figure 1.6**). Perhaps the best known contribution of photoreconnaissance within the cold war

FIGURE 1.6 A U.S. Air Force intelligence officer using a stereoscope to examine aerial photography, Korean conflict, July 1951. From U.S. Air Force, National Archives and Records Administration, ARC 542288.

era came during the Cuban Missile Crisis. In 1962, U.S. photointerpreters were able to detect with confidence the early stages of the Soviet Union's introduction of missiles into Cuba, far earlier than Soviet strategists had anticipated, thereby setting the stage for initiating one of the most serious incidents of the cold war era (see Brugioni, 1991). Such events created an environment for further development of advanced reconnaissance techniques. As newer, more sophisticated instruments were developed, the superseded technologies were released for wider, nondefense applications in the civilian economy (**Figures 1.7 and 1.8**).

Robert Colwell's Research

Among the most significant developments in the civilian sphere were the works of Robert N. Colwell (see Colwell, 1956, 1966, 1983; **Figure 1.9**), who applied color infrared film (then popularly known as "camouflage detection film," originally developed for use in World War II) to advance its applications for practice of aerial agriculture and forestry. During the war, Colwell served as an officer in the U.S. Naval Reserve, with expertise in applications of aerial photography for analysis of terrain, vegetation, and coastal environments. Later, Colwell joined the faculty of the University of California–Berkeley School of Forestry, where he developed applications of photographic interpretation to the practice of forestry. His research expanded to examine applications of the near-infrared spectrum to monitor vegetation health in the context of forestry and agricultural crops as well (**Figure 1.10**). Colwell's research defined the practice of multispectral remote sensing that today forms the basis for much of modern aerial imaging.

FIGURE 1.7 A 1950s forester examining aerial photography to delineate landscape units. By the 1950s, aerial photography and related forms of imagery had become integrated into the day-to-day operations of a multitude of businesses and industries throughout the world. From Forest History Society, Durham, North Carolina. Used by permission.

FIGURE 1.8 A *Twinplex* stereoscopic plotting instrument designed to prepare topographic maps to derive accurate elevation data from aerial photography (1957). During much of the twentieth century, photogrammetric analyses depended on optical–mechanical instruments such as shown here, designed to extract information by controlling the physical orientation of the photograph and optical projection of the image. By the end of the century, such processes were conducted in the digital domain using electronic instruments. From Photographic Library, USGS. Photograph by E. F. Patterson, no. 223.

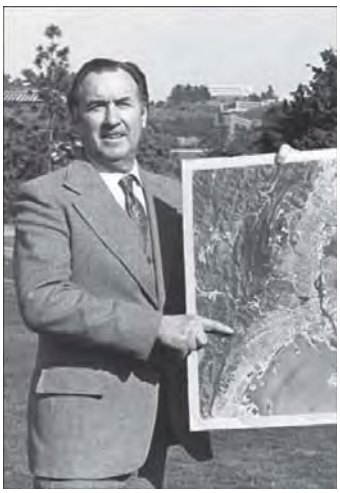


FIGURE 1.9 Robert Colwell, as photographed in 1956, Berkeley, California. Professor Colwell (1918–2005) served much of his career as a faculty member at the University of California–Berkeley, where his research defined key concepts in the practice of remote sensing, especially practical applications in forestry and agriculture. He is recognized internationally for his leadership in pioneering technological advances in the field of remote sensing of Earth resources. Here, Professor Colwell holds an aerial image depicting northeastern San Francisco. From University of California–Berkeley Library (Carol Ness, Berkeley News editor, *U.C. Berkeley News*), April 13, 2016. Used by permission.

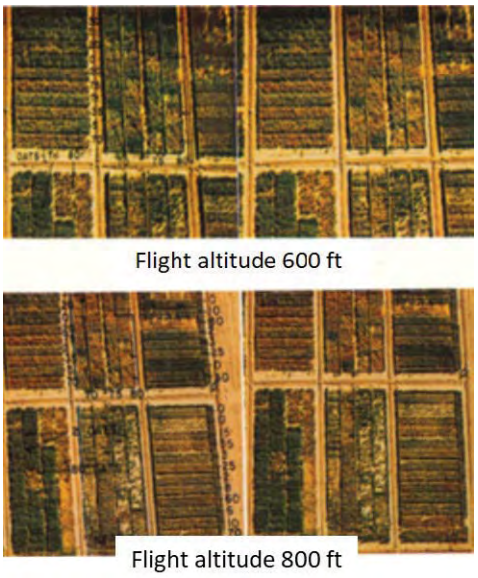


FIGURE 1.10 Robert Colwell's aerial photographs of experimental plots as seen from differing altitudes. From Colwell (1956). Copyright © 1956 Regents of the University of California. Used by permission.

Civil Applications of Aerial Imagery

During the 1950s, government and civil society accepted aerial photography as an important resource for applications in agriculture, forestry, and broad-scale resource inventory and as a source of cartographic information (Figures 1.7 and 1.8). During this period, some remote sensing capabilities were available for civilian use as superseded military applications were released for civil applications. For example, some instruments extended the reach of aerial observation outside the visible spectrum into the infrared and microwave regions; over subsequent decades, these instruments, including imaging radars, multispectral imaging, and related instruments, developed into especially effective sensor systems.

Remote Sensing

It was in this context that analysts began to accept the term *remote sensing* as the common vocabulary to describe the family of nonphotographic sensors. Evelyn Pruitt (1918–2000) (Figure 1.11), a scientist working for the U.S. Navy's Office of Naval Research (ONR), coined this term when she recognized that the term *aerial photography* no longer accurately described the many forms of imagery collected using radiation outside the visible spectrum. Pruitt studied geography at the University of California–Los Angeles, but also had interests in geology, meteorology, and related disciplines (Walker, 2006). In 1943, she began employment with the U.S. Coast and Geodetic Survey (C&GS) in Washington, D.C., and then worked for the newly formed ONR, which had been organized to conduct research to improve scientific knowledge of coastal environments. Her initial work for ONR focused on arctic studies but soon included broadly based research in coastal systems and processes in coastal regions worldwide.

As Pruitt worked with her colleagues to develop new photointerpretation techniques, she realized that continued use of phrases such as *photointerpretation* to describe imagery from nonphotographic sensors was off the mark. Thus, Pruitt has been credited with

initial use of the term *remote sensing* (Larson, 2004, in Walker, 2006) for this family of instruments. ONR, as one of the few organizations regularly engaged in use of multiple sensor systems, provided the context for this contribution.

Early in the 1960s, the U.S. National Aeronautics and Space Administration (NASA) established a research program in remote sensing—a program that, during the next decade, was to support remote sensing research at institutions throughout the United States. During this same period, a committee of the U.S. National Academy of Sciences (NAS) studied opportunities for application of remote sensing in the fields of agriculture and forestry. In 1970, the NAS reported the results of their work in a document that outlined many of the opportunities offered by this emerging field of inquiry (National Academy of Sciences, 1970).

Figures 1.12 and 1.13 show the manual interpretation of aerial imagery using techniques that would later be superseded by automated technologies.

Satellite Remote Sensing

In 1972, the launch of Landsat 1 (initially named the Earth Resources Technology Satellite [ERTS]), the first of many Earth-orbiting satellites specifically designed for observation of the Earth's land areas, provided another advancement of remote sensing technologies. Landsat provided, for the first time, systematic, repetitive observation of the Earth's land areas. Each Landsat image depicted large areas of the Earth's surface in several regions of the electromagnetic spectrum and yet provided modest levels of detail sufficient for practical applications in many fields.

The Landsat program evolved from the vision, skills, and engineering insight of several organizations, especially NASA's technical program and the U.S. Geological Survey (USGS), to anticipate the value of capturing images of the Earth's surface. William Pecora, chief geologist and later director of the USGS, made many notable contributions to the Landsat vision. His contributions addressed tasks of national scope, especially for his advocacy of satellite systems capable of observing Earth's surface—not only for mineral resources, but also for monitoring broad-scale views of the Earth's surface. His advocacy of such programs, supported also by other institutions, led to the formation of the Landsat satellite program.

Although it is difficult to recognize Landsat's full significance, it is possible to recognize three of its most important contributions. First, routine availability of multispec-

FIGURE 1.11 Evelyn Pruitt (1918–2000) in about 1973, probably at the time of her retirement from ONR. She is praised for her research in applying remote sensing analysis to study coastal environments. During her long career for ONR, she recognized that the term *photointerpretation* poorly described imagery from nonphotographic sensors used by ONR. She has been credited with use of the term *remote sensing* to describe this broader family of aerial imagery. From ONR.

FIGURE 1.12 A USGS cartographic technician using an airbrush to depict relief, as interpreted from aerial photographs, 1961. Within a few decades, computer cartography and GIS would routinely create this effect by applying hill-shading algorithms to create digital elevation models. From Photographic Library, USGS. Photograph by E. F. Patterson, no. 1024.

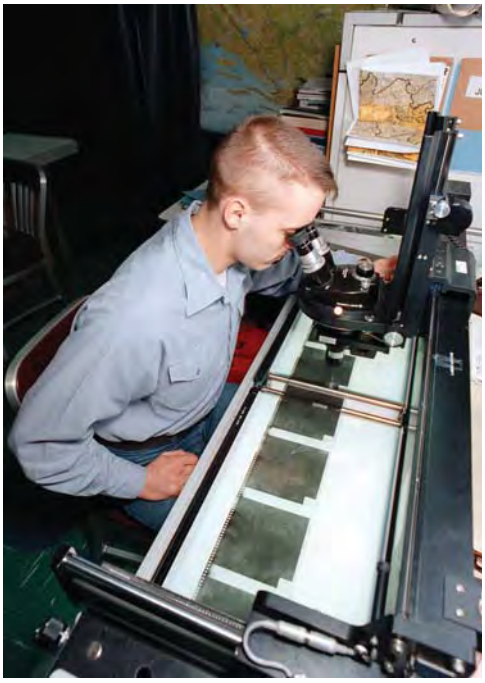


FIGURE 1.13 U.S. Navy Intelligence Specialist 3rd Class John Yanc using a binocular stereoscope to examine a strip of aerial photographs positioned on a light table on board the U.S. Navy's nuclear-powered aircraft carrier USS George Washington (CVN 73) (2010). From U.S. Navy.

tral data for large regions of the Earth's surface greatly expanded the ability of analysts of all varieties to acquire multispectral satellite imagery. Such data, though previously available, were largely confined to specialized research laboratories. The broad reach of Landsat's data, ease of access, and its open availability have greatly expanded the population of analysts who can acquire multispectral satellite data at varied dates and locations.

A second pivotal contribution of the Landsat program was the creation of an incentive for the rapid and broad expansion of digital analysis for remote sensing (**Figure 1.14**). Prior to Landsat, image analyses were largely completed by visual examination of prints and transparencies of aerial images. Analyses of digital images by computer were possible mainly in specialized research institutions; personal computers and the variety of image analysis programs that we now regard as commonplace, were not routinely available for most analyses. Routine availability of digital data in a standard format created the context that spurred growth of digital analyses and set the stage for developing image analysis software that is now commonplace. A third contribution of the Landsat program was its role as a model for developing other land observation satellites designed and operated by diverse organizations throughout the world.

Virginia Norwood

Finally, we note the value of satellite sensor systems, which, generally speaking, provide reliable observations of the Earth's surface. Many of us recognize the U.S. Landsat system as one of the oldest, most reliable, and established satellite observation systems. First,

FIGURE 1.14 NASA–Goddard staff working with Landsat/TIRS thermal imaging, March 13, 2012. Landsat's thermal instruments can record landscape temperatures from satellite altitudes to monitor water usage and water resources. Data from these instruments have provided reliable information documenting the availability of water resources for southwestern agriculture. From NASA–Goddard.

we recognize Virginia Norwood (Figure 1.15) for her significant contributions to the early Landsat systems. As a Hughes Aerospace engineer, she proposed using digital scanning technology for the first Landsat mission (launched in 1972). Her insight led to use of the Multispectral Scanner System (MSS), which, although intended as an experimental instrument, became the primary sensor for Landsat 1 when the primary sensor failed to perform as planned. The MSS's reliability and high-quality imagery formed a model for later Landsat instruments. This instrument therefore made a little-known contribution to the success of the Landsat system.

Free and Open Landsat Data Policy

Although U.S. policy long allowed distribution of satellite imagery to the public without cost, Landsat data formed an exception. For many years, individual Landsat data were distributed at costs varying from $600 to several thousand dollars, depending on various policies in effect during the period from 1973 to 2008. However, in 2008, the U.S. federal government implemented a policy that now distributes Landsat data without cost to the customer. This change in policy has greatly increased the number of downloads of Landsat imagery, creating a rapid expansion of scientific and operational applications, serving government, private sector, and civil society. Within U.S. governmental programs, the availability of imagery has greatly expanded and accelerated the scope of analyses. The Landsat Program is a model for other nations and space agencies (e.g., the European Copernicus Program), demonstrating the value of open access for Earth observation data, and has motivated other nations to implement similar policies (Zhu et al., 2016).

Unmanned Aerial Vehicles

Unmanned aerial vehicle (UAV) technologies and applications are a rapidly evolving area of remote sensing. UAVs, also known as "drones," are aircraft guided remotely by human pilots or by onboard programming (Figure 1.16). Hobbyists have a long history of interest in recreational aircraft, typically radio-controlled fixed-wing aircraft. Likely beginning in the 1990s, improvised UAV designs were created using miniaturized technologies derived from some of the technologies originally designed for mobile phones and similar

FIGURE 1.15 Virginia Norwood, a Hughes Aerospace engineer, whose insight and initiative formed the foundations for satellite observation. Photograph by Steve Covington of Aerospace Corp.

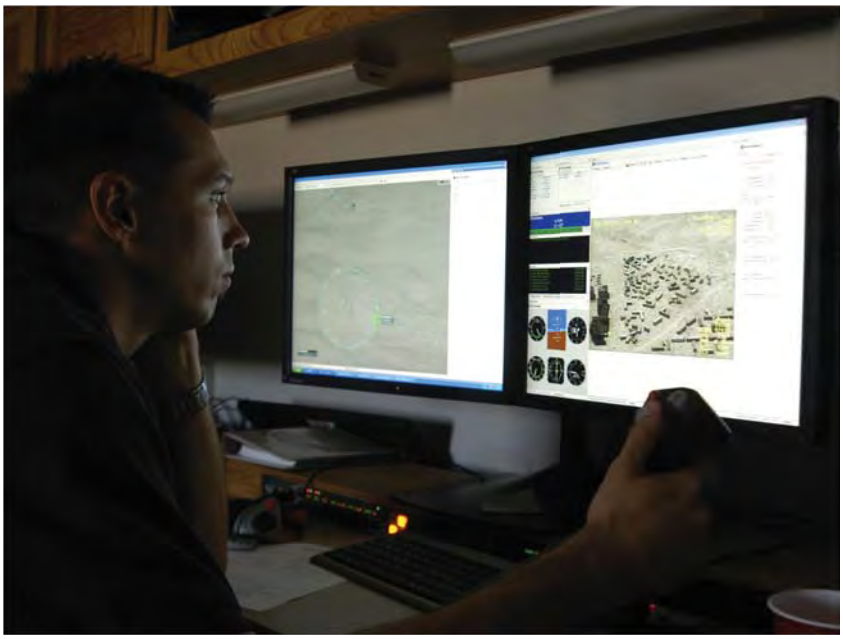


FIGURE 1.16 A systems analyst examining data and video transmitted by a U.S. Navy UAV to acquire intelligence, conduct surveillance, and collect reconnaissance data. From U.S. Marine Corps. Photograph by Cpl. Michael P. Snody.

instruments, including GPS, inertial measurement units (IMUs), infrared cameras, lasers, and remote control.

Although they vary widely, the principal UAV components include:

- Overall body (or frame)
- Power supply
- Computing
- Software
- Sensors
- Flight control
- Remote networking

UAV systems can include additional capabilities and can have a copter or fixed-wing design. Sizes and capabilities vary greatly according to purpose. For our discussions here, we think of UAVs of rather modest sizes (perhaps dimensions up to a meter), modest costs, modest capabilities, and limited range. These characteristics are especially practical for many civilian applications (Table 1.2). Because they can be rapidly deployed and are flexible in their acquisition parameters, UAV applications tend to focus on high-resolution acquisitions, particularly on-demand survey or monitoring tasks that may also complement more routine broader-scale remote sensing operations.

TABLE 1.2 Example UAV Applications

Aerial mapping	Precision agriculture	Mapping and surveying
Aerial photography	Forestry	Bridge and infrastructure monitoring
Weather survey	Irrigation	Land cover mapping
Search and rescue	Wildlife counts	Disaster management

1.4 SUMMARY

This chapter has introduced the basics of remote sensing, in part by reviewing technological changes over time as innovations provide new capabilities and resources for assessing and recording the status of the Earth's surface. Broadly stated, remote sensing monitors a landscape, city, or region by recording reflected or emitted radiation using one or more of the varied instruments that record radiation emitted from the Earth's surface. This chapter illustrates key historic contributions through innovation, insight, and experimentation.

We provide insight into how remote sensing technologies have matured from rough improvisations into systems that support collection of data, detect changes, assess landscapes, and maintain counts (vehicles, numbers of soldiers, and such, if we take the early days of World War I as an example). Postconflict developments saw remote sensing applications tailored to civic applications (forestry, agriculture, urban systems, and coastlines, for example) by applying approaches to a wider range of tasks supporting the broader needs of society.

As we progress through this book, we will explore the common tools, strategies and applications that form basic principles for the practice of remote sensing. We organize this discussion through chapters on common themes related to image acquisition, analysis, analytical techniques, and the use of remote sensing for several environmental applications.

REVIEW QUESTIONS

1. Satellite observation of the Earth provides many advantages over aircraft-borne sensors. Consider fields such as agronomy, forestry, or hydrology. For one such field of study, list as many of the advantages as you can. Can you suggest some disadvantages?
2. Much (but not all) information derived from remotely sensed data comes from spectral information. To understand how spectral data may not always be as reliable as one might first think, briefly describe the spectral properties of a maple tree and a cornfield. How might these properties change over the period of a year? Or a day?
3. All remotely sensed images observe the Earth from above. Can you list some advantages of the overhead view (as opposed to ground-level views) that make remote sensing images inherently advantageous for many purposes? List some disadvantages of the overhead view.
4. Remotely sensed images show the combined effects of many landscape elements, including vegetation, topography, illumination, soils, and drainage. In your view, is this diverse combination an advantage or a disadvantage? Explain.

5. List ways in which remotely sensed images differ from maps. Also, list the advantages and disadvantages of each. List some of the tasks for which each might be more useful.
6. This chapter emphasizes how the field of remote sensing is formed by knowledge and perspectives derived from many different disciplines. Examine the undergraduate catalog for your college or university and prepare a comprehensive program of study in remote sensing from courses listed. Identify gaps—courses or subjects that would be desirable but are not offered.
7. Inspect library copies of some of the remote sensing texts and journals listed in the references for this chapter. Examine the tables of contents, selected chapters, and lists of references. Many of these volumes may form useful references for future study or research in the field of remote sensing.

REFERENCES

Alföldi, T., P. Catt, and P. Stephens. 1993. Definitions of Remote Sensing. *Photogrammetric Engineering and Remote Sensing,* Vol. 59, pp. 611–613.

Avery, T. E., and G. L. Berlin. 1992. *Fundamentals of Remote Sensing and Airphoto Interpretation*. Upper Saddle River, NJ: Prentice-Hall, 472 pp.

Babington-Smith, C. 1957. *Air Spy: The Story of Photo Intelligence in World War II*. New York: Ballantine, 190 pp. (Reprinted 1985. Bethesda, MD: American Society for Photogrammetry and Remote Sensing, 266 pp.)

Barrett, E. C., and C. F. Curtis. 1976. *Introduction to Environmental Remote Sensing*. New York: Macmillan, 472 pp.

Brugioni, D. A. 1991. *Eyeball to Eyeball: The Inside Story of the Cuban Missile Crisis*. New York: Random House, 622 pp.

Campbell, J. B. 2008. Origins of Aerial Photographic Interpretation, U.S. Army, 1916–1918. *Photogrammetric Engineering and Remote Sensing,* Vol. 74, pp. 77–93.

Campbell, J. B., and V. V. Salomonson. 2010. Remote Sensing—A Look to the Future. Chapter 25 in *Manual of Geospatial Science and Technology* (2nd ed.) (J. D. Bossler, ed.). Boca Raton, FL: CRC Press,, pp. 487–510.

Campbell, J., T. Hardy, and R. C. Barnard. 2010. Emerging Markets for Satellite and Aerial Imagery. Chapter 22 in *Manual of Geospatial Science and Technology* (2nd ed.). (J. D. Bossler, ed.). Boca Raton, FL: CRC Press,, pp. 423–437.

Colwell, R. N. 1956. Determining the Prevalence of Certain Cereal Crop Diseases by Means of Aerial Photography. *Hilgardia,* Vol. 26, No. 5, pp. 223–286.

Colwell, R. N. 1966. Uses and Limitations of Multispectral Remote Sensing. In *Proceedings of the Fourth Symposium on Remote Sensing of Environment*. Ann Arbor: Institute of Science and Technology, University of Michigan, pp. 71–100.

Colwell, R. N. (ed.). 1983. *Manual of Remote Sensing* (2nd ed.). Falls Church, VA: American Society of Photogrammetry, 2 vols., 2240 pp.

Curran, P. 1987. Commentary: On Defining Remote Sensing. *Photogrammetric Engineering and Remote Sensing,* Vol. 53, pp. 305–306.

Day, D. A., J. M. Logsdon, and B. Latell (eds.). 1998. *Eye in the Sky: The Story of the Corona Spy Satellites*. Washington, DC: Smithsonian Institution Press, 303 pp.

Estes, J. E., J. R. Jensen, and D. S. Simonett. 1977. The Impact of Remote Sensing on United States' Geography: The Past in Perspective, Present Realities, Future Potentials. In *Proceedings of the Eleventh International Symposium on Remote Sensing of Environment*. Ann Arbor: Institute of Science and Technology, University of Michigan, pp. 101–121.

Executive Office of the President. 1973. *Report of the Federal Mapping Task Force on Mapping, Charting, Geodesy, and Surveying,* Washington, DC: Government Printing Office.
Fischer, W. A. (ed.). 1975. History of Remote Sensing. Chapter 2 in *Manual of Remote Sensing* (R. G. Reeves, ed.). Falls Church, VA: American Society of Photogrammetry, pp. 27–50.
Fischer, W. A., W. R. Hemphill, and A. Kover. 1976. Progress in Remote Sensing. *Photogrammetria,* Vol. 32, pp. 33–72.
Foody, G. M., T. A. Warner, and M. D. Nellis. 2009. A Look to the Future. Chapter 34 in *The Sage Handbook of Remote Sensing* (T. A. Warner, M. D. Nellis, and G. M. Foody, eds.). Washington, DC: SAGE, pp. 475–481.
Fussell, J., D. Rundquist, and J. A. Harrington. 1986. On Defining Remote Sensing. *Photogrammetric Engineering and Remote Sensing,* Vol. 52, pp. 1507–1511.
Goddard, G. W. 1969. *Overview: A Life-Long Adventure in Aerial Photography.* New York: Doubleday & Company, 415 pp.
Gorrell, E. S. 1917–1919. *Gorrell's History of the US Army Air Service.* Microfilmed typescript, T-619 (58 rolls). College Park, MD: U.S. National Archives and Records Administration.
Hall, S. S. 1993. *Mapping the Next Millennium: How Computer-Driven Cartography Is Revolutionizing the Face of Science.* New York: Random House, 360 pp.
Jensen, J. R. 2002. Remote Sensing—Future Considerations. Chapter 23 in *Manual of Geospatial Sciences and Technology* (J. D. Bossler, ed.). Boca Raton, FL: CRC Press, pp. 389–398.
Jensen, J. R. 2007. *Remote Sensing of the Environment: An Earth Resource Perspective.* Upper Saddle River, NJ: Prentice-Hall, 608 pp.
Landsat Legacy Project Team. 2017. *Landsat's Enduring Legacy: Pioneering Global Land Observations from Space.* Bethesda, MD: American Society for Photogrammetry and Remote Sensing.
Lee, W. T. 1922. *The Face of the Earth as Seen from the Air* (American Geographical Society Special Publication No. 4). New York: American Geographical Society.
Lillesand, T. M., R. W. Kiefer, and J. W. Chipman. 2008. *Remote Sensing and Image Interpretation.* New York: John Wiley, 756 pp.
Lintz, J., and D. S. Simonett. 1976. *Remote Sensing of Environment.* Reading, MA: Addison-Wesley, 694 pp.
Monmonier, M. 2002. Aerial Photography at the Agricultural Adjustment Administration: Acreage Controls, Conservation Benefits, and Overhead Surveillance in the 1930s. *Photogrammetric Engineering and Remote Sensing,* Vol. 68, pp. 1257–1261.
National Academy of Sciences. 1970. *Remote Sensing with Special Reference to Agriculture and Forestry.* Washington, DC: National Academy of Sciences, 424 pp.
Parker, D. C., and M. F. Wolff. 1965. Remote Sensing. *International Science and Technology,* Vol. 43, pp. 20–31.
Parker, H. D. 1988. The Unique Qualities of a Geographic Information System: A Commentary. *Photogrammetric Engineering and Remote Sensing,* Vol. 54, pp. 1547–1549.
Pedlow, G. W., and D. E. Welzenbach. 1992. *The Central Intelligence Agency and Overhead Reconnaissance: The U-2 and OXCART Programs, 1954–1974.* Washington, DC: Central Intelligence Agency.
Ray, R. G. 1960. *Aerial Photographs in Geological Interpretation and Mapping* (U.S. Geological Survey Professional Paper 373). Washington, DC: U.S. Geological Survey, 230 pp.
Reeves, R. G. (ed.). 1975. *Manual of Remote Sensing.* Falls Church, VA: American Society of Photogrammetry, 2 vols., 2144 pp.
Ryerson, B., and S. Aronoff. 2010. *Why "Where" Matters: Understanding and Profiting from GPS, GIS, and Remote Sensing.* Manotick, ON: Kim Geomatics, 378 pp.
Simonett, D. S. 1983. Development and Principles of Remote Sensing. Chapter 1 in *Manual of Remote Sensing* (R. N. Colwell, ed.). Falls Church, VA: American Society of Photogrammetry, pp. 1–35.

Swain, P. H., and S. M. Davis. 1978. *Remote Sensing: The Quantitative Approach.* New York: McGraw-Hill, 396 pp.

Thenkabail, P. S. (ed.). 2018. *Remote Sensing Handbook*. Boca Raton, FL: Taylor & Francis, 3 vols., 2304 pp.

Xue, Y., A. P. Cracknell, and H. D. Guo. 2002. The Integration of Remote Sensing, Geographic Information Systems, Global Positioning Systems and Telecommunication. *International Journal of Remote Sensing,* Vol. 23, pp. 1851–1893.

Walker, J. H. 2006. Evelyn Lord Pruitt, 1918–2000. *Annals of the Association of American Geographers,* Vol. 96, pp. 432–439.

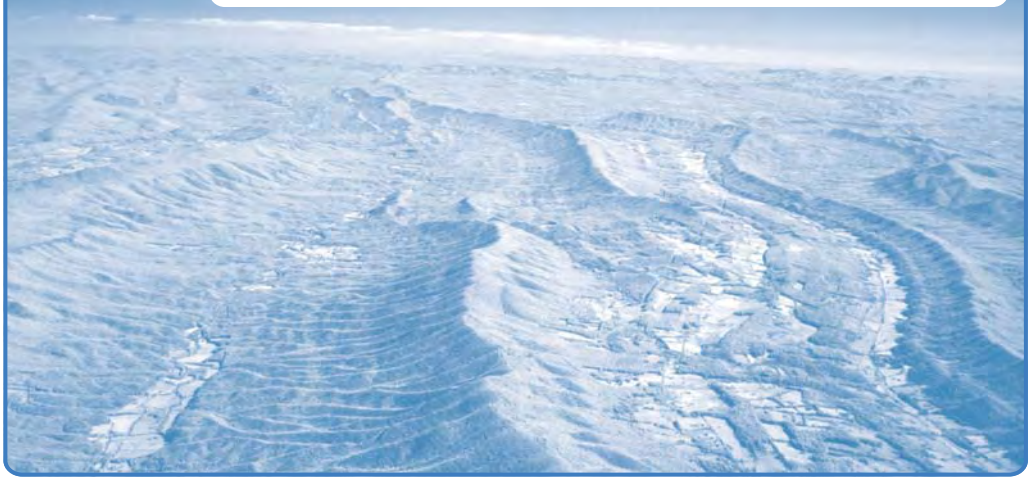
Warner, T. A., M. D. Nellis, and G. M. Foody (eds.). 2009. *The Sage Handbook of Remote Sensing.* Washington, DC: SAGE, 504 pp.

White, L. P. 1977. *Aerial Photography and Remote Sensing for Soil Survey.* Oxford: Clarendon Press, 104 pp.

Woodcock, C. E., R. Allen, M. Anderson, A. Belward, R. Bindschadler, W. Cohen, . . . and R. Wynne. 2008, May 23. Free Access to Landsat Imagery. *Science*, Vol. 320 (5879), p. 1011.

Zhu, Z., M. A. Wulder, D. P. Roy, C. E. Woodcock, M. C. Hansen, V. C. Radeloff, . . . and T. A. Scambos. 2016. The Benefits of the Free and Open Landsat Data Policy. *Remote Sensing of Environment,* Vol. 224, pp. 382–385.

2 Electromagnetic Radiation



MAJOR TOPICS TO UNDERSTAND

- The Electromagnetic Spectrum
- Major Divisions of the Electromagnetic Spectrum
- Radiation Laws
- Interactions with the Atmosphere
- Interactions with Surfaces

2.1 INTRODUCTION

With the exception of objects at absolute zero, all objects emit electromagnetic radiation. Objects also reflect radiation that has been emitted by other objects. By recording emitted or reflected radiation and applying knowledge of its behavior as it passes through the Earth's atmosphere and interacts with objects, remote sensing analysts develop knowledge of the character of features such as vegetation, structures, soils, rock, or water bodies on the Earth's surface. Interpretation of remote sensing imagery depends on a sound understanding of electromagnetic radiation and its interaction with surfaces and

the atmosphere. The discussion of electromagnetic radiation in this chapter builds a foundation that will permit development in subsequent chapters of the many other important topics within the field of remote sensing.

The most familiar form of electromagnetic radiation is visible light, which forms only a small (but very important) portion of the full electromagnetic spectrum. The large segments of this spectrum that lie outside the range of human vision require our special attention because they may behave in ways that are quite foreign to our everyday experience with visible radiation.

2.2 THE ELECTROMAGNETIC SPECTRUM

Electromagnetic energy is generated by several mechanisms, including changes in the energy levels of electrons, acceleration of electrical charges, decay of radioactive substances, and the thermal motion of atoms and molecules. Nuclear reactions within the Sun produce a full spectrum of electromagnetic radiation, which is transmitted through space without experiencing major changes. As this radiation approaches the Earth, it passes through the atmosphere before reaching the Earth's surface. Some is reflected upward from the Earth's surface; it is this radiation that forms the basis for photographs and similar images. Other solar radiation is absorbed at the surface of the Earth and is then reradiated as thermal energy. This thermal energy can also be used to form remotely sensed images, although they differ greatly from the aerial photographs formed from reflected energy. Finally, human-made radiation, such as that generated by imaging radars, is also used for remote sensing.

Electromagnetic radiation consists of an electrical field (E) that varies in magnitude in a direction perpendicular to the direction of propagation (Figure 2.1). In addition, a magnetic field (H) oriented at right angles to the electrical field is propagated in phase with the electrical field.

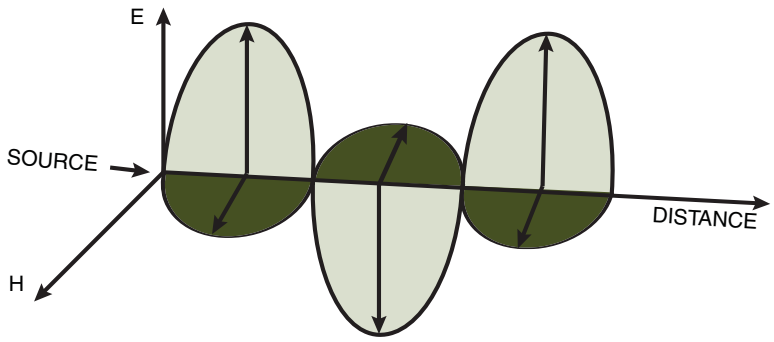


FIGURE 2.1 Electric (E) and magnetic (H) components of electromagnetic radiation. The electric and magnetic components are oriented at right angles to one another and vary along an axis perpendicular to the axis of propagation. Image by Susmita Sen.

Electromagnetic energy can be characterized by several properties (Figure 2.2):

1. *Wavelength* is the distance from one wave crest to the next. Wavelength can be measured in everyday units of length, although very short wavelengths have such small distances between wave crests that extremely short (and therefore less familiar) measurement units are required (Table 2.1).
2. *Frequency* is measured as the number of crests passing a fixed point in a given period of time. Frequency is often measured in *hertz*, each of which is equivalent to one cycle per second (Table 2.2), and multiples of hertz.
3. *Amplitude* is equivalent to the height of each peak (see Figure 2.2). Amplitude is often measured as energy levels (formally known as *spectral irradiance*), expressed as watts per square meter per micrometer (i.e., as energy level per wavelength interval).
4. In addition, the *phase* of a waveform specifies the extent to which the peaks of one waveform align with those of another. Phase is measured in angular units, such as degrees or radians. If two waves are aligned, they oscillate together and are said to be "in phase" (a phase shift of 0 degrees). However, if a pair of waves is aligned such that the crests match with the troughs, they are said to be "out of phase" (a phase shift of 180 degrees) (see Figure 2.2).

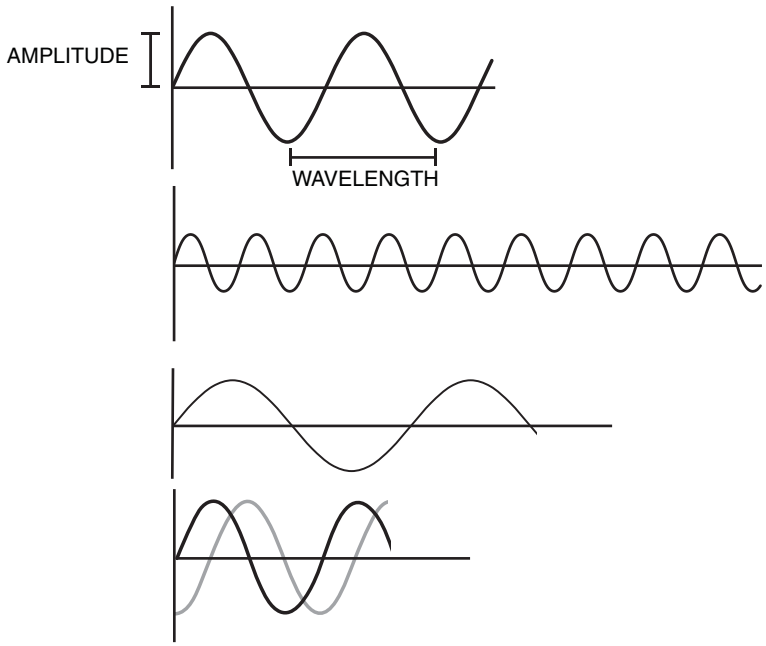


FIGURE 2.2 Amplitude, frequency, and wavelength. The second diagram represents high frequency, short wavelength; the third, low frequency, long wavelength. The bottom diagram illustrates two waveforms that are out of phase. Image by Susmita Sen.

TABLE 2.1 Units of Length Used in Remote Sensing

Unit	Distance
Kilometer (km)	1,000 m
Meter (m)	1.0 m
Centimeter (cm)	0.01 m = 10^{-2} m
Millimeter (mm)	0.001 m = 10^{-3} m
Micrometer (μm)[a]	0.000001 m = 10^{-6} m
Nanometer (nm)	10^{-9} m
Ångström unit (Å)	10^{-10} m

[a]Formerly called the micron (μ); the term *micrometer* is now used by agreement of the General Conference on Weights and Measures.

TABLE 2.2 Frequencies Used in Remote Sensing

Unit	Frequency (cycles per second)
Hertz (Hz)	1
Kilohertz (kHz)	10^3 (= 1,000)
Megahertz (MHz)	10^6 (= 1,000,000)
Gigahertz (GHz)	10^9 (= 1,000,000,000)

The speed of electromagnetic energy (c) is constant at 299,792 kilometers (km) per second. Frequency (ν) and wavelength (λ) are related:

$$c = \lambda\nu \qquad \text{(EQ. 2.1)}$$

Therefore, characteristics of electromagnetic energy can be specified using either frequency or wavelength. Varied disciplines and varied applications follow different conventions for describing electromagnetic radiation, using either wavelength (measured in Ångström units [Å], microns/micrometers, nanometers, millimeters, etc., as appropriate) or frequency (using hertz, kilohertz, megahertz, etc., as appropriate). Although there is no authoritative standard, a common practice in the field of remote sensing is to define regions of the spectrum on the basis of wavelength, often using micrometers (each equal to one one-millionth of a meter, symbolized as μm), millimeters (mm), and meters (m) as units of length. Departures from this practice are common; for example, electrical engineers who work with microwave radiation traditionally use frequency to designate subdivisions of the spectrum. In this book, we usually employ wavelength designations. The student should, however, be prepared to encounter different usages in scientific journals and in references.

2.3 MAJOR DIVISIONS OF THE ELECTROMAGNETIC SPECTRUM

Major divisions of the electromagnetic spectrum (Table 2.3) are, in essence, arbitrarily defined. In a full spectrum of solar energy there are no sharp breaks at the divisions, as indicated graphically in Figure 2.3. Subdivisions are established for convenience and by traditions within different disciplines, so do not be surprised to find different definitions in other sources or in references pertaining to other disciplines.

Table 2.3 does not show two important categories: the optical and reflective spectra. The *optical spectrum*, from 0.30 to 15 μm, defines those wavelengths that can be reflected and refracted with lenses and mirrors. The *reflective spectrum* extends from about 0.38 to 3.0 μm and defines that portion of the solar spectrum used directly for remote sensing. As an introduction, Figure 2.3 presents subdivisions for the electromagnetic spectrum, as often defined within the field of remote sensing; specifics are discussed in subsequent chapters.

- *Visible:* The short range of wavelengths is defined by the sensitivity of the human visual system for films, sensors, and imagery.
- *NIR (near-infrared region):* The NIR (about 0.7 μm–1.4 μm), outside the visible spectrum, is valuable mainly because it is not subject to atmospheric scattering and because of its effectiveness in detecting and monitoring living vegetation.
- *VNIR* (*visible and near-infrared region*): The VNIR (0.4–1.4 μm) is defined by the full visible spectrum, extended to include the adjacent region of the infrared spectrum.
- *SWIR* (*shortwave infrared region*): The SWIR (sometimes defined as 0.9–1.7 μm or as 0.7–2.5 μm) is effective because it is not subject to atmospheric scattering, and yet it is effective for the mapping of minerals, fires, crop health, and surface moisture.
- *Thermal:* Thermal radiation, designated here by its broad range of wavelengths within the 8 μm–14 μm atmospheric window, conveys temperature information.

TABLE 2.3 Principal Divisions of the Electromagnetic Spectrum

Division	Limits
Gamma rays	< 0.03 nm
X-rays	0.03–300 nm
Ultraviolet radiation	0.30–0.38 μm
Visible light	0.38–0.72 μm
Infrared radiation	
Near infrared	0.72–1.30 μm
Mid infrared	1.30–3.00 μm
Far infrared	7.0–1,000 μm (1 mm)
Microwave radiation	1 mm–30 cm
Radio	≥ 30 cm

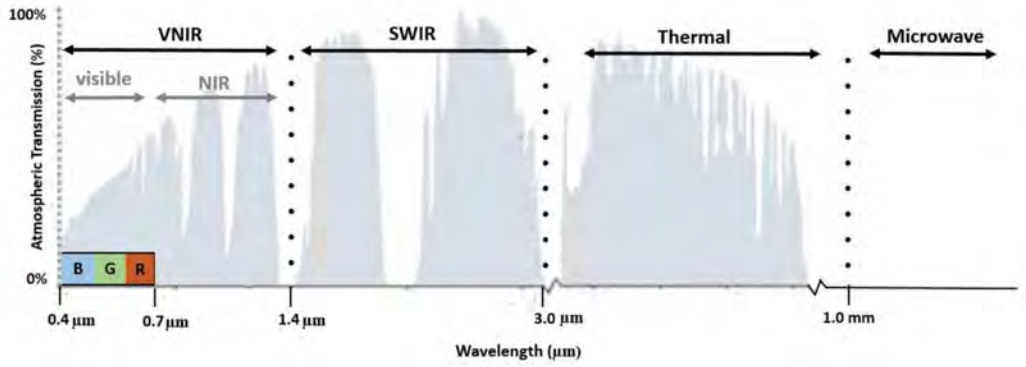


FIGURE 2.3 Some important features of the electromagnetic spectrum, labeled to identify key aspects for the practice of remote sensing (not presented at scale). VNIR = visible and near-infrared portion of the electromagnetic spectrum; SWIR = shortwave and near-infrared region. Image by Susmita Sen.

(Technically, thermal radiation covers a wide range of wavelengths, but as a practical matter, it is defined here more narrowly to focus wavelengths that are largely free of atmospheric scattering.)

- *Microwave:* Microwave radiation is often defined as 1.0 µm–1.0 m. Shorter wavelengths of the microwave region can convey thermal information, although the microwave region is especially significant for active remote sensing (using imaging radars; see Chapter 8) and both aircraft and satellite sensors.

The Ultraviolet Spectrum

For practical purposes, radiation of significance for remote sensing can be said to begin with the ultraviolet region, a zone of short-wavelength radiation that lies between the X-ray region and the limit of human vision. Often, the ultraviolet region is subdivided into the near ultraviolet (sometimes known as UV-A; 0.32–0.40 µm), the far ultraviolet (UV-B; 0.28–0.32 µm), and the extreme ultraviolet (UV-C; below 0.28 µm). The ultraviolet region was discovered in 1801 by the German scientist Johann Wilhelm Ritter (1776–1810). Literally, ultraviolet means "beyond the violet," designating it as the region just outside the violet region, the shortest wavelengths visible to humans. Near-ultraviolet radiation is known for its ability to induce fluorescence (emission of visible radiation) in some materials; it has significance for a specialized form of remote sensing (see Section 2.6). However, ultraviolet radiation is easily scattered by the Earth's atmosphere, so it is not generally used for remote sensing of Earth materials.

The Visible Spectrum

Although the visible spectrum constitutes a very small portion of the spectrum, it has obvious significance in remote sensing. Colors of features depend on the physical characteristics of objects and the circumstances of their perception. For example, color can be defined by the nature of light from surfaces, usually depending on the spectrum of

incident radiation, and viewing and illumination angles. Perceived colors are associated with wavelets as reflected from objects and their physical properties such as light absorption and emission spectra.

Limits of the visible spectrum are defined by the sensitivity of the human visual system. Optical properties of visible radiation were first investigated by Isaac Newton (1641–1727), who, during 1665 and 1666, conducted experiments revealing that visible light can be divided (using prisms, or, in our time, diffraction gratings) into three segments. Today we know these segments as the *additive primaries,* defined approximately from 0.4 to 0.5 μm (blue), 0.5 to 0.6 μm (green), and 0.6 to 0.7 μm (red) (**Figure 2.4**). *Primary colors* are defined such that no single primary can be formed from a mixture of the other two and that all other colors can be formed by mixing the three primaries in appropriate proportions. Equal proportions of the three additive primaries combine to form white light.

The color of an object is defined by the color of the light that it reflects (**Figure 2.4**). Thus, a "blue" object is "blue" because it reflects blue light. Intermediate colors are formed when an object reflects two or more of the additive primaries, which combine to create the sensation of "yellow" (red and green), "purple" (red and blue), or other colors. The additive primaries are significant whenever we consider the colors of light, as, for example, in the exposure of photographic films.

In contrast, *representations* of colors in films, paintings, and similar images are formed by combinations of the three *subtractive primaries* that define the colors of pigments and dyes. Each of the three subtractive primaries absorbs a third of the visible spectrum (**Figure 2.4**). *Yellow* absorbs blue light (and reflects red and green), *cyan* (a greenish-blue) absorbs red light (and reflects blue and green), and *magenta* (a bluish red) absorbs green light (and reflects red and blue light). A mixture of equal proportions of pigments of the three subtractive primaries yields black (complete absorption of the

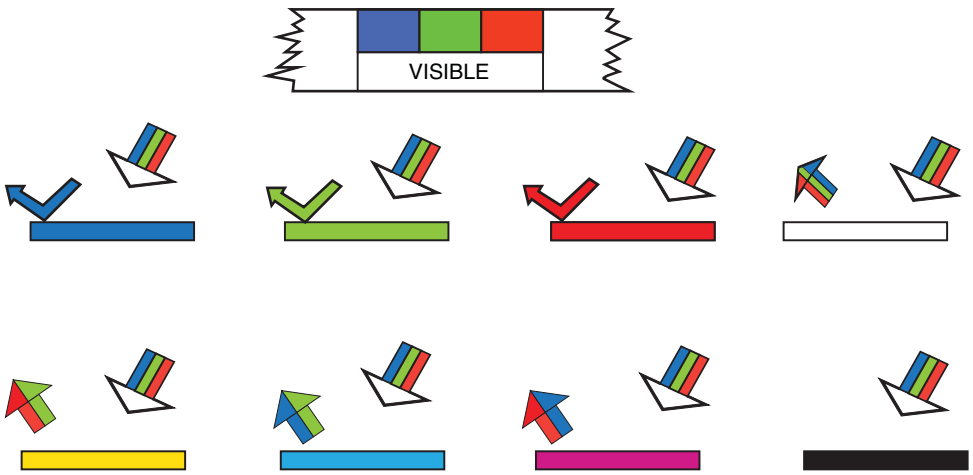


FIGURE 2.4 Color in the visible spectrum. Color is characterized by behavior of light within the visible spectrum, defined by behavior of the *additive primaries* (top row; colors of light, as reflected from objects), and the *subtractive primaries* (bottom row; colors of pigments and dyes as they reflect light). Image by Susmita Sen.

visible spectrum). The additive primaries are of interest in matters concerning radiant energy, whereas the subtractive primaries specify colors of the pigments and dyes used in reproducing colors on films, photographic prints, and other images.

The Infrared Spectrum

Wavelengths longer than the red portion of the visible spectrum are designated as the infrared region, discovered in 1800 by the British astronomer William Herschel (1738–1822). This segment of the spectrum is very large relative to the visible region, as it extends from 0.72 to 15 μm, making it more than 40 times as wide as the visible light spectrum. Because of its broad range, it encompasses radiation with varied properties. Two important categories can be recognized here. The first consists of *near-infrared* and *mid-infrared* radiation, defined as those regions of the infrared spectrum closest to the visible. Radiation in the near-infrared region behaves, with respect to optical systems, in a manner analogous to radiation in the visible spectrum. Therefore, remote sensing in the near-infrared region can use films, filters, and cameras with designs similar to those intended for use with visible light.

The second category of infrared radiation is the *far-infrared* region, consisting of wavelengths well beyond the visible, extending into regions that border the microwave region (Table 2.3). This radiation is fundamentally different from that in the visible and the near-infrared regions. Whereas near-infrared radiation is essentially solar radiation reflected from the Earth's surface, far-infrared radiation is emitted by the Earth. In everyday language, the far infrared consists of "heat," or "thermal energy." Sometimes this portion of the spectrum is referred to as the *emitted infrared*.

Microwave Energy

The longest wavelengths commonly used in remote sensing are those from about 1 mm to 1 μm in wavelength. The shortest wavelengths in this range have much in common with the thermal energy of the far infrared. The longer wavelengths of the microwave region merge into the radio wavelengths used for commercial broadcasts. Our knowledge of the microwave region originates from the work of the Scottish physicist James Clerk Maxwell (1831–1879) and the German physicist Heinrich Hertz (1857–1894).

2.4 RADIATION LAWS

The propagation of electromagnetic energy follows certain physical laws. In the interest of conciseness, some of these laws are outlined in abbreviated form because our interest here is the basic relationships they express rather than the formal derivations that are available to the student in more comprehensive sources.

Isaac Newton was among the first to recognize the dual nature of light (and, by extension, all forms of electromagnetic radiation), which simultaneously displays behaviors associated with both discrete and continuous phenomena. Newton maintained that light is a stream of minuscule particles ("corpuscles") that travel in straight lines. This notion is consistent with the modern theories of Max Planck (1858–1947) and Albert Einstein (1879–1955). Planck discovered that electromagnetic energy is absorbed and emitted in discrete units called *quanta,* or *photons*. The size of each unit is directly

proportional to the frequency of the energy's radiation. Planck defined a constant (h) to relate frequency (υ) to radiant energy (Q):

$$Q = h\upsilon \quad \text{(EQ. 2.2)}$$

His model explains the photoelectric effect, the generation of electric currents by the exposure of certain substances to light, as the effect of the impact of these discrete units of energy (quanta) on surfaces of certain metals, causing the emission of electrons.

Newton knew of other phenomena, such as the refraction of light by prisms, that are best explained by assuming that electromagnetic energy travels in a wave-like manner. James Clerk Maxwell was the first to formally define the wave model of electromagnetic radiation. His mathematical definitions of the behavior of electromagnetic energy are based on the assumption from classical (mechanical) physics that light and other forms of electromagnetic energy propagate as a series of waves. The wave model best explains some aspects of the observed behavior of electromagnetic energy (e.g., refraction by lenses and prisms and diffraction), whereas quantum theory provides explanations of other phenomena (notably, the photoelectric effect).

The rate at which photons (quanta) strike a surface is the *radiant flux* (ϕ_e), measured in watts (W); this measure specifies energy delivered to a surface in a unit of time. We also need to specify a unit of area; the *irradiance* (E_e) is defined as radiant flux per unit area (usually measured as watts per square meter). Irradiance measures radiation that strikes a surface, whereas the term *radiant exitance* (M_e) defines the rate at which radiation is emitted from a unit area (also measured in watts per square meter).

All objects with temperatures above absolute zero have temperature and emit energy. The amount of energy and the wavelengths at which it is emitted depend on the temperature of the object. As the temperature of an object increases, the total amount of energy emitted also increases, and the wavelength of maximum (peak) emission becomes shorter. These relationships can be expressed formally using the concept of the *blackbody*. A blackbody is a hypothetical source of energy that behaves in an idealized manner. It absorbs all incident radiation; none is reflected. A blackbody emits energy with perfect efficiency; its effectiveness as a radiator of energy varies only as temperature varies.

The blackbody is a hypothetical entity because in nature all objects reflect at least a small proportion of the radiation that strikes them and thus do not act as perfect reradiators of absorbed energy. Although truly perfect blackbodies cannot exist, their behavior can be approximated using laboratory instruments. Such instruments have formed the basis for the scientific research that has defined relationships between the temperatures of objects and the radiation they emit. *Kirchhoff's law* states that the ratio of emitted radiation to absorbed radiation flux is the same for all blackbodies at the same temperature. This law forms the basis for the definition of *emissivity* (ε), the ratio between the emittance of a given object (M) and that of a blackbody at the same temperature (M_b):

$$\varepsilon = M/M_b \quad \text{(EQ. 2.3)}$$

The emissivity of a true blackbody is 1, and that of a perfect reflector (a *whitebody*) would be 0. Blackbodies and whitebodies are hypothetical concepts, approximated in the laboratory under contrived conditions. In nature, all objects have emissivities that fall between these extremes (*graybodies*). For these objects, emissivity is a useful measure of their effectiveness as radiators of electromagnetic energy. Those objects that tend to

absorb high proportions of incident radiation and then reradiate this energy will have high emissivities. Those that are less effective as absorbers and radiators of energy have low emissivities (i.e., they return much more of the energy that reaches them). (In Chapter 10, further discussion of emissivity explains that emissivity of an object can vary with its temperature.)

The *Stefan–Boltzmann law* defines the relationship between the total emitted radiation (W) (often expressed in watts · cm^{-2}) and temperature (T) (absolute temperature, K):

$$W = \sigma T^4 \qquad \text{(EQ. 2.4)}$$

Total radiation emitted from a blackbody is proportional to the fourth power of its absolute temperature. The constant (σ) is the Stefan–Boltzmann constant (5.6697×10^{-8}) (watts · m^{-2} · K^{-4}), which defines unit time and unit area. In essence, the Stefan–Boltzmann law states that hot blackbodies emit more energy per unit area than do cool blackbodies.

Wien's displacement law specifies the relationship between the wavelength of radiation emitted and the temperature of a blackbody:

$$\lambda = 2{,}897.8/T \qquad \text{(EQ. 2.5)}$$

where λ is the wavelength at which radiance is at a maximum and T is the absolute temperature (K). As blackbodies become hotter, the wavelength of maximum emittance shifts to shorter wavelengths (Figure 2.5).

All three of these radiation laws are important for understanding electromagnetic radiation. They have special significance later in discussions of detection of radiation in the far-infrared spectrum (Chapter 10).

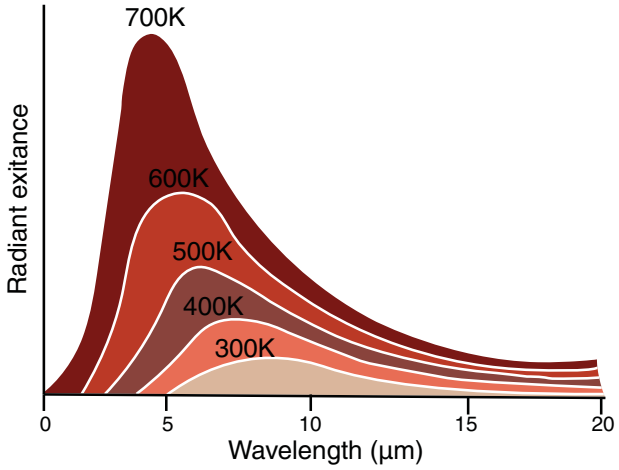


FIGURE 2.5 Wien's displacement law. For blackbodies at high temperatures, maximum radiation emission occurs at short wavelengths. Blackbodies at low temperatures emit maximum radiation at longer wavelengths. Image by Susmita Sen.

2.5 INTERACTIONS WITH THE ATMOSPHERE

All radiation used for remote sensing must pass through the Earth's atmosphere. If the sensor is carried by a low-flying aircraft, the effects of the atmosphere on image quality may be negligible. In contrast, energy that reaches sensors carried by Earth satellites (Chapter 7) must pass through the *entire depth* of the Earth's atmosphere. Under these conditions, atmospheric effects may have substantial impact on the quality of images and data that the sensors generate. Therefore, the practice of remote sensing requires knowledge of interactions of electromagnetic energy with the atmosphere.

In cities we are often acutely aware of the visual effects of dust, smoke, haze, and other atmospheric impurities due to their high concentrations. We easily appreciate their effects on brightnesses and colors we see. But even in clear air, the visual effects of the atmosphere are numerous, though so commonplace that we may not recognize their significance. In both settings, as solar energy passes through the Earth's atmosphere, it is subject to modification by several physical processes, including (1) scattering, (2) refraction, and (3) absorption.

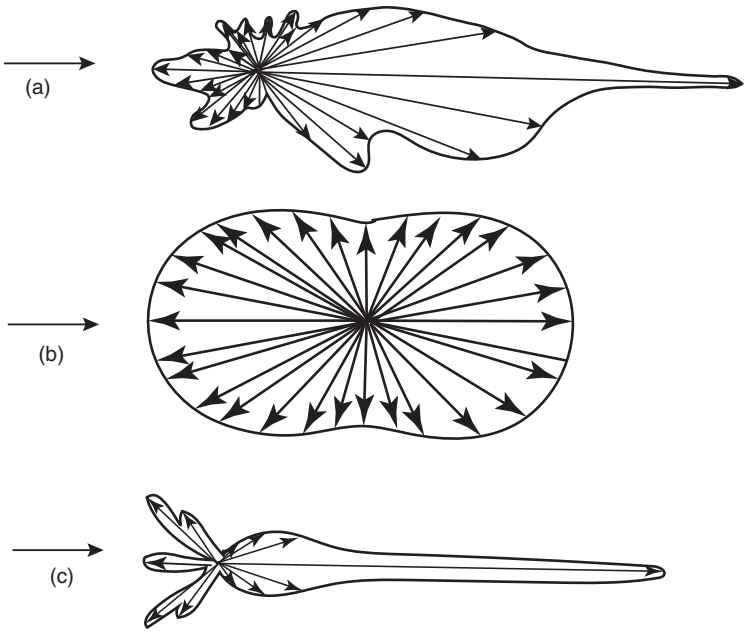


FIGURE 2.6 Scattering behaviors of three classes of atmospheric particles. (a) Atmospheric dust and smoke form rather large, irregular particles that create a strong forward-scattering peak, with a smaller degree of backscattering. (b) Atmospheric molecules are more nearly symmetric in shape, creating a pattern characterized by preferential forward- and backscattering, but without the pronounced peaks observed in the first example. (c) Large water droplets create a pronounced forward-scattering peak, with smaller backscattering peaks. From Lynch and Livingston (1995). Used by permission of Cambridge University Press.

Scattering

Scattering is the redirection of electromagnetic energy by particles suspended in the atmosphere or by large molecules of atmospheric gases (Figure 2.6). The amount of scattering that occurs depends on the sizes of these particles, their abundance, the wavelength of the radiation, and the depth of the atmosphere through which the energy is traveling. The effect of scattering is to redirect radiation so that a portion of the incoming solar beam is directed back toward space, as well as toward the Earth's surface.

A common form of scattering was discovered by the British scientist Lord J. W. S. Rayleigh (1824–1919) in the late 1890s. He demonstrated that a perfectly clean atmosphere, consisting only of atmospheric gases, causes scattering of light in a manner such that the amount of scattering increases greatly as the wavelength becomes shorter. *Rayleigh scattering* occurs when atmospheric particles have diameters that are very small relative to the wavelength of the radiation. Typically, such particles could be very small specks of dust or some of the larger molecules of atmospheric gases, such as nitrogen (N_2) and oxygen (O_2). These particles have diameters that are much smaller than the wavelength (λ) of visible and near-infrared radiation (on the order of diameters less than λ).

Because Rayleigh scattering can occur in the absence of atmospheric impurities, it is sometimes referred to as *clear atmosphere scattering*. It is the dominant scattering process high in the atmosphere, up to altitudes of 9–10 km, which is the upper limit for atmospheric scattering. Rayleigh scattering is *wavelength-dependent,* meaning that the amount of scattering changes greatly as one examines different regions of the spectrum (Figure 2.7). Blue light is scattered about four times as much as is red light, and ultraviolet light is scattered almost 16 times as much as is red light. *Rayleigh's law* states that this form of scattering is in proportion to the inverse of the fourth power of the wavelength.

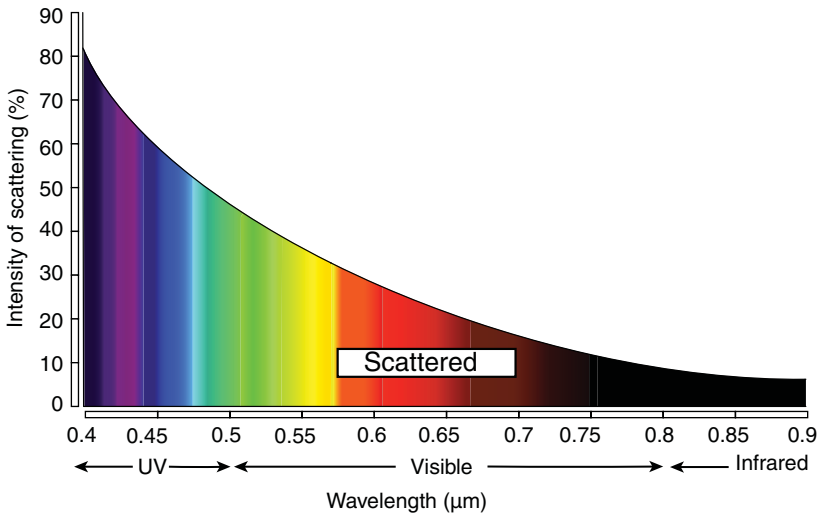


FIGURE 2.7 Rayleigh scattering. The effect of Rayleigh scattering is much higher at shorter wavelengths. Image by Susmita Sen.

Rayleigh scattering is the cause of both the blue color of the sky and the brilliant red and orange colors often seen at sunset. At midday, when the Sun is high in the sky, the atmospheric path of the solar beam is relatively short and direct, so an observer at the Earth's surface sees mainly the blue light preferentially redirected by Rayleigh scatter. At sunset, observers on the Earth's surface see only those wavelengths that pass through the longer atmospheric path caused by the low solar elevation; because only the longer wavelengths penetrate this distance without attenuation by scattering, we see only the reddish component of the solar beam. Variations of concentrations of fine atmospheric dust or of tiny water droplets in the atmosphere may contribute to variations in atmospheric clarity and therefore to variations in the colors of sunsets.

Although Rayleigh scattering forms an important component of our understanding of atmospheric effects on transmission of radiation in and near the visible spectrum, it applies only to a rather specific class of atmospheric interactions. In 1906, the German physicist Gustav Mie (1868–1957) published an analysis that describes atmospheric scattering involving a broader range of atmospheric particles. *Mie scattering* is caused by large atmospheric particles, including dust, pollen, smoke, and water droplets. Such particles may seem to be very small by the standards of everyday experience, but they are many times larger than those responsible for Rayleigh scattering. Those particles that cause Mie scattering have diameters that are roughly equivalent to the wavelength of the scattered radiation. Mie scattering can influence a broad range of wavelengths in and near the visible spectrum; Mie's analysis accounts for variations in the size, shape, and composition of such particles. Mie scattering is wavelength-dependent, but not in the simple manner of Rayleigh scattering; it tends to be greatest in the lower atmosphere (0 to 5 km), where larger particles are abundant.

Nonselective scattering is caused by particles that are much larger than the wavelength of the scattered radiation. For radiation in and near the visible spectrum, such particles might be larger water droplets or large particles of airborne dust. "Nonselective" means that scattering is not wavelength-dependent, so we observe it as a whitish or grayish haze; all visible wavelengths are scattered equally.

Effects of Scattering

Scattering causes the atmosphere to have a brightness of its own. In the visible portion of the spectrum, shadows are not jet black (as they would be in the absence of scattering) but are merely dark; we can see objects in shadows because of light redirected by particles in the path of the solar beam. The effects of scattering are also easily observed in vistas of landscapes; the colors and brightnesses of objects are altered as they are positioned at locations more distant from the observer. Landscape artists take advantage of this effect, called *atmospheric perspective,* to create the illusion of depth by painting more distant features in subdued colors and those in the foreground in brighter, more vivid colors.

Scattering has several important consequences for remote sensing (**Figure 2.8**). Because of the wavelength dependency of Rayleigh scattering, radiation in the blue and ultraviolet regions of the spectrum (which are most strongly affected by scattering) is usually not considered useful for remote sensing. Images that record these portions of the spectrum tend to record the brightness of the atmosphere rather than the brightness of the scene itself. For this reason, remote sensing instruments often exclude short-wave radiation (blue and ultraviolet wavelengths) by use of filters or by decreasing the sensitivities of films to these wavelengths. (However, some specialized applications of remote

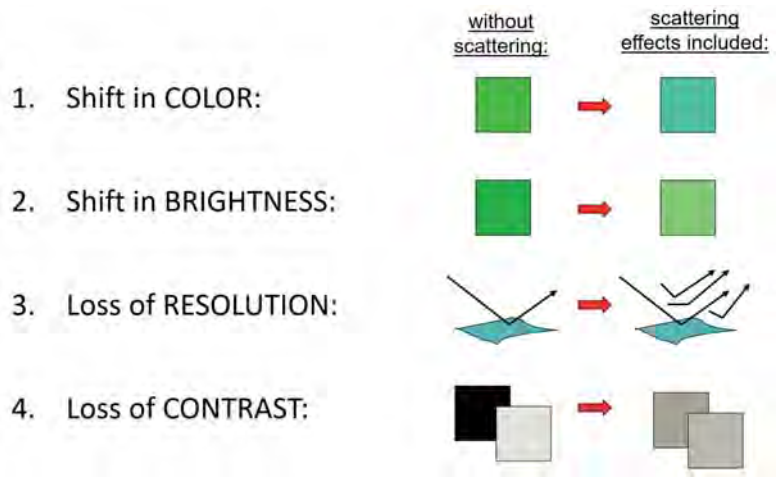


FIGURE 2.8 Effects of atmospheric scattering. Image by Susmita Sen.

sensing, not discussed here, do use ultraviolet radiation.) Scattering also directs energy from outside the sensor's field of view toward the sensor's aperture, thereby decreasing the spatial detail recorded by the sensor. Furthermore, scattering tends to make dark objects appear brighter than they would otherwise be, and bright objects appear darker, thereby decreasing the *contrast* recorded by a sensor (Chapter 4). Because "good" images preserve the range of brightnesses present in a scene, scattering degrades the quality of an image.

Some of these effects are also illustrated in **Figure 2.9**. Observed radiance at the sensor, I, is the sum of I_S, radiance reflected from the Earth's surface, conveying information about surface reflectance; I_O, radiation scattered from the solar beam directly to the sensor without reaching the Earth's surface; and I_D, diffuse radiation, directed first to the ground, then to the atmosphere, before reaching the sensor.

The effects of these components are additive within a given spectral band (Kaufman, 1984):

$$I = I_S + I_O + I_D \qquad \text{(EQ. 2.6)}$$

I_S varies with differing surface materials, topographic slopes and orientation, and angles of illumination and observation. I_O is often assumed to be more or less constant over large areas, although most satellite images represent areas large enough to encompass atmospheric differences sufficient to create variations in I_O. Diffuse radiation, I_D, is expected to be small relative to other factors, but it varies from one land surface type to another, so in practice it would be difficult to estimate. We should note the special case presented by shadows in which $I_S = 0$, because the surface receives no direct solar radiation. However, shadows have their own brightness, derived from I_D, and their own spectral patterns, derived from the influence of local land cover on diffuse radiation. Remote sensing is devoted to the examination of I_S at different wavelengths to derive information about the Earth's surface. **Figure 2.10** illustrates how I_D, I_S, and I_O vary with wavelength for surfaces of differing brightness.

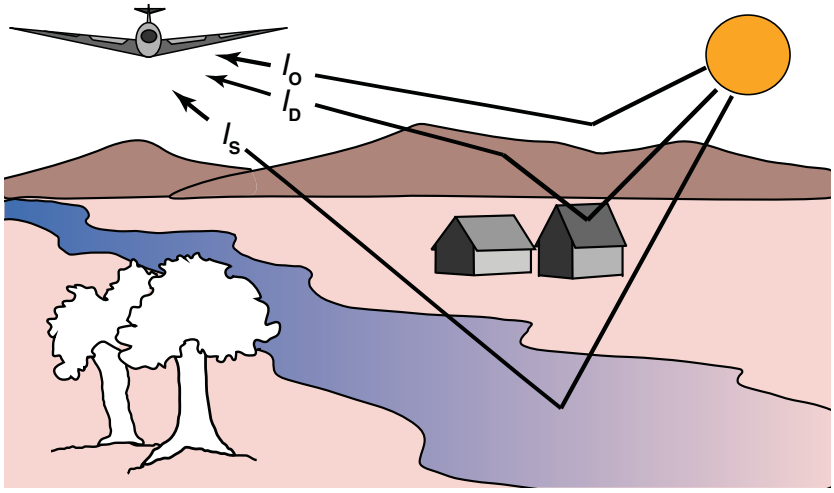


FIGURE 2.9 Principal components of observed brightness. I_S represents radiation reflected from the ground/water surface, I_O is energy scattered by the atmosphere directly to the sensor, and I_D represents diffuse light directed to surface features, then to the atmosphere, before reaching the sensor. This diagram describes the behavior of radiation in and near the visible region of the spectrum. Based on Campbell and Ran (1993). Image by Susmita Sen.

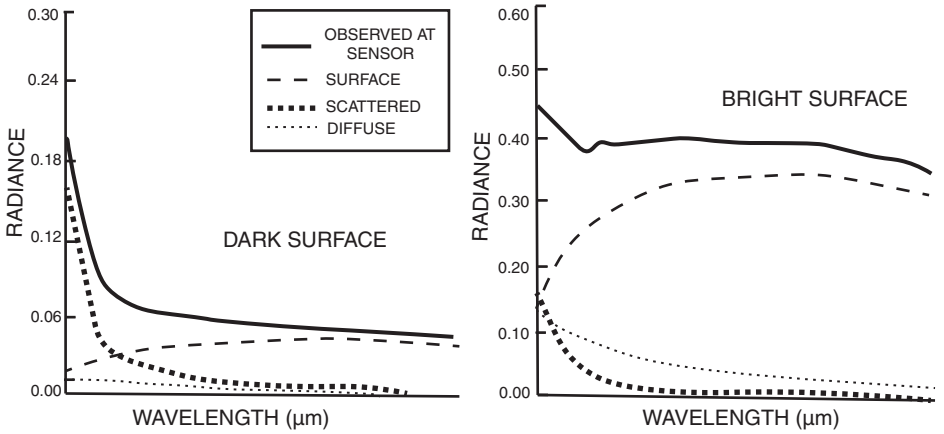


FIGURE 2.10 Changes in reflected, diffuse, scattered, and observed radiation over wavelength for dark (left) and bright (right) surfaces. The diagram shows the magnitude of the components illustrated in **Figure 2.9**. Atmospheric effects constitute a larger proportion of observed brightness for dark objects than for bright objects, especially at short wavelengths. (Here, radiance has been normalized. Note also the differences in scaling of the vertical axes for the two diagrams.) Adapted from Kaufman (1984). Image by Susmita Sen. Used by permission of the author and the Society of Photo-Optical Instrumentation Engineers.

Refraction

Refraction is the bending of light rays at the contact area between two media that transmit light. Familiar examples of refraction are the lenses of cameras or magnifying glasses (Chapter 4), which bend light rays to project or enlarge images, and the apparent displacement of objects submerged in clear water. Refraction also occurs in the atmosphere as light passes through atmospheric layers of varied clarity, humidity, and temperature. These variations influence the density of atmospheric layers, which in turn causes a bending of light rays as they pass from one layer to another. An everyday example is the shimmering appearance on hot summer days of objects viewed in the distance as light passes through hot air near the surface of heated highways, runways, and parking lots. The index of refraction (n) is defined as the ratio between the velocity of light in a vacuum (c) to its velocity in the medium (c_n):

$$n = c/c_n \qquad \text{(EQ. 2.7)}$$

Assuming uniform media, as the light passes into a denser medium, it is deflected toward the *surface normal,* a line perpendicular to the surface at the point at which the light ray enters the denser medium, as represented by the solid line in Figure 2.11. The angle that defines the path of the refracted ray is given by *Snell's law*:

$$n \sin \theta = n' \sin \theta' \qquad \text{(EQ. 2.8)}$$

where n and n' are the indices of refraction of the first and second media, respectively, and θ and θ' are angles measured with respect to the surface normal, as defined in Figure 2.11.

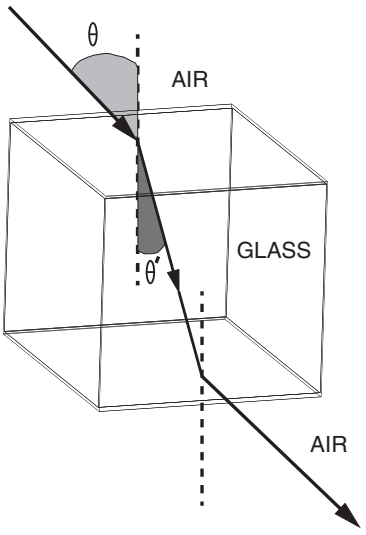


FIGURE 2.11 Refraction. This diagram represents the path of a ray of light as it passes from one medium (air) to another (glass) and again as it passes back to the first. Differing densities of the atmosphere and the glass deflect the path of light rays. Image by Susmita Sen.

Absorption

Absorption of radiation occurs when the atmosphere prevents, or strongly attenuates, transmission of radiation or its energy through the atmosphere. (Energy acquired by the atmosphere is subsequently reradiated at longer wavelengths.) Three gases are responsible for most absorption of solar radiation. Ozone (O_3) is formed by the interaction of high-energy ultraviolet radiation with oxygen molecules (O_2) high in the atmosphere (maximum concentrations of ozone are found at altitudes of about 20–30 km in the stratosphere). Although naturally occurring concentrations of ozone are quite low (perhaps 0.07 parts per million at ground level, 0.1 to 0.2 parts per million in the stratosphere), ozone plays an important role in the Earth's energy balance. Absorption of the high-energy, short-wavelength portions of the ultraviolet spectrum (mainly less than 0.24 μm) prevents transmission of this radiation to the lower atmosphere.

Carbon dioxide (CO_2) also occurs in low concentrations (about 0.03% by volume of a dry atmosphere), mainly in the lower atmosphere. Aside from local variations caused by volcanic eruptions and human activities, the distribution of CO_2 in the lower atmosphere is probably relatively uniform (although human activities that burn fossil fuels have apparently contributed to increases during the past 100 years or so). Carbon dioxide is important in remote sensing because it is effective in absorbing radiation in the mid- and far-infrared regions of the spectrum. Its strongest absorption occurs in the region from about 13 to 17.5 μm in the mid infrared.

Finally, water vapor (H_2O) is commonly present in the lower atmosphere (below about 100 km) in amounts that vary from 0 to about 3% by volume. (Note the distinction between *water vapor,* discussed here, and droplets of *liquid* water, mentioned previously.) From everyday experience we know that the abundance of water vapor varies greatly from time to time and from place to place. Consequently, the role of atmospheric water vapor, unlike those of ozone and carbon dioxide, varies greatly with time and location. It may be almost insignificant in a desert setting or in a dry air mass, but it may be highly significant in humid climates and in moist air masses. Furthermore, water vapor is several times more effective in absorbing radiation than are all other atmospheric gases combined. Two of the most important regions of absorption are in several bands between 5.5 and 7.0 μm and above 27.0 μm; absorption in these regions can exceed 80% if the atmosphere contains appreciable amounts of water vapor.

Atmospheric Windows

Thus, the Earth's atmosphere is by no means completely transparent to electromagnetic radiation because these gases together form important barriers to transmission of electromagnetic radiation through the atmosphere. It selectively transmits energy of certain wavelengths; those wavelengths that are relatively easily transmitted through the atmosphere are referred to as *atmospheric windows* (Figure 2.12). Positions, extents, and effectiveness of atmospheric windows are determined by the absorption spectra of atmospheric gases.

Atmospheric windows are of obvious significance for remote sensing; they define those wavelengths that can be used for forming images. Energy at other wavelengths, not within the windows, is severely attenuated by the atmosphere and therefore cannot be effective for remote sensing. In the far-infrared region, the two most important windows extend from 3.5 to 4.1 μm and from 10.5 to 12.5 μm. The latter is especially important

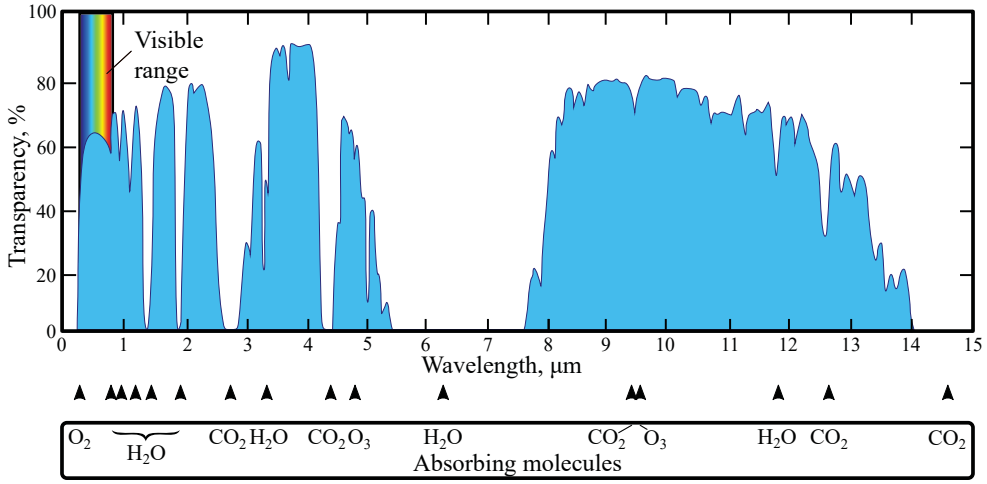


FIGURE 2.12 Atmospheric windows. This schematic representation illustrates the most prominent windows (i.e., transparent regions of the atmosphere that permit radiation to pass through the atmosphere). The shaded regions represent wavelengths regions clear to transmit radiation, especially at the 8 to 14 micrometer region (8–14 μm), and for much of the 0.2–5.5 μm (visible–mid-infrared) region. From U.S. Navy.

because it corresponds approximately to wavelengths of peak emission from the Earth's surface. A few of the most important atmospheric windows are given in Table 2.4; other, smaller windows are not given here but are listed in reference books.

Overview of Energy Interactions in the Atmosphere

Remote sensing is conducted in the context of all the atmospheric processes discussed thus far, so it is useful to summarize some of the most important points by outlining a perspective that integrates much of the preceding material. Figure 2.13 is an idealized diagram of the Earth's energy balance created by NASA from data produced by satellite measurements (NASA Langley Research Center, 2021). The energy balance quanti-

TABLE 2.4 Major Atmospheric Windows

	Ultraviolet and visible	0.30–0.75 μm 0.77–0.91 μm
	Near infrared	1.55–1.75 μm 2.05–2.4 μm
	Thermal infrared	8.0–9.2 μm 10.2–12.4 μm
	Microwave	7.5–11.5 mm 20.0+ mm

Note: Data from Fraser and Curran (1976, p. 35). Used by permission of Addison-Wesley Publishing Co., Inc.

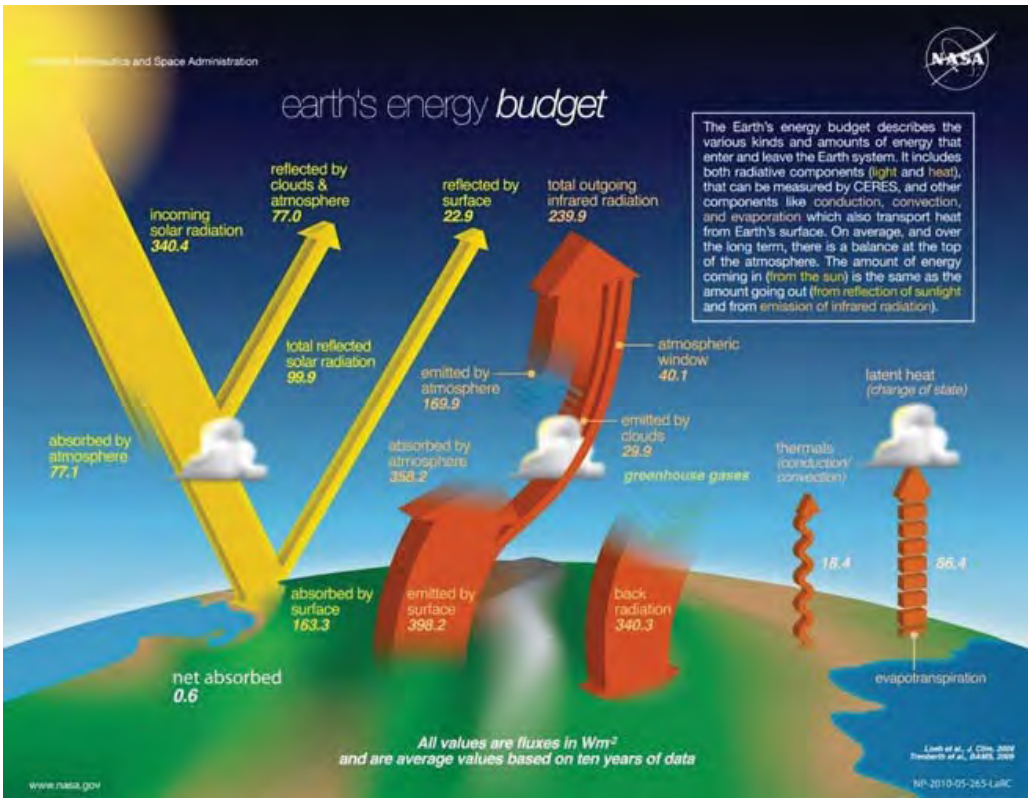


FIGURE 2.13 Earth's energy budget. A depiction of average incoming and outgoing radiation from Earth, created by NASA and many scientists using satellite measurements. From NASA.

fies the way energy moves to the Earth from the Sun and away from the Earth through reflectance and emission. The yellow portion of the diagram represents the behavior of "shortwave" radiation (defined loosely here to include radiation with wavelengths less than 4.0 μm), which is the primary energy that reaches Earth from the Sun. It is true that the Sun emits a broad spectrum of radiation, but the maximum intensity is emitted at approximately 0.5 μm within this region, and little solar radiation at longer wavelengths reaches the ground surface.

In the balance presented in **Figure 2.13**, of the 340 W/m^2 average shortwave radiation that reaches the outer edge of the Earth's atmosphere, roughly half is either absorbed by the atmosphere or reflected from the atmosphere and Earth's surface back into space. For remote sensing in the visible spectrum, it is the portion reflected from the Earth's surface that is of primary interest (see **Figure 2.3**), although knowledge of the quantity scattered is also important. The other half (roughly) of the incoming radiation is ultimately absorbed and then reradiated at the Earth's surface. From Wien's displacement law (Equation 2.5), we know that the Earth, being much cooler than the Sun, must emit radiation at much longer wavelengths than does the Sun. The Sun, at 6,000 K, has its maximum intensity at 0.5 μm (in the green portion of the visible spectrum); the Earth, at 300 K, emits with maximum intensity near 10 μm, in the far-infrared spectrum.

Terrestrial radiation, with wavelengths longer than 10 μm, is represented in Figure 2.13 by the orange arrows. The bulk of emitted energy is absorbed by the atmosphere and either emitted into space as infrared radiation or reradiated back to Earth. A smaller portion of energy (about 40 W/m^2) is emitted directly into space at wavelengths that correspond to atmospheric windows (chiefly 8–13 μm). For meteorology, it is reradiated energy that is the primary interest because it is the source of energy for heating the Earth's atmosphere. For remote sensing, the energy that passes through the atmospheric windows is also of significance, as it is this radiation that conveys information concerning the radiometric properties of features on the Earth's surface.

2.6 INTERACTIONS WITH SURFACES

As electromagnetic energy reaches the Earth's surface, it must be reflected, absorbed, or transmitted. The proportions accounted for by each process depend on the nature of the surface, the wavelength of the energy, and the angle of illumination.

Reflection

Reflection occurs when a ray of light is redirected as it strikes a nontransparent surface. The nature of the reflection depends on sizes of surface irregularities (roughness or smoothness) in relation to the wavelength of the radiation considered. If the surface is smooth relative to wavelength, *specular* reflection occurs (Figure 2.14a). Specular reflection redirects all, or almost all, of the incident radiation in a single direction. For such surfaces, the angle of incidence is equal to the angle of reflection (i.e., in Equation 2.8, the two media are identical, so $n = n'$, and therefore $\theta = \theta'$). For visible radiation, specular reflection can occur with surfaces such as a mirror, smooth metal, or a calm water body.

If a surface is rough relative to wavelength, it acts as a *diffuse,* or *isotropic,* reflector. Energy is scattered more or less equally in all directions. For visible radiation, many natural surfaces might behave as diffuse reflectors, including, for example, uniform grassy surfaces. A perfectly diffuse reflector (known as a *Lambertian surface*) would have equal brightnesses when observed from any angle (Figure 2.14b).

The idealized concept of a perfectly diffuse reflecting surface is derived from the work of Johann H. Lambert (1728–1777), who conducted many experiments designed to

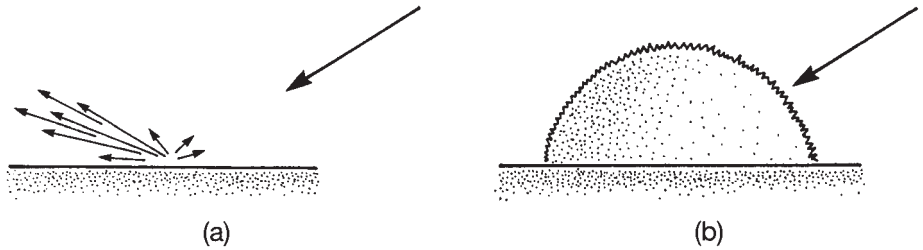


FIGURE 2.14 Specular (a) and diffuse (b) reflection. *Specular reflection* occurs when a smooth surface tends to direct incident radiation in a single direction (mirror-like reflection). *Diffuse reflection* occurs when a rough surface tends to scatter energy more or less equally in all directions.

describe the behavior of light. One of Lambert's laws of illumination states that the perceived brightness (radiance) of a perfectly diffuse surface does not change with the angle of view. This is Lambert's cosine law, which states that the observed brightness (I) of such a surface is proportional to the cosine of the incidence angle (θ), where I is the brightness of the incident radiation as observed at zero incidence:

$$I' = I/\cos\theta \quad \text{(EQ. 2.9)}$$

This relationship is often combined with the equally important inverse square law, which states that observed brightness decreases according to the square of the distance from the observer to the source:

$$I' = (I/D^2)(\cos\theta) \quad \text{(EQ. 2.10)}$$

Both the cosine law and the inverse square law are depicted in Figure 2.15.

Bidirectional Reflectance Distribution Function

Because of its simplicity and directness, the concept of a Lambertian surface is frequently used as an approximation of the optical behavior of objects observed in remote sensing. However, the Lambertian model does not hold precisely for many, if not most, natural surfaces. Actual surfaces exhibit complex patterns of reflection determined by details of surface geometry (e.g., the sizes, shapes, and orientations of plant leaves). Some surfaces may approximate Lambertian behavior at some incidence angles but exhibit clearly non-Lambertian properties at other angles.

Reflection characteristics of a surface are described by the *bidirectional reflectance distribution function* (BRDF). The BRDF is a mathematical description of the optical behavior of a surface with respect to angles of illumination and observation, given that it has been illuminated with a parallel beam of light at a specified azimuth and elevation. (The function is "bidirectional" in the sense that it accounts for both the angle of illumination and the angle of observation.) The BRDF for a Lambertian surface has the shape depicted in Figure 2.14b, with even brightnesses as the surface is observed from any angle. Actual surfaces have more complex behavior. Descriptions of BRDFs for actual, rather than idealized, surfaces permit assessment of the degrees to which they approach the ideals of specular and diffuse surfaces (Figure 2.16).

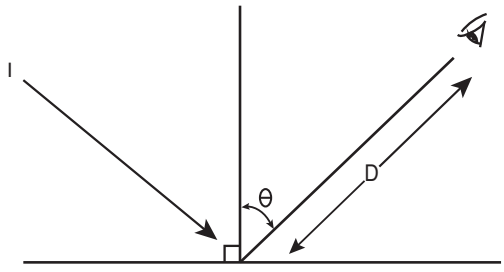


FIGURE 2.15 Inverse square law and Lambert's cosine law. Image by Susmita Sen.

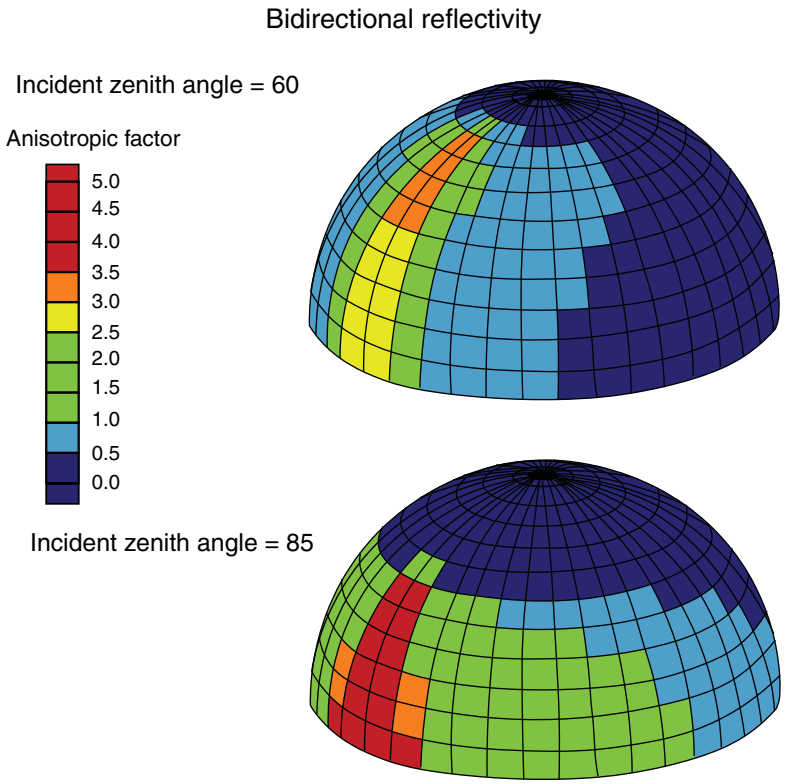


FIGURE 2.16 BRDFs for two surfaces. The varied shading represents the effects of differing angles of illumination as indicated in the annotations (calculated by Pierre Villeneuve). Image by Susmita Sen.

Transmission

Transmission of radiation occurs when radiation passes through a substance without significant attenuation (Figure 2.17). From a given thickness, or depth, of a substance, the ability of a medium to transmit energy is measured as the transmittance (t):

$$t = \frac{\text{Transmitted radiation}}{\text{Incident radiation}} \qquad \text{(EQ. 2.11)}$$

In the field of remote sensing, the transmittance of films and filters is often important. With respect to naturally occurring materials, we often think only of water bodies as capable of transmitting significant amounts of radiation. However, the transmittance of many materials varies greatly with wavelengths, so our direct observations in the visible spectrum do not transfer to other parts of the spectrum. For example, plant leaves are generally opaque to visible radiation but transmit significant amounts of radiation in the infrared.

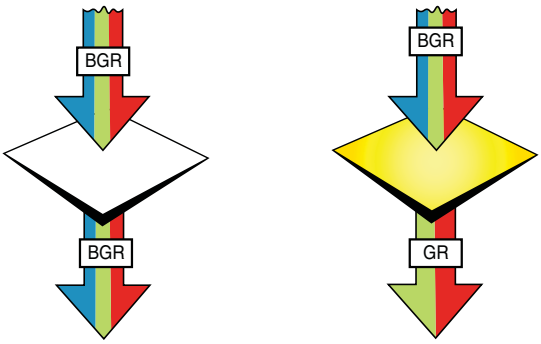


FIGURE 2.17 Transmission. Incident radiation passes through an object without significant attenuation (left), or selectivity transmitted (right). The object on the right would act as a yellow ("minus blue") filter, as it would transmit visible radiation, except for blue light, which is blocked by the filter. Image by Susmita Sen.

Fluorescence

Fluorescence occurs when an object illuminated with radiation of one wavelength emits radiation at a different wavelength. The most familiar examples are some sulfide minerals, which emit visible radiation when illuminated with ultraviolet radiation. Other objects also fluoresce, although observation of fluorescence requires very accurate and detailed measurements that are not now routinely available for most applications. Chlorophyll fluorescence is an important phenomenon to examine with remote sensing for vegetation, as it is an indication of photosynthetic activity and has been shown through spectroscopy to reveal differences between healthy and stressed leaves.

Polarization

The *polarization* of electromagnetic radiation denotes the orientation of the oscillations within the electric field of electromagnetic energy (**Figure 2.18**). A light wave's electric field (traveling in a vacuum) is typically oriented perpendicular to the wave's direction of travel (i.e., the energy propagates as a *transverse* wave); the field may have a preferred orientation, or it may rotate as the wave travels. Although polarization of electromagnetic radiation is too complex for a full discussion here, it is possible to introduce some of the basics and highlight its significance.

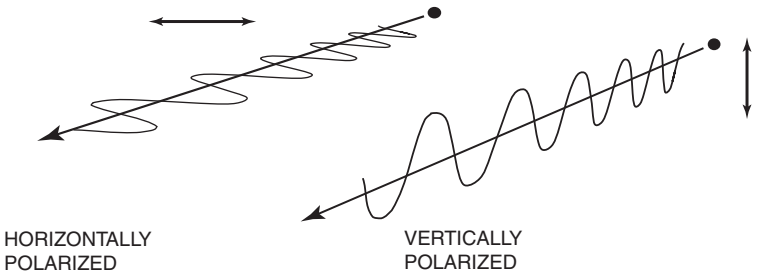


FIGURE 2.18 Schematic representation of horizontally and vertically polarized radiation. The arrows signify the orientations of the electric fields. Image by Susmita Sen.

An everyday example of the effect of polarized light is offered by polarizing sunglasses, which are specifically designed to reduce glare. Typically, sunlight within the atmosphere has a mixture of polarizations; when it illuminates surfaces at steep angles (i.e., when the Sun is high in the sky), the reflected radiation tends to also have a mixture of polarizations. However, when the Sun illuminates a surface at low angles (i.e., the Sun is near the horizon), many surfaces tend to preferentially reflect the horizontally polarized component of solar radiation. The polarizing sunglasses are manufactured with lenses that include molecules that can preferentially absorb the horizontally polarized bright radiation, thereby reducing glare. Polarization has broader significance in the practice of remote sensing. Within the atmosphere, polarization of light is related to the nature and abundance of atmospheric aerosols and atmospheric clarity. Chapter 8 introduces the use of polarized radiation in the design of active microwave sensors.

Reflectance

For many applications of remote sensing, the brightness of a surface is best represented not as irradiance but rather as *reflectance*. Reflectance (R_{rs}) is expressed as the relative brightness of a surface as measured for a specific wavelength interval:

$$\text{Reflectance} = \frac{\text{Observed brightness}}{\text{Irradiance}} \qquad \text{(EQ. 2.12)}$$

As a ratio, it is a dimensionless number (between 0 and 1), but it is commonly expressed as a percentage. In the usual practice of remote sensing, R_{rs} is not directly measurable because normally we can observe only the observed brightness and must estimate irradiance. Strategies devised for estimation of reflectance are discussed in Chapter 11.

Spectral Properties of Objects

Remote sensing consists of the study of radiation emitted and reflected from features at the Earth's surface. In the instance of emitted (far-infrared) radiation, the object itself is the immediate source of radiation. For reflected radiation, the source may be the Sun, the atmosphere (by means of scattering of solar radiation), or human-made radiation (chiefly imaging radars).

A fundamental premise in remote sensing is that we can learn about objects and features on the Earth's surface by studying the radiation reflected and/or emitted by these features. Using cameras and other remote sensing instruments, we can observe the brightnesses of objects over a range of wavelengths, so that there are numerous points of comparison between the brightnesses of separate objects. A set of such observations or measurements constitutes a spectral response pattern, sometimes called the *spectral signature* of an object (**Figure 2.19**). In the ideal, detailed knowledge of a spectral response pattern might permit identification of features of interest, such as separate kinds of crops, forests, or minerals. This idea has been expressed as follows:

> Everything in nature has its own unique distribution of reflected, emitted, and absorbed radiation. These spectral characteristics can—if ingeniously exploited—be used to distinguish one thing from another or to obtain information about shape, size, and other physical and chemical properties. (Parker and Wolff, 1965, p. 21)

This statement expresses the fundamental concept of the spectral signature, the notion that features display unique spectral responses that permit clear identification, from spectral information alone, of individual crops, soils, and so on from remotely sensed images. In practice, it is now recognized that spectra of features change both over time (e.g., as a cornfield grows during a season) and over distance (e.g., as proportions of specific tree species in a forest change from place to place).

Nonetheless, study of the spectral properties of objects forms an important part of remote sensing. Some research has been focused on examination of the spectral properties of different classes of features. Thus, although it may be difficult to define unique signatures for specific kinds of vegetation, we can recognize distinctive spectral patterns for vegetated and nonvegetated areas and for certain classes of vegetation, and we can sometimes detect the existence of diseased or stressed vegetation. In other instances, we may be able to define spectral patterns that are useful within restricted geographic and temporal limits as a means of studying the distributions of certain plant and soil characteristics. Chapter 14 describes how detailed spectral measurements permit application of some aspects of the concept of the spectral signature.

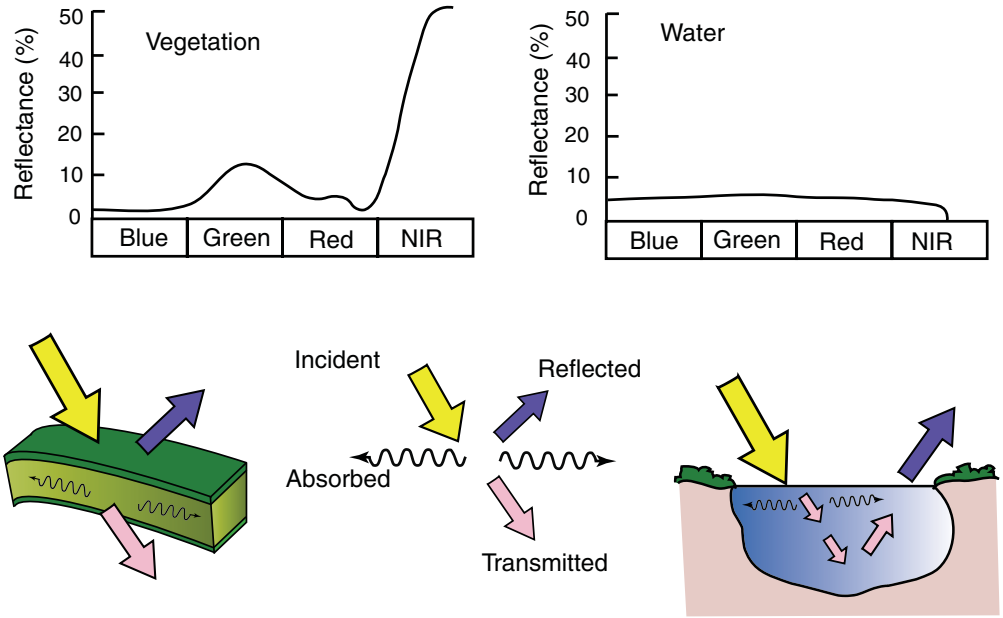


FIGURE 2.19 Spectral response curves for vegetation and water. These curves represent contrasting relationships between brightness (vertical axis) and wavelength (horizontal axis) for two common surfaces: living vegetation and open water. The sketches represent schematic views of a cross section of a living leaf (left) and a pond with clear, calm water (right). The large arrows represent incident radiation from the Sun; the small lateral arrows represent *absorbed radiation*; the downward arrows represent *transmitted energy*; and the upward arrows represent energy directed upward to the sensor (known as *reflectance*) that form the spectral response patterns illustrated at the top of the diagram. Fuller discussions of both topics are presented in later chapters. Image by Susmita Sen.

2.7 SUMMARY

Remote sensing typically takes one of three basic forms, depending on the wavelengths of energy detected and the purposes of the study. In the simplest form, one records the reflection of solar radiation from the Earth's surface (Figure 2.20). This is the kind of remote sensing that is most nearly similar to everyday experience. For example, film in a camera records radiation from the Sun after it is reflected from the objects of interest, regardless of whether one uses a simple handheld camera to photograph a family scene or a complex aerial camera to photograph a large area of the Earth's surface. This form of remote sensing mainly uses energy in the visible and near-infrared portions of the spectrum. Key variables include atmospheric clarity, spectral properties of objects, angle and intensity of the solar beam, choices of films and filters, and others explained in Chapter 4.

A second strategy for remote sensing is to record radiation *emitted* (rather than *reflected*) from the Earth's surface. Because emitted energy is strongest in the far-infrared spectrum, this kind of remote sensing requires special instruments designed to record these wavelengths. (There is no direct analog to everyday experience for this kind of remote sensing.) Emitted energy from the Earth's surface is mainly derived from shortwave energy from the Sun that has been absorbed, then reradiated at longer wavelengths (Figure 2.21).

Emitted radiation from the Earth's surface reveals information concerning the thermal properties of materials, which can be interpreted to suggest patterns of moisture, vegetation, surface materials, and human-made structures. Other sources of emitted radiation (of secondary significance here, but often of primary significance elsewhere) include geothermal energy and heat from steam pipes, power plants, buildings, and forest fires. This example also represents "passive" remote sensing because it employs instruments designed to sense energy emitted by the Earth, not energy generated by a sensor.

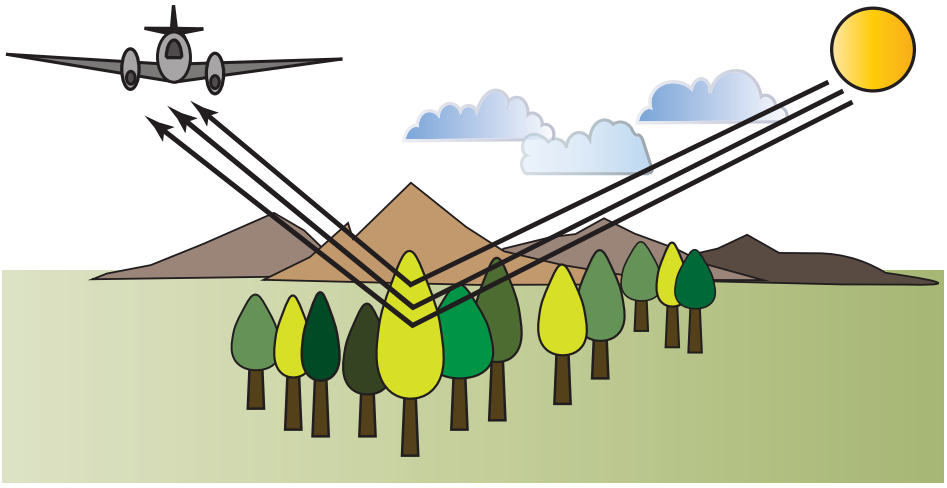


FIGURE 2.20 Remote sensing using reflected solar radiation. The sensor detects solar radiation that has been reflected from features at the Earth's surface. Image by Susmita Sen.

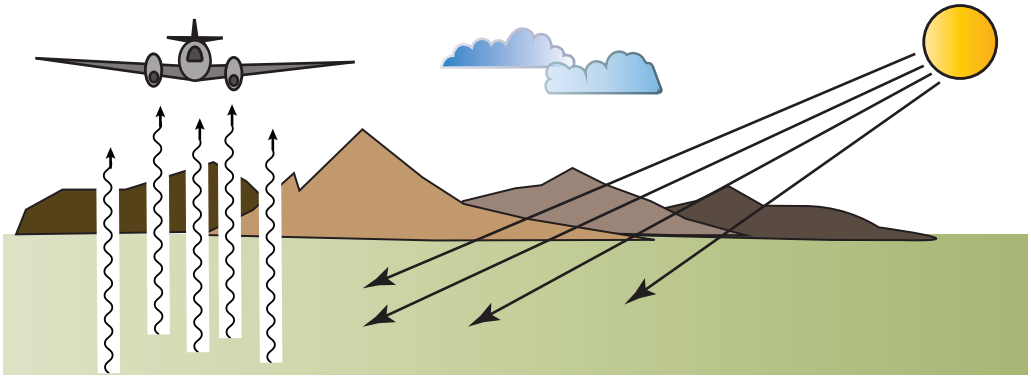


FIGURE 2.21 Remote sensing using emitted terrestrial radiation. The sensor records solar radiation that has been absorbed at the Earth's surface and is then reemitted as thermal infrared radiation. (See also **Figure 2.13**.) Image by Susmita Sen.

Finally, sensors belonging to a third class of remote sensing instruments generate their own energy, then record the reflection of that energy from the Earth's surface (Figure 2.22). These are "active" sensors—"active" in the sense that they provide their own energy, so they are independent of solar and terrestrial radiation. As an everyday analogy, a camera with a flash attachment can be considered to be an active sensor. In practice, active sensors are best represented by imaging radars and lidars (Chapters 8 and 9), which transmit energy toward the Earth's surface from an aircraft or satellite, then receive the reflected energy to form an image. Because they sense energy provided directly by the sensor itself, such instruments have the capability to operate at night and during cloudy weather.

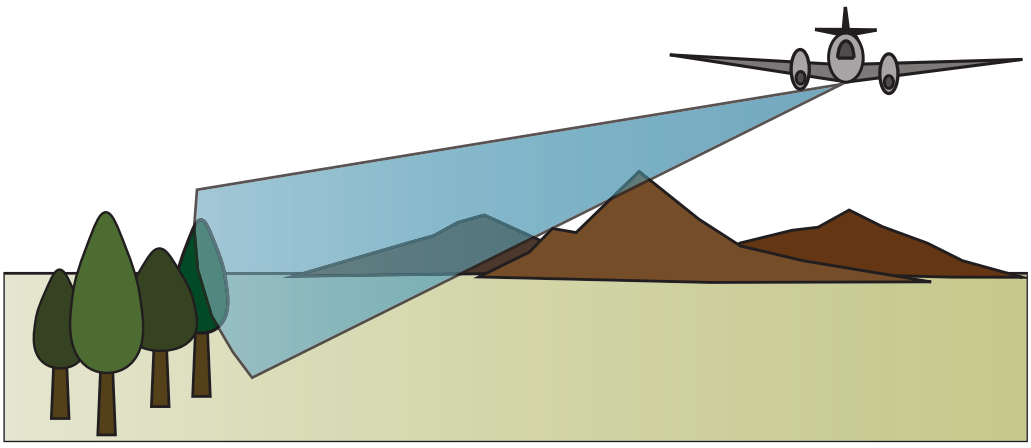


FIGURE 2.22 Active remote sensing. The airborne radar system illuminates terrain with microwave energy and then receives and records the reflected energy as it has been reflected by the Earth's surface. Image by Susmita Sen.

SOME TEACHING AND LEARNING RESOURCES

- Quantum Mechanics
  www.youtube.com/watch?v=l_t8dn4c6_g
- XNA Atmospheric Scattering
  www.youtube.com/watch?v=W0ocgQd_huU
- Atmospheric Rayleigh Scattering in 3D with Unity 3D
  www.youtube.com/watch?v=PNBnfqUycto
- How a Sunset Works
  www.youtube.com/watch?v=BdNQ1xB34QI&feature=related
- Tour of the EMS 04—Infrared Waves
  www.youtube.com/watch?v=i8caGm9Fmh0
- IR Reflection
  www.youtube.com/watch?v=h2n9WQCH1ds&feature=related
- Emissivity Makes a Temperature Difference: Blackbody Calibrator
  www.youtube.com/watch?v=JElKE-ADXr8&feature=related

REVIEW QUESTIONS

1. Using books provided by your instructor or available through your library, examine reproductions of landscape paintings to identify artistic use of atmospheric perspective. Perhaps some of your own photographs of landscapes illustrate the optical effects of atmospheric haze.
2. Some streetlights are deliberately manufactured to provide illumination with a reddish color. From material presented in this chapter, can you suggest why?
3. Although this chapter has largely dismissed ultraviolet radiation as an important aspect of remote sensing, there may well be instances in which it might be effective, despite the problems associated with its use. Under what conditions might it prove practical to use ultraviolet radiation for remote sensing?
4. The human visual system is most nearly similar to which model of remote sensing?
5. Can you identify analogs from the animal kingdom for each of the models for remote sensing discussed in Section 2.7?
6. Examine **Figure 2.13**, which shows the radiation balance of the Earth's atmosphere. Explain how it can be that there are more units of radiation emitted from the ground than enter through the incoming solar radiation.
7. Examine **Figure 2.13** again. Discuss how the values in this figure might change in different environments, including (a) desert, (b) the Arctic, and (c) an equatorial climate. How might these differences influence our ability to conduct remote sensing in each region?
8. Spectral signatures can be illustrated using values indicating the brightness in several spectral regions.

	UV	Blue	Green	Red	NIR
Forest	28	29	36	27	56
Water	22	23	19	13	8
Corn	53	58	59	60	71
Pasture	40	39	42	32	62

Assume for now that these signatures are influenced by effects of the atmosphere. Can all categories be reliably separated, based on these spectral values? Which bands are most useful for distinguishing between these classes?

9. Describe ideal atmospheric conditions for remote sensing.

REFERENCES

Bohren, C. F. 1987. *Clouds in a Glass of Beer: Simple Experiments in Atmospheric Physics*. New York: John Wiley, 195 pp.

Campbell, J. B., and L. Ran. 1993. CHROM: A C Program to Evaluate the Application of the Dark Object Subtraction Technique to Digital Remote Sensing Data. *Computers and Geosciences,* Vol. 19, pp. 1475–1499.

Chahine, M. T. 1983. Interaction Mechanisms within the Atmosphere. Chapter 5 in *Manual of Remote Sensing* (R. N. Colwell, ed.). Falls Church, VA: American Society of Photogrammetry, pp. 165–230.

Chameides, W. L., and D. D. Davis. 1982. Chemistry in the Troposphere. *Chemical and Engineering News,* Vol. 60, pp. 39–52.

Clark, R. N., and T. L. Roush. 1984. Reflectance Spectroscopy: Quantitative Analysis Techniques for Remote Sensing Applications. *Journal of Geophysical Research,* Vol. 89, pp. 6329–6340.

Estes, J. E. 1978. The Electromagnetic Spectrum and Its Use in Remote Sensing. Chapter 2 in *Introduction to Remote Sensing of Environment* (B. F. Richason, ed.). Dubuque, IA: Kendall-Hunt, pp. 15–39.

Fraser, R. S., and R. J. Curran. 1976. Effects of the Atmosphere on Remote Sensing. Chapter 2 in *Remote Sensing of Environment* (C. C. Lintz and D. S. Simonett, eds.). Reading, MA: Addison-Wesley, pp. 34–84.

Goetz, A. F. H., J. B. Wellman, and W. L. Barnes. 1985. Optical Remote Sensing of the Earth. *Proceedings of the IEEE,* Vol. 73, pp. 950–969.

Hariharan, T. A. 1969. Polarization of Reflected Solar Radiation over Land, Sea and Cloud Surfaces. *Pure and Applied Physics,* Vol. 77, pp. 151–159.

Kaufman, Y. J. 1984. Atmospheric Effects on Remote Sensing of Surface Reflectance. Special Issue: *Remote Sensing* (P. N. Slater, ed.). *Proceedings, SPIE,* Vol. 475, pp. 20–33.

Kaufman, Y. J. 1989. The Atmospheric Effect on Remote Sensing and Its Correction. Chapter 9 in *Theory and Applications of Optical Remote Sensing* (G. Asrar, ed.). New York: John Wiley, pp. 336–428.

Lynch, D. K., and W. Livingston. 1995. *Color and Light in Nature*. New York: Cambridge University Press, 254 pp.

Minnaert, M. 1954. *The Nature of Light and Color* (revision by H. M. Kremer-Priest; translation by K. E. Brian Jay). New York: Dover, 362 pp.

Mobley, C. D. 1999. Estimation of the Remote-Sensing Reflectance from Above-Surface Measurements. *Applied Optics,* Vol. 38, pp. 7442–7455.

NASA Langley Research Center, Science Directorate Education and Public Outreach. 2021, August 13. *The NASA Earth's Energy Budget*. Available at *https://science-edu.larc.nasa.gov/energy_budget*.

Parker, D. C., and M. F. Wolff. 1965. Remote Sensing. *International Science and Technology, Vol*. 43, pp. 20–31.

Rinker, J. N. 1994. ISSSR Tutorial I: Introduction to Remote Sensing. In *Proceedings of the International Symposium on Spectral Sensing Research '94*. Alexandria, VA: U.S. Army Topographic Engineering Center, pp. 5–43.

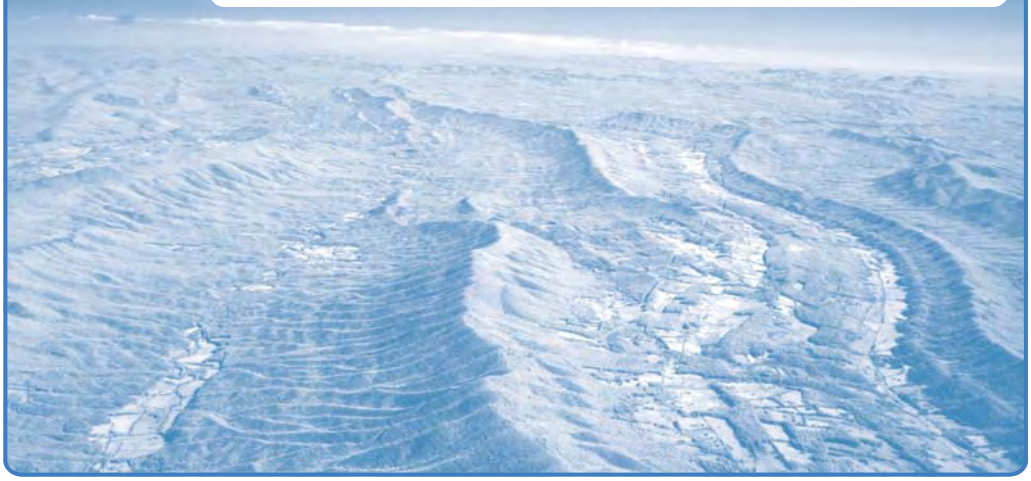
Schaepman-Strub, G., M. E. Schaepaman, J. Martonchik, T. Painter, and S. Dangel. 2009. Radiometry and Reflectance: From Terminology Concepts to Measured Quantities. Chapter 15 in *Sage Handbook of Remote Sensing* (T. M. Warner, M. D. Nellis, and G. M. Foody, Eds.). London: SAGE, pp. 215–228.

Slater, P. N. 1980. *Remote Sensing: Optics and Optical Systems*. Reading, MA: Addison-Wesley, 575 pp.

Stimson, A. 1974. *Photometry and Radiometry for Engineers*. New York: John Wiley, 446 pp.

Turner, R. E., W. A. Malila, and R. F. Nalepka. 1971. Importance of Atmospheric Scattering in Remote Sensing. In *Proceedings of the 7th International Symposium on Remote Sensing of Environment*. Ann Arbor, MI: Willow Run Laboratories, University of Michigan, pp. 1651–1697.

3 Remote Sensing Platforms



MAJOR TOPICS TO UNDERSTAND

- Platforms
- Fixed-Wing Aircraft
- Helicopters
- Satellite Systems
- Unmanned Aerial Systems
- Tethered Balloons
- Mobile Collection of Field Data

3.1 INTRODUCTION

This chapter introduces an important dimension to the practice of remote sensing—*platforms*—remote sensing's generic name for the aerial (sometimes orbital) vehicles used to place sensors in appropriate positions to acquire imagery to satisfy a specific need. These vehicles cover a wide range of technologies and applications. Remote sensing systems acquire information about the Earth's surface and landscape features as observed from

aerial perspectives at varied altitudes, without contact with objects. Remote sensing platforms presented here illustrate capabilities such as ecology, land survey, geology, geography, cities, agriculture, and forest fires. ***Active remote sensing*** emits energy from surfaces or sensors that scan features detecting and measuring radiation reflected from the Earth's surface. In contrast, ***passive remote sensing*** collects radiation emitted or reflected from a landscape surface. Aside from sunlight, the most common sources of radiation are features such as infrared radiation, photography, and other features characterized by passive sensors.

3.2 PLATFORMS

Remote sensing *platforms* refer to the various vehicles that carry sensor systems that collect electromagnetic radiation we use to form images of the Earth's surface. They provide the overhead perspective for the map-like views of the Earth's surface we see in remotely sensed imagery. Platforms include vehicles such as fixed-wing aircraft, helicopters, unmanned aerial systems (UAS), satellites, and balloons—vehicles that provide the means to control the altitude, orientation, trajectory, and timing necessary to acquire useful imagery. Although the choice of platform may seem to be a secondary concern for the practice of remote sensing, even cursory consideration reveals that it is central to the practice of remote sensing. Often, the choice of sensor for a remotely sensed image is closely connected to the choice of platform; the unavailability of a suitable platform may change the entire design strategy for a remote sensing project. Thus, matching a sensor to an appropriate platform forms a key aspect of any remote sensing mission.

Platforms vary greatly with respect to size, range, maneuverability, expense, and altitudinal range. Here, we briefly consider the most commonly used platforms, including fixed-wing aircraft, rotorcraft (chiefly helicopters), unmanned aerial systems (UASs), satellites, and tethered balloons. Each provides its own distinct advantages and disadvantages and has its own niche in the practice of remote sensing.

3.3 FIXED-WING AIRCRAFT

Fixed-wing aircraft are what we usually picture when we think of an airplane—a cylindrical fuselage with rigid wings, powered by propeller, jet engines, or turbo-props. Fixed-wing aircraft have many distinctive capabilities for remote sensing, including long flight range, precise navigation, ability to accommodate varied sensor systems, and the ability to lift heavy loads to support a variety of alternative remote sensing missions. Aerial survey firms will carefully select aircraft models to meet specific operational requirements for their business models with respect to supporting sensors, flight range, sensors, and fuel economy. Although some aircraft were designed specifically for aerial survey and observation, current systems are based mainly on specialized models of general-purpose aircraft tailored for the aerial survey mission.

Fixed-wing aircraft have disadvantages of high costs for purchase, maintenance, and operation. They can incur costs for weather delays and for expenses incurred for transit from one project site to another. In addition, they require supporting staff for maintenance, flight planning, IT support, navigation, and sensor operation and maintenance.

Small Aircraft

Figure 3.1 illustrates the Cessna 172, as an example of the smaller fixed-wing aerial platform—a single-engine aircraft, small in size, capacity, and range relative to other systems. It features four seats and a high-wing design (favoring oblique photography, as it provides a largely unobstructed lateral view). Because of their widespread availability, such aircraft are often used for handheld photography, news photography or videos, and aerial photography for real estate advertising. The pod illustrated in Figure 3.2 modifies the aircraft to provide a capability for vertical aerial photography, and for use of other sensors.

FIGURE 3.1 The Cessna 172, an example of a small aircraft with capabilities for aerial photography. It is one of the most widely used general aviation aircraft, employed also as a trainer by the U.S. Air Force, the U.S. Army, and federal agencies for aerial search and patrol. From NOAA.

FIGURE 3.2 The camera pod, visible as the dark capsule positioned below the fuselage just below the wing strut, is approved for use on the Cessna 172 and other aircraft. This detachable pod is suitable for small- and medium-format camera systems, as well as compact lidar systems and thermal sensors. From Airborne Scientific. Used by permission.

Midsized Aircraft

Larger aircraft, perhaps turbo-prop or jet-powered, offer an increased capability for payload and for crew and passengers. They have higher purchase and operating costs but provide the advantages of longer flight range and larger payload capacity, and they are available in models that support aerial survey and remote sensing equipment. Figures 3.3 and 3.4 illustrate the DeHavilland Twin Otter (DHC-6) and the Beechcraft King Air 90, both larger aircraft used by the U.S. National Oceanic and Atmospheric Administration (NOAA) to support their remote sensing and aerial survey missions.

Aerial survey firms use midsized aircraft for a variety of tasks, staffed by crews with expertise for navigation, extended flights, and instrument operations. Such aircraft and their instrument payloads are supported by additional staff for ground support, including IT services, maintenance, flight planning, and image analysis and interpretation. Businesses must operate not only to acquire imagery locally, but also to cover costs for transit of aircraft and crew from one project site to another, and for weather delays. Collection of aerial imagery is usually guided by a statement of work (SOW), which is basically a contract between the customer and the aerial survey firm, with precise specifications as to the nature of the imagery, its quality, date, time of day, and season (leaf-off/leaf-on). Aerial survey firms acquire imagery to support a wide variety of projects, including highway and building construction, urban planning, coastal management, and agricultural management.

Large Fixed-Wing Aircraft

Less frequently, much larger aircraft are configured for remote sensing missions, which are often those requiring the use of experimental sensors, larger crews, multisensor systems, longer missions, or operations at higher altitudes. An example is the Douglas DC-8 design, which is a four-engine, long-range, narrow-body jet airliner that was manufactured from 1958 to 1972, initially used in passenger and cargo service. Later, many of the DC-8s were reconfigured for special-purpose missions, such as NASA's remote sensing programs. The DC-8 can operate at an altitude of 41,000 ft, with a range of 5,000 mi.

FIGURE 3.3 The DeHavilland Twin Otter (DHC-6) midsize aircraft, used by NOAA for aerial survey missions because of its maneuverability, versatility, and ability to accommodate varied remote sensing instruments. From NOAA, Northeast Fisheries Science Center. Photograph by Christin Khan.

FIGURE 3.4 The Beechcraft King Air 90 midsize aircraft, available in a version (model 200T) designed specifically to support remote sensing and aerial observation missions. Modifications support imaging radars and provide dome-shaped windows in the rear fuselage that allow vertical observation directly below the aircraft path. From NOAA.

Figures 3.5 and 3.6 show the AIRSAR DC-8–72, with *synthetic aperture radar* (SAR) antennae installed in the aft fuselage, with a special port to accommodate radar antennae. (*Note:* Specifics of radar systems are presented in Chapter 8.)

One example use of the DC-8 is the NASA AIRSAR, which has been employed since 1988 as the NASA DC-8 Airborne Laboratory Program (NASA/Ames Research Center, CA, *http://airsar.jpl.nasa.gov/documents/genairsar/overview.html*) to support airborne science research. AIRSAR is configured to accommodate imaging radars (specifically SARs, described in Chapter 8), with the flexibility to operate a wide variety of other instruments and sensors. This flexibility provides a platform for a wide range of experiments, supporting many disciplines for NASA, federal, state, academic, and foreign investigators. As an example, the AIRSAR serves as a testbed for NASA's evaluation of new radar technologies, processing techniques, and exploration of new application missions.

3.4 HELICOPTERS

Helicopters are rotary-wing aircraft with rotor blades functioning as wings to provide lift. They are the best-known examples of the class of aerial vehicles labeled as *rotorcraft* (aircraft characterized by horizontally oriented blades). Helicopters provide specific capabilities to support remote sensing activities, including a unique ability to hover, fly laterally, and operate in confined surroundings, and they have the maneuverability to navigate complex flight plans. They are expensive to operate and maintain, and gener-

FIGURE 3.5 NASA's DC-8 large aircraft, configured mainly for testing and developmental missions. The port for the imaging radar antenna is visible at the aft fuselage, just aft of the left wing. From NASA.

ally have shorter ranges than fixed-wing aircraft. Helicopter designs vary greatly in size, range, and cargo capacity.

Helicopters can accommodate a wide range of remote sensing instruments. Larger models can be flown with cargo doors open to allow unobstructed visibility to the sides, and they can also accommodate bulky equipment and instruments within the cabin. Smaller models can be transported by trailer to project sites, and for longer distances, they can be transported in cargo aircraft. They are well suited for a wide variety of remote sensing missions because of their versatility and maneuverability, ability to oper-

FIGURE 3.6 A closer view of the radar antenna port in the DC-8 fuselage, showing antenna panels that transmit microwave pulses from the imaging radar system. From NASA.

ate in confined circumstances, such as forested regions, uneven terrain, or other settings unsuitable for fixed-wing aircraft. Disadvantages include high purchase prices and operating expenses, and shorter ranges of operation, which may restrict helicopters to smaller, specialized, remote sensing projects. Helicopters can operate at low flight speeds relative to other aircraft, so they can be effective for observing sites from multiple perspectives and for tracking wildlife.

Aerial photography and videography (provided that they are protected from vibration associated with helicopters) are effectively supported by helicopters, often for TV news, documentary films, and scientific studies. Helicopters are often used when maneuverability is important or low flight speeds are advantageous. Examples include their use in rugged terrain for wildlife mapping/monitoring and environmental disasters, or in airspace where fixed-wing flight operations are impractical (see **Figures 3.7 and 3.8**).

3.5 SATELLITE SYSTEMS

Today artificial Earth satellites (i.e., human-built) are so commonplace that it can be difficult to recognize how significant they are for observing the Earth, acquiring environmental data, and functioning in communications and data systems. Here, we will refer specifically to artificial Earth satellites designed for Earth observation.

With the launch of the first U.S. Landsat satellite in 1972, civil remote sensing began to use satellites as platforms to observe the Earth's surface. (At that time, previous satellite systems had been used for specialized, unpublicized, strategic reconnaissance photography and for meteorological observations.) Land observation satellite systems are

FIGURE 3.7 The Bell 412 helicopter, often identified as the NOAA Ocean Explorer. Designed for transport, aerial survey, and other capabilities, it is often used for remote sensing applications. The original version of the Bell 412 was designed in the 1970s, with subsequent versions developed as late as 2013. From NOAA.

FIGURE 3.8 The Huey rescue helicopter (Bell UH-1, nicknamed the "Huey") in flight. The Huey family of helicopters have both military and civil applications, including rescue, reconnaissance, medical evacuation, crop dusting, cargo lifting, and aerial firefighting. Photograph by Alan Radecki.

unique in their focus on land observation—systematic collection of imagery that records site-specific data about the Earth's land areas, including, for example, agriculture, water bodies, urban systems, forests, and rangelands.

From this beginning, today numerous corporations and national governments operate satellite remote sensing systems that are specifically designed for observation of the Earth's surface. Campbell and Salomonson (2010) list nations that are now operating land observation satellites that collect data in the optical region, and others that operate their own radar imaging satellites. Such satellite platforms offer unique capabilities, including an ability to provide synoptic perspective (observation of large areas in a single image), fine detail, and systematic, repetitive coverage. These capabilities are well suited to maintaining an up-to-date, worldwide, cartographic infrastructure, and for monitoring changes in the many broad-scale environmental issues that the world faces today.

Satellite systems contribute important capabilities to the overall remote sensing mission by doing the following:

- Providing systematic collection on established, repetitive, schedules
- Using consistent data formats
- Depending on specifics, providing a capability for worldwide coverage
- Archiving data to provide resources for change detection
- Offering a revisit capability (ability to observe the same region on a repetitive basis)
- Providing a broad customer base for remote sensing
- Allowing the routine image acquisition of remote, hazardous, or inaccessible sites

Remote sensing satellite systems also have shortcomings, including:

- Substantial costs for development, manufacturing, and operations
- Requirement of long lead times to develop the systems
- Difficulties in recovering from system or equipment failures
- Difficulty in responding to unexpected or fast-moving events for which imagery may be desired because of the systematic coverage cycles

The Basic Satellite Platform

A remote sensing satellite is a specialized version of a standard instruments platform that provides support services for the remote sensing payload. The satellite's overall frame, often known as the *bus,* provides a structure to hold and organize a collection of devices necessary to support the satellite mission. The *power supply* is composed of solar panels that accumulate energy on the sunlit side of the orbit, and batteries that retain energy for use as the satellite passes through the Earth's shadow. A *thermal control* subsystem balances the effects of solar heating and the heat generated from the satellite. A *computer subsystem* coordinates the satellite's varied systems and monitors system status. An *attitude control subsystem* tracks the course of the satellite and ensures that it is aimed correctly. Finally, a *propulsion system* adjusts the satellite's orientation and makes minor course corrections (once in orbit, the satellite does not require propulsion to maintain its motion, but it does require energy to periodically adjust its orbital track). These components, though differing in specifics, form a standard inventory of key components of the satellite bus.

A remote sensing satellite system acquires its distinctive identity through its design, sensors, and specific orbit, which together define its unique capabilities. Another unique feature of a satellite remote sensing satellite system is its *ground component,* which provides a specialized capability for monitoring and guiding satellite operations and for archiving, processing, and distributing specialized products tailored to support specific scientific and business applications.

Satellite Orbits

A key characteristic of any remote sensing satellite system is its *orbit*—its repetitive trajectory in space. Each satellite orbit can be uniquely described by six properties, known as *Classical Orbital Elements* (COEs). While all six of these elements are important to spacecraft design and operation, for our present purposes, two orbital parameters are sufficient: (1) *inclination* and (2) *altitude*.

An important characteristic of a satellite orbit is *inclination,* which specifies an orbit's angle with respect to the equator (**Figure 3.9**). For example, an inclination of 0 degrees would specify a perfectly equatorial orbit, while an orbit with an inclination of 90 degrees would pass directly over the North and South Poles. Orbits with inclinations greater than 90 degrees are said to be *retrograde* because the satellite orbits counter to the Earth's rotation, while orbits with inclinations between 0 and 90 degrees are said to be *prograde* because they travel in the direction of the Earth's rotation.

Each satellite orbit is designed for an inclination that will best support its mission. For example, the Tropical Rainfall Measuring Mission (TRMM) satellite, designed to monitor tropical rainfall, is placed in an orbit with a relatively low inclination (35 degrees), permitting frequent observation of equatorial regions. Other orbits likewise

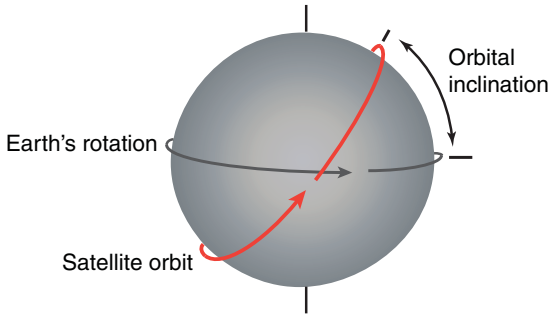


FIGURE 3.9 Orbital inclination. From NASA.

provide specific capabilities matched to satellite missions. For example, *polar orbits* pass close to both poles on each revolution (inclinations at or close to 90°). Sometimes the designation *near-polar orbit* signifies remote sensing satellites positioned within a few degrees of truly polar. Such satellites, positioned in near-polar orbits, indicate that the satellite initially progresses northward as it approaches on one side of the Earth, then southward as it passes to the opposite side.

Orbital altitude is the second important parameter for describing orbits. Satellites orbit at all altitudes, and by Newton's Laws, the speed of a satellite is a direct function of its altitude. It is noteworthy that orbital altitude is traditionally considered in orbital mechanics textbooks as the *semi-major axis*.

Not all orbits are useful for remote sensing purposes. Many factors are considered when selecting an orbit for a remote sensing mission, including, but not limited to, the required spatial resolution, field of view, and the illumination angle of features on the surface. A relatively small set of orbital categories suffice to describe orbits for remote sensing satellites (see **Figure 3.10** for some examples).

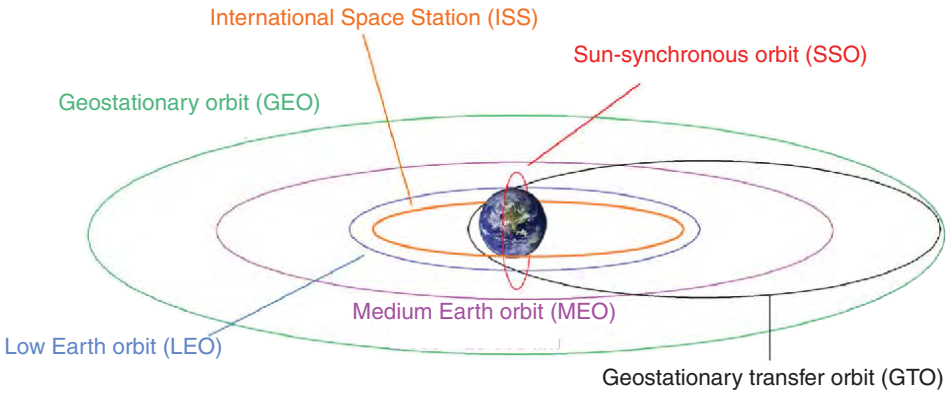


FIGURE 3.10 Several distinctive orbital units, including the International Space Station, the geostationary transfer orbit, and the sun-synchronous orbit, together contributing to our understanding of orbital systems.

1. *Low Earth orbit (LEO):* LEOs operate within the region closest to Earth, roughly described as starting at the top of the atmosphere, extending to about 2,000 km altitude—a region above most of the Earth's atmosphere and below 2,000 km. This zone favors economical launch and maintenance of satellite orbits. LEO is used often for scientific satellites, including remote sensing satellites, and for many weather satellites, which are often placed in nearly circular LEOs. For imaging systems, the LEO's low altitudes favor capture of high-resolution imagery.

2. *Sun-synchronous orbit (SSO):* The SSO is a special type of LEO orbit inclined at or near 98 degrees. Not all LEO satellites are sun-synchronous, but *all* sun-synchronous satellites operate in LEO. Of course, the sun-synchronous satellites are important enough to merit individual discussion. The defining property of a sun-synchronous orbit is that, due to its slightly off-polar inclination, it passes over locations on the ground at the same solar time every day, owing to a property called *orbital precession*. This characterization is a particularly useful feature for missions that use passive sensors to gather reflected sunlight, since it reduces discrepancies due to illumination differences between consecutive imaging passes. The list of satellites operating in SSO is quite extensive as it is perhaps the most popular orbit for scientific remote sensing satellites. Some particularly well-known examples of SSO satellites include Landsat, Polar Operational Environmental Satellite (with the Advanced Very-High-Resolution Radiometer [AVHRR] instruments), and TIROS Operational Vertical Sounder (TOVS).

3. *Medium Earth orbit (MEO):* While MEO can sometimes be used to describe the broad region between LEO and GEO (geosynchronous Earth orbit), spanning the altitude ranges of 2,000 km to 35,780 km, it is almost always used to refer to a much narrower band of particularly useful orbits: the *semisynchronous orbits*. These orbits are characterized by an orbital period of approximately 12 hours (hence the name because they orbit the Earth twice for every rotation) and fall close to 20,200 km altitude. These MEO orbits are used almost exclusively for Global Navigation Satellite Systems (GNSS), including the United States' Global Positioning System (GPS), Russia's GLONASS, China's Compass, and the European Union's Galileo systems. From our geospatial perspective, the principal satellites of interest in MEO are navigation satellites; the GPS is composed of 24 satellites that orbit at an altitude of 20,000 km above the Earth. Differences in time delay for signals received from four of the satellites form the basis for assessing the precise locations of a GPS receiver on the Earth's surface.

4. *Geosynchronous Earth orbit (GEO):* At an altitude of 35,786 km and an inclination at or near 0 degrees, GEO orbits are very popular. Satellites in this orbit have the useful property of orbiting at the same angular rate as the Earth rotates, so they appear to be stationary with respect to the surface of the Earth. While the GEO is highly valued as a location for communication satellites, it is also used extensively for remote sensing missions. Such orbits are valuable because a satellite in such an orbit can view the same location. For example, communication satellites are positioned to provide telecommunication links and television signals, and to gather meteorological data. Weather satellites in geosynchronous orbits are especially valuable for their ability to provide views of the progress of weather systems for an entire hemisphere.

Weather satellites such as the U.S. Geostationary Operational Environmental Satellite (GOES) and the European Meteosat satellites operate from GEO. GEO is also unique in that it consists exclusively of equatorial (i.e., 0-degree-inclination) orbits. A satellite at

the GEO altitude of 35,786 km but at any other inclination would drift relative to Earth. Occasionally, this effect happens to GEO satellites as a result of perturbations, but it is not a desirable situation and may even result in the early termination of a mission. The ability to remain stationary over a fixed location on the equator is the most important feature of this category of orbits. The main disadvantage of GEO orbits from a remote sensing perspective is that they are quite far from the Earth. The trade-off comes in the form of resolution for field of view. A satellite in GEO can observe almost an entire hemisphere but at the cost of relatively low resolution, while a satellite in LEO can image at a comparatively high resolution but has a much smaller field of view.

Many other types of orbits exist aside from those discussed above, but they generally are not of significance for remote sensing applications.

Image Footprint

In operation, satellite sensors view a strip of land centered on the system's ground track (**Figures 3.11 and 3.12**), forming the system's *swath width*. The swath is typically subdivided into quadrilaterals (**Figure 3.12**) that form *scenes*, the basic image units for a specific remote sensing system. Scenes are typically identified by unique scene IDs that identify dates, times, and locations of images. Sizes of scenes vary with the specific satellite system, ranging from a few kilometers on a side for high-resolution commercial satellite systems to much larger dimensions for broad-scale systems. Each scene is composed of picture elements ("pixels"—the uniform cells that systematically subdivide the image)

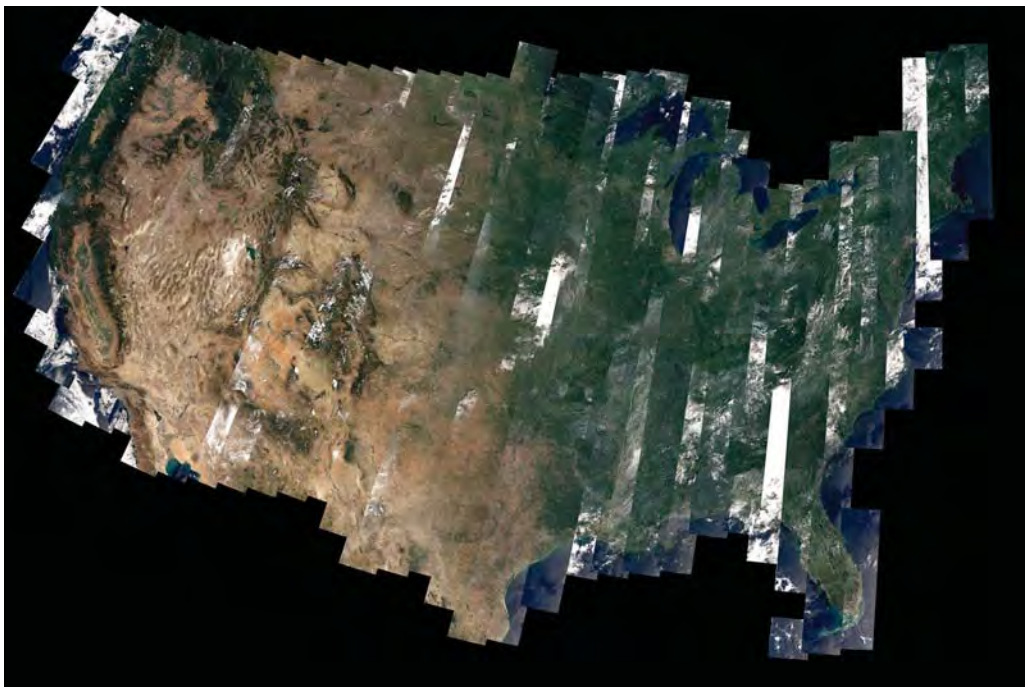


FIGURE 3.11 Composite of Landsat image tracks for the 48 U.S. states, illustrating the NE–SW orientation of the satellite track and its sun-synchronous orbit (on its descending track), and width of the image coverage track, known as *swath width*. From NASA.

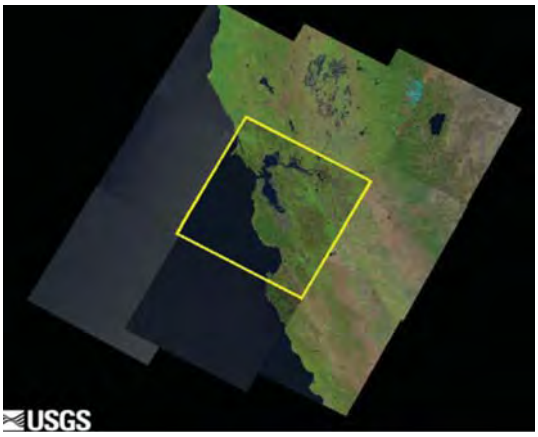


FIGURE 3.12 A single Landsat image footprint (the yellow outline), central California (San Francisco, and its region). The Landsat image footprint is about 185 × 185 km (115 × 115 mi.) in size. From U.S. Geological Survey.

forming the image's basic informational units. Small-footprint satellite systems might have fine-resolution pixels of a few meters or less. Systems with much larger footprints use larger pixels, providing much coarser spatial detail.

Satellite Constellations

When a group of satellites work in concert, they are referred to as a *satellite constellation*. In some cases, constellations allow for more complete and continuous coverage of areas on the Earth as the satellites travel along their orbits. There are common examples of these constellations that are relevant for remote sensing applications, including the well-known GPS constellation mentioned earlier. These are essentially multiple instances of the same or similar satellites, operating as part of the same system. In other cases, satellites with different but complementary systems may closely follow each other along the same orbit path. Each satellite is independent, but the constellation allows acquisition of data from different sensors that are collected within very close timing over the same location on Earth. An example of this type of configuration is the international afternoon constellation, which consists of a series of satellites that collect information on the atmosphere and Earth's surface (e.g., water, aerosols, temperature, clouds, wind). This information can be used synergistically to better understand atmospheric dynamics (see *https://atrain.nasa.gov* for additional information). Finally, constellations are also used for CubeSats. These are miniaturized satellites (typically 10-cm cubes) that are much cheaper than other systems, and they often use off-the-shelf components. They are usually deployed in clusters and are frequently launched as a secondary payload off a launch vehicle for a different mission (or, in some cases, they may be launched from the International Space Station).

3.6 UNMANNED AERIAL SYSTEMS

An unmanned aerial system (UAS) is a powered aerial vehicle, using aerodynamic lift, which is remotely or autonomously piloted. Although their long history is rooted mainly in military applications, current uses reach deep into a broad range of civil applications, including science, meteorology, agriculture, and operations in environments that are

dangerous for humans. These systems have other designations such as unmanned aerial vehicles (UAVs) or drones. The usage UAS has been officially adopted by the U.S. Federal Aviation Administration, Department of Defense, and similar organizations in other nations and will be used henceforth in this book.

Fixed-Wing UAS

For many decades, the civil interest in UAS resided in dedicated hobbyists; often, they were members of local clubs who specialized in gasoline-powered, fixed-wing, aerial vehicles, often built or modified by the operators for their specific avocations (**Figure 3.13**). These systems can be relatively large, awkward to handle, and difficult to operate, often requiring skill and experience both to construct and to operate. These UASs were usually piloted from the ground by radio link. In recent decades, technological advances (chiefly, improvements in batteries, miniaturization of inertial navigation systems, navigation components, and software) have permitted manufacturers to design the large variety of inexpensive, compact, battery-powered UASs now in wide use (**Figure 3.14**). These systems can be smaller, lighter vehicles, some of Styrofoam construction, which is feasible because of the availability of lighter battery and sensor technologies (**Figures 3.14 and 3.15**). In this form, the UAS is inexpensive, compact, and flexible, widely used for both practical applications and recreational activities.

Rotorcraft/Helicopters

The conventional helicopter design has been used for UAS, and often for larger UASs designed for military applications. A popular UAS design is the *quadcopter,* a version of the helicopter design powered by two pairs of rotors, positioned at the extremities of four arms (**Figure 3.16**). Quadcopters are characterized by their compact size, maneuverability, and ability to operate in confined spaces, leading to their role as a favored design for a large proportion of UAS systems in operation. The four rotary blades (two rotors rotating clockwise and two counterclockwise) favor effective lift, stability, safety, and maneuverability. Such characteristics are feasible because of technological advances in the designs of inexpensive, lightweight flight controllers, accelerometers (IMU), global

FIGURE 3.13 Flight preparation for gasoline-powered, fixed-wing drone/UAS.

FIGURE 3.14 Flight check for battery-powered fixed-wing SenseFly eBee UAS prior to launch.

FIGURE 3.15 Trimble UX5 fixed wing UAS. This style of fixed-wing UAS is relatively inexpensive but has limited range, and it can be subject to loss in trees, construction sites, and the like. From Trimble Navigation Limited. Used by permission.

positioning systems, and compact sensors. The quadcopter system independently varies speeds of each rotor to control the vertical and lateral orientation and motion of the vehicle. This design results in stability and ability to launch vertically, permitting use in confined spaces (e.g., forested zones) without the clear horizontal track required for launch of fixed-wing vehicles.

UAS Sensors

The UASs described here can accommodate sensors comparable to those carried by other platforms, in sizes comparable to cell phone cameras, capturing imagery in the visible (blue, green, and near infrared). UASs can also carry specialized sensors, thermal infrared, and multispectral (i.e., beyond visible region). Although miniaturized versions of hyperspectral, lidar, and SARs have been developed for UASs (these sensors are described in subsequent chapters), their effectiveness is still being tested for many applications, and thus they have not yet been widely adopted for commercial purposes.

FIGURE 3.16 Trimble ZX5 quadcopter UAS, illustrating a more expensive design, but one with increased capabilities. From Trimble Navigation Limited. Used by permission.

UAS Flight Planning and Operations

Current UAS systems provide the ability to program detailed flight plans (autonomous flight planning) by delineating an area of interest, setting the flight altitude, and outlining the specifics of coverage. The analyst can program flight lines to follow a predetermined strategy or can set the program to optimize coverage through orientation of flight lines (Figure 3.17). A UAS flight plan can set a *geofence*. In the context of UAS, geofencing is a software program that establishes a virtual barrier for an in-flight UAS, using GPS to define geographic boundaries within the flight plan (Figure 3.17). The geofence prevents an in-flight UAS from leaving a predetermined region, preventing encroachment into approaching prohibited/undesired areas or entering restricted airspace. Commonly, flight programming detects when battery power declines to unacceptably low levels and directs the UAS to return to the launch point to prevent loss of the vehicle due to battery failure. UAS battery life is one of the main limitations for UAS applications and must be carefully planned for.

Applications

Although a list of current UAS applications would be far too long, some of the most frequently reported uses indicate that UAS technology is flexible and may apply to a broad range of applications:

- Agriculture—inspect and monitor fields and rangeland
- Precision agriculture—provide detailed maps of within-field soil and moisture patterns
- Forestry—monitor forest growth, occurrence of disease, and insect infestations
- Emergency response—floods, fire, extreme weather, transportation accidents, industrial accidents
- Infrastructure monitoring and maintenance—monitoring pipelines, powerlines, towers, etc.
- Mobile collection of field data (described in Section 3.8)

FIGURE 3.17 Left: UAS flight plan, showing geofence (thick circular outline) boundary and flight paths. Right: UAS flight plan, highlighting curved tracks positioned outside the coverage zone to align successive paths to build systematic photographic coverage.

Most people are usually quick to realize the advantages that UASs offer for economical acquisition of imagery with fine spatial detail, but they are often slower to recognize their value for quick responses to record details of events such as floods, tornado damage, fire, or coastal erosion, when critical details of the events may vanish within a few hours.

UASs contribute to a trend in the practice of remote sensing that has seen remote sensing move from a field requiring specialized knowledge and software that could be mastered only by people willing to devote considerable effort to learning specialized techniques to one that is increasingly accessible for a wide variety of applications.

3.7 TETHERED BALLOONS

Tethered balloons are unusual remote sensing platforms, but they are practical for specialized applications. The technical but infrequently used term *aerostat* designates a tethered balloon of strong construction, typically using helium for buoyancy, without its own means of propulsion. The tether is usually a strong cable, or set of cables, controlled by a winch, to anchor the balloon in fixed position.

Tethered balloons have a long history in remote sensing (Vierling et al., 2006), initially in military service, but now include increasing numbers of civil applications. Although various types of balloons were used as platforms for early aerial photography, their systematic use began in World War I for military use. Prior to the common use of aircraft, tethered balloons were used for artillery observation and, to a lesser extent, for aerial photography (Campbell, 2008). Once the airplane came into routine use, the balloon was vulnerable to attack by hostile aircraft, so use of balloons as photographic platforms declined. Later, especially during World War II, balloons were used in antisubmarine warfare because of their ability to linger over designated areas to observe the same region for extended intervals.

Tethered balloons are currently in use to support research, offering, as Vierling et al. (2006) have noted, opportunities for continuous service over weeks, as well as the advantages of precise control over altitude and ease of use in remote regions, when aircraft use is restricted or difficult. They offer benefits of low costs and freedom from the low-frequency vibration that can create problems for other platforms, especially helicopters. As remote sensing platforms, tethered balloons can lift heavy equipment (including cameras and other sensors). The balloon described by Vierling et al. (2006) measures, when in service, 12 m in length and 4.3 m in width, with sensors controlled remotely by wireless communications, powered by batteries (Figure 3.18).

In remote sensing contexts, tethered balloons have been used in numerous applications, including monitoring wildlife (Nosal et al., 2013) and collecting hyperspectral data (Chen and Vierling, 2006). Perhaps the longest established use of tethered balloons for aerial observation has been in the field of archaeology (Verhoeven, 2009). The ability to fix the position of the balloon permits archaeologists to monitor the progress of an archaeological excavation using sequential photography that provides day-by-day (or perhaps hour-by hour) records of an excavation's progress. Thus, in archaeology, the balloon provides the ability to measure locations and configurations of excavated objects. Under development now are much larger balloons designed for use as either fixed-point or mobile units for surveillance of very large areas, such as border zones, traffic patterns, oil fields, or wildlife (Laskas, 2016).

(a) (b) (c)

FIGURE 3.18 Tethered balloon: (a) Inflating a balloon in preparation for launch. (b) Tethered balloon (on the left) deployed for flight, anchored by a cable, with the instrument platform visible on the right. (c) Closer view of the instrument platform during flight preparation, with a selection of sensors visible near the upper portion of the platform. Photographs by Lee Vierling. Used by permission.

3.8 MOBILE COLLECTION OF FIELD DATA

Mobile data collection refers to the use of devices (usually mobile phones or tablets) that enable the analyst to collect *in situ* data in real time (i.e., data can be collected in the field as they are observed—locations, time, measurement, characterization) without the intermediate steps of transcribing based on use of portable devices.

The current availability of mobile technologies provides opportunities to use devices such as smartphones (Android or IOS) with GPS capabilities to record location, time, and date. Many different mobile apps are available for download to a mobile phone or tablet. Often, they include a menu to code identifying information (e.g., tree species, crop types, forests, or geology) or the analyst's notes. Such information can be posted to a layer within a geographic information system (GIS) to support geospatial analysis. Examples include collection of data for accuracy assessment, training data for image classification, and validation data collected directly in the field. Other examples include labeling point features, updating databases, and recording time-sensitive data. Such applications bypass the usual steps that require separate operations for recording, transporting, recoding, and manual data entry. Mobile data collection apps enable analysts to update data in the field, integrate field observations with other data, and immediately share with other analysts. Although there are many more apps, four common mobile field data collection apps are briefly described next (Figure 3.19).

- *Collector for ArcGIS* App (*http://doc.arcgis.com/en/collector*) is a recent introduction to the Environmental Systems Research Institute's suite of ArcGIS products. Collector enables real-time collection of point, line, and polygon features, together with immediate entry to ArcGIS Online.

- *Nature's Notebook* App (*https://usanpn.org/natures_notebook*) is designed for entry of phenological data, as observed on-site. Its use with a mobile phone can record geographical locations, as well as dates and times that match to field observations, building a phenological record in time and space.

FIGURE 3.19 Examples of four mobile apps for collecting data field *in situ.* Clockwise from upper left: Avenza Maps logo, ESRI Collector app, Nature's Notebook app, Fulcrum app.

- *Avenza Maps* (*www.avenza.com*) permits users to enter field data, using mobile phones or tablets, on digital maps prepared in advance by the user or selected from a collection available online. The maps form the framework for entering field information, as observed in the field. Because other workers can simultaneously use the same template, the PDF map system is effective for team efforts. At any time, the maps can be observed to evaluate the overall data collection effort (**Figure 3.20**).

- The *Fulcrum App* (*http://fulcrumapp.com*) is an Android app that permits mobile collection of field data. It provides options for several application themes (including agriculture, archaeology, construction, and environment). For each template, alternative collection schemes (e.g., for environmental surveys, invasive species, and land cover survey) are

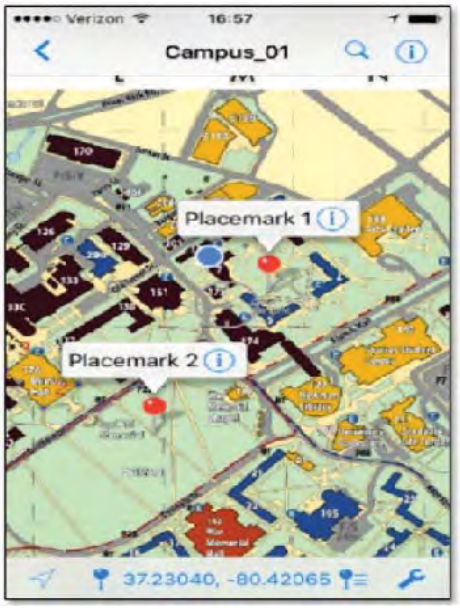


FIGURE 3.20 Sample screen display for Avenza Maps on the IOS platform. Users can place and label markers.

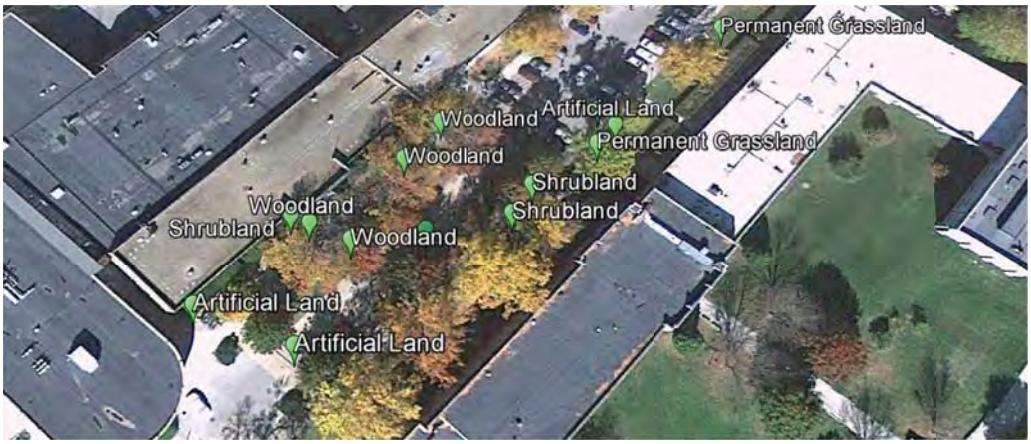


FIGURE 3.21 Data points labeled using the Fulcrum mobile app (Android).

available. The app registers data as a point, but it can include photographs to document the landscape. Data can be exported to a .kml file to use with other GIS software (Figure 3.21).

3.9 SUMMARY

This chapter briefly reviewed the common platforms on which remote sensing instruments are mounted, within the contexts of the types of remote sensing described in this book. Of course, the platforms mentioned can be explored in much greater detail within their own specific disciplines. The fields of aeronautics, geodetics, and orbital mechanics have all made significant contributions to the practical utility of remote sensing, but they are well beyond the scope of this book. For our purposes, the concepts of revisit (systematic or otherwise), altitude, and swath/coverage have major impacts on the type, resolution, timing, characteristics, and utility of the remote sensing imagery that are used for applications described in later chapters.

REVIEW QUESTIONS

1. This chapter discusses numerous platforms that can be used for remote sensing. How would you decide which type of platform is most suitable for a given application?
2. CubeSats were described as miniaturized, low-cost satellites that are often launched in clusters. Under what circumstances might CubeSats be desired over the larger, more complex satellites typically run by space agencies or companies?
3. What type of platform is best suited to track a large weather system, such as a hurricane, and why? What type of platform would be most useful to monitor forest fires in the Amazon and why? For each of these applications, what are some trade-offs that must be considered if only the platform could be chosen?

4. UASs have become increasingly popular in recent years. What are some new remote sensing applications that have emerged with them, that might not be possible with other types of remote sensing platforms?
5. Mobile field data collection platforms allow for the immediate integration of field data with other data, including remote sensing data. What are some applications for which this would be a major advantage of the more time-consuming methods previously available to us?

REFERENCES

Campbell, J. B. 2008. Origins of Aerial Photographic Interpretation, U.S. Army, 1916 to 1918. *Photogrammetric Engineering and Remote Sensing,* Vol. 74, pp. 77–93.

Campbell, J. B., and V. V. Salomonson. 2010. Remote Sensing: A Look to the Future. Chapter 25 in *Manual of Geospatial Science and Technology* (2nd ed.) (J. D. Bossler, ed.). Boca Raton, FL: CRC Press, pp. 487–509.

Chen, X., and L. Vierling. 2006. Spectral Mixture Analyses of Hyperspectral Data Acquired Using a Tethered Balloon. *Remote Sensing of Environment,* Vol. 103, pp. 338–350.

Gertler, J. 2012. *US Unmanned Aerial Systems*. Report R42136. Washington, DC: Library of Congress, Congressional Research Service.

Laskas, J. M. 2016, February 21. Helium Dreams. *The New Yorker*, pp. 28–34.

Nosal, P., D. C. Cartamil, J. W. Long, M. Lührmann, N. C. Wegner, and J. B. Graham. 2013. Demography and Movement Patterns of Leopard Sharks (*Triakis Semifasciata*) Aggregating Near the Head of a Submarine Canyon Along the Open Coast of Southern California, USA. *Environmental Biology of Fishes,* Vol. 96, pp. 865–878.

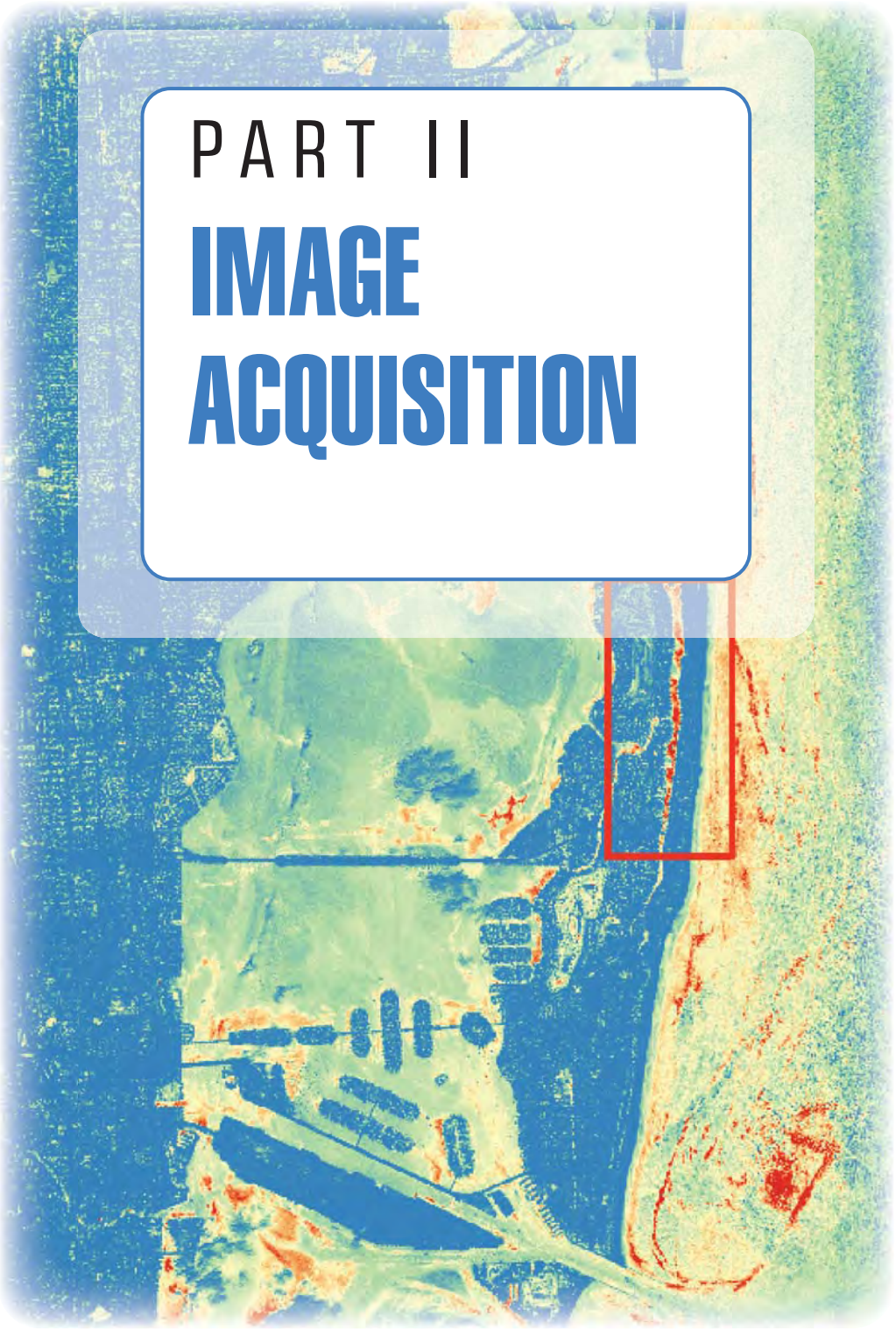
Shahbazi, M., J. Theau, and P. Ménard. 2014. Recent Applications of Unmanned Aerial Imagery in Natural Resource Management. *GIScience and Remote Sensing,* Vol. 51, pp. 339–365.

Verhoeven, G. J. J. 2009. Providing an Archeological Bird's-Eye View—An Overall Picture of Ground-Based Means to Execute Low-Altitude Aerial Photography (LLAP) in Archeology. *Archaeological Prospection,* Vol. 16, pp. 233–239.

Verhoeven, G. J. J., J. Loenders, F. Vermeulen, and R. Docter. 2009. Helikite Aerial Photography—A Versatile Means of Unmanned, Radio Controlled, Low-Altitude Aerial Archeology. *Archaeological Prospection,* Vol. 16, pp. 125–138.

Vierling, L., M. Fersdahl, X. Chen, Z. Li, and P. Zimmerman. 2006. The Short Wave Aerostat-Mounted Imager (SWAMI): A Novel Platform for Acquiring Remotely Sensed Data from a Tethered Balloon. *Remote Sensing of Environment,* Vol. 103, pp. 255–264.

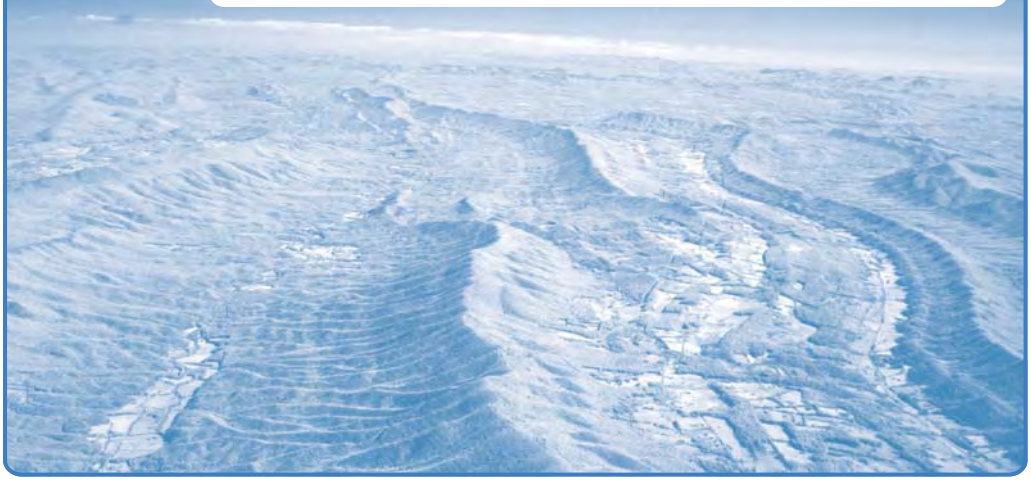
Xiao, X., P. Dorovskoy, C. M. Biradar, and E. Bridge. 2011. A Library of Geo-Referenced Field Photos from the Field. *EOS Transactions,* Vol. 92(49), pp. 453–454.



PART II

IMAGE ACQUISITION

4 Digital Mapping Cameras



MAJOR TOPICS TO UNDERSTAND

- Fundamentals of the Aerial Photograph
- Geometry of the Vertical Aerial Photograph
- Digital Aerial Cameras
- Digital Scanning of Analog Images
- Spectral Sensitivity
- Band Combinations: Optical Imagery
- Coverage by Multiple Photographs
- Photogrammetry
- Sources of Aerial Photography
- Your Own Infrared Photographs
- Your Own 3D Photographs
- Your Own Kite Photography

4.1 INTRODUCTION

This chapter introduces sensors used for acquiring aerial photographs. Although cameras are the oldest form of remote sensing instrument, they have changed dramatically

in recent decades, yet nonetheless exhibit continuity with respect to their fundamental purposes. Cameras designed for aerial use capture imagery that provides high positional accuracy and fine spatial detail. Despite the many other forms of remotely sensed imagery in use today, aerial photography remains a widely used form of aerial imagery, employed for a wide variety of tasks by local and state governments, private businesses, and federal agencies to gather information to support planning, environmental studies, construction, transportation studies, routing of utilities, and many other tasks. The versatility of these images accounts for a large part of their enduring utility over the decades, even as fundamental technological shifts have transformed the means by which these images are acquired and analyzed. It is noteworthy that, especially in the United States, there is a large archive of aerial photographs acquired over the decades that form an increasingly valuable record of landscape changes since the 1930s.

During recent decades, the cameras, films, and related components that long formed the basis for traditional photographic systems (known as *analog* technologies) are rapidly being replaced by *digital* instruments that provide imagery with comparable characteristics that is acquired using electronic technologies. Here we introduce basic concepts that apply to these sensors, characterized by their use of aircraft as a platform and of the visible and near-infrared spectrum, and by their ability to produce imagery with fine detail and robust geometry. Although the majority of this chapter presents broad, generally applicable concepts without reference to specific instruments, it does introduce a selection of specific systems now used for acquiring aerial imagery.

The transition from analog-to-digital aerial cameras has been under way for several decades and is now maturing with respect to collection of imagery, for analysis, and storage. Yet, digital systems are still evolving, with a variety of systems in use and under development, with uncertain standards, and with discussion of relative merits of alternative systems. The following sections, therefore, present a snapshot of the transition from analog-to-digital technologies.

4.2 FUNDAMENTALS OF THE AERIAL PHOTOGRAPH

Systems for acquiring aerial images rely on the basic components common to the familiar handheld cameras we all have used for everyday photography: (1) a lens to gather light to form an image; (2) a light-sensitive surface to record the image; (3) a shutter that controls entry of light; and (4) a camera body—a light-tight enclosure that holds the components together in their correct positions (Figure 4.1). Aerial cameras include these components in a structure that differs from that encountered in our everyday experience with cameras: (1) a film magazine, (2) a drive mechanism, and (3) a lens cone (Figure 4.1). This structure characterizes the typical design for the analog aerial camera that has been used (in its many variations) for aerial photography starting in the early 1900s. Although alternative versions of analog cameras were tailored to optimize specific capabilities, for our discussion, it is the *metric*, or *cartographic*, camera that has the greatest significance. Whereas other cameras may have been designed to acquire images (for example) of very large areas or under unfavorable operational conditions, the design of the metric camera is optimized to acquire high-quality imagery of high positional fidelity; it is the metric camera that forms the current standard for aerial photography.

For most of the history of remote sensing, aerial images were recorded as photographs or photograph-like images. A photograph forms a physical record—paper or film

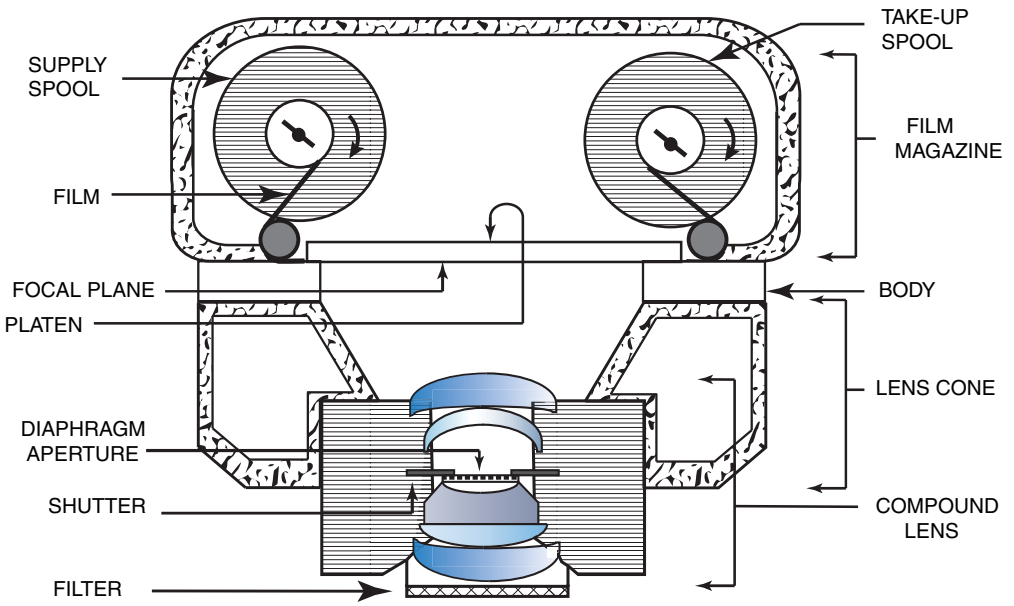


FIGURE 4.1 Schematic diagram of an aerial framing camera, cross-sectional view. Image by Susmita Sen.

with chemical coatings that portray the patterns of the images. Such images are referred to as *analog* images because the brightness of a photograph is proportional (i.e., analogous) to the brightness in a scene. Although photographic media have value for recording images, in the context of remote sensing, their disadvantages, including difficulties of storage, transmission, searching, and analysis, set the stage for replacement by digital media. Digital technologies, in contrast, record image data as arrays of individual values that convey the pattern of brightness within an image.

Although a digital aerial camera shares many of the components and characteristics outlined above, in detail its design differs significantly from that of the analog camera. Because the image is captured by digital technology, digital cameras do not require the film and the complex mechanisms for manipulating the film. Furthermore, digital cameras often include many capabilities that were not fully developed during the analog era, including links to positional and navigational systems and elaborate systems for annotating images.

The Lens

The *lens* gathers reflected light and focuses it on the focal plane to form an image. In its simplest form, a lens is a glass disk carefully ground into a shape with nonparallel curved surfaces (Figure 4.2). The change in optical densities as light rays pass from the atmosphere to the lens and back to the atmosphere causes refraction of light rays; the sizes, shapes, arrangements, and compositions of camera lenses are carefully designed to control refraction of light to maintain color balance and to minimize optical distortions. Optical characteristics of lenses are determined largely by the refractive index of the glass

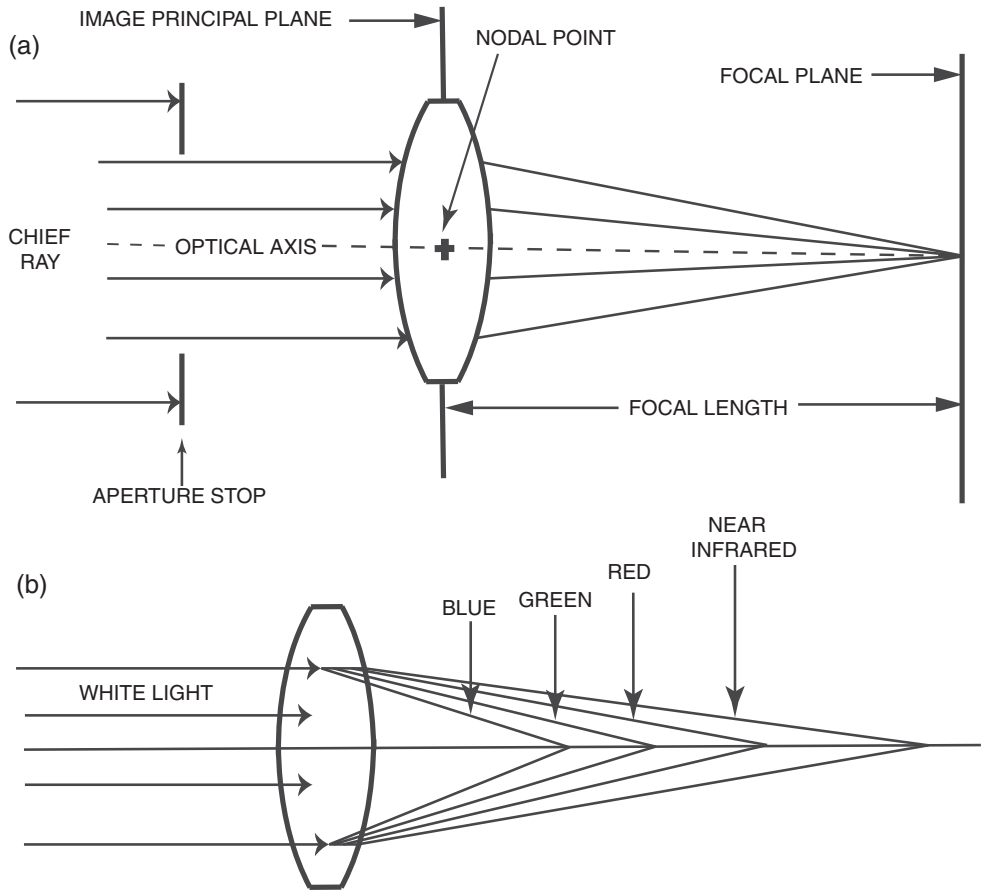


FIGURE 4.2 (a) Cross-sectional view of a simple lens as formed by a (b) chromatic aberration. Light of differing wavelengths is brought to focus at differing distances from the lens. The more complex lenses designed for aerial cameras can bring the differing wavelengths to a common focal point. Image by Susmita Sen.

(Chapter 2) and the degree of curvature. The quality of a lens is determined by the quality of its glass, the precision with which that glass is shaped, and the accuracy with which it is positioned within a camera. Imperfections in lens shape contribute to *spherical aberration*, a source of error that distorts images and causes loss of image clarity. For modern aerial photography, spherical aberration is usually not a severe problem because most modern aerial cameras use lenses of very high quality. Figure 4.2a shows the simplest of all lenses: a simple positive lens. Such a lens is formed from a glass disk with equal curvature on both sides; light rays are refracted at both edges to form an image.

Most aerial cameras use *compound lenses*, formed from many separate lenses of varied sizes, shapes, and optical properties. These components are designed to correct for errors that may be present in any single component, so the whole unit is much more accurate than any single element. For present purposes, consideration of a simple lens will

be sufficient to define the most important features of lenses, even though a simple lens differs greatly from those actually used in modern aerial cameras. The *optical axis* joins the centers of curvature of the two sides of the lens. Although refraction occurs throughout a lens, a plane passing through the center of the lens, known as the *image principal plane,* is considered to be the center of refraction within the lens (Figure 4.2a). The image principal plane intersects the optical axis at the *nodal point*.

Parallel light rays reflected from an object at a great distance (at an "infinite" distance) pass through the lens and are brought to focus at the principal *focal point*—the point at which the lens forms an image of the distant object. The *chief ray* passes through the nodal point without changing direction; the paths of all other rays are deflected by the lens. A plane passing through the focal point parallel to the image principal plane is known as the *focal plane*. For handheld cameras, the distance from the lens to the object is important because the image is brought into focus at distances that increase as the object is positioned closer to the lens. For such cameras, it is important to use lenses that can be adjusted to bring each object to a correct focus as the distance from the camera to the object changes. For aerial cameras, the scene to be photographed is always at such large distances from the camera that the focus can be fixed at infinity, with no need to change the focus of the lens.

In a simple positive lens, the focal length is defined as the distance from the center of the lens to the focal point and is usually measured in inches or millimeters. (For a compound lens, the definition is more complex.) For a given lens, the focal length is not identical for all wavelengths. Blue light is brought to a focal point at a shorter distance than are red or infrared wavelengths (Figure 4.2b). This effect is the source of *chromatic aberration*. Unless corrected by lens design, chromatic aberration would cause the individual colors of an image to be out of focus. In high-quality aerial cameras, chromatic aberration is corrected to ensure that radiation used to form the image is brought to a common focal point.

A lens's field of view can be controlled by a *field stop,* a mask positioned just in front of the focal plane. An *aperture stop* is usually positioned near the center of a compound lens; it consists of a mask with a circular opening of adjustable diameter (Figure 4.3). An aperture stop can control the intensity of light at the focal plane but does not influence the field of view or the size of the image. Manipulation of the aperture stop controls only the brightness of the image without changing its size. Usually, aperture size is measured as the diameter of the adjustable opening that admits light to the camera.

Relative aperture is defined as:

$$f = \frac{\text{Focal length}}{\text{Aperture size}} \qquad \text{(EQ. 4.1)}$$

where focal length and aperture are measured in the same units of length and f is the *f number,* the relative aperture. A large f number means that the aperture opening is small relative to focal length; a small f number means that the opening is large relative to focal length.

Why use f numbers rather than direct measurements of aperture? One reason is that standardization of aperture with respect to focal length permits specification of aperture sizes using a value that is independent of camera size. Specification of an aperture as "23

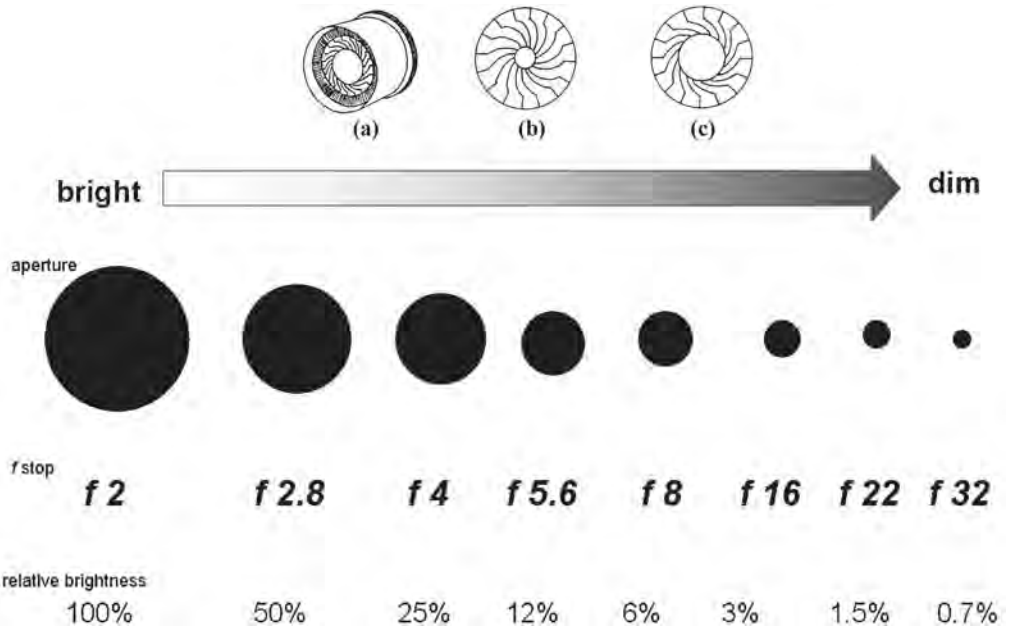


FIGURE 4.3 Diaphragm aperture stop. (a) Perspective view; (b) narrow aperture; (c) wide aperture. *f* stops are represented in the lower portion of the diagram.

mm" has no practical meaning unless we also know the size (focal length) of the camera. Specification of an aperture as "*f* 4" has meaning for cameras of all sizes; we know that it is one-fourth of the focal length for any size camera.

The standard sequence of apertures is: *f* 1, *f* 1.4, *f* 2, *f* 2.8, *f* 4, *f* 5.6, *f* 8, *f* 11, *f* 16, *f* 22, *f* 32, *f* 64 (and so forth). This sequence is designed to change the amount of light by a factor of 2 as the *f* stop is changed by one position. For example, a change from *f* 2 to *f* 2.8 halves the amount of light entering the camera; a change from *f* 11 to *f* 8 doubles the amount of light. A given lens, of course, is capable of using only a portion of the range of apertures mentioned above.

Lenses for aerial cameras typically have rather wide fields of view. As a result, light reaching the focal plane from the edges of the field of view is typically dimmer than light reflected from objects positioned near the center of the field of view. This effect creates a dark rim around the center of the aerial photograph—an effect known as *vignetting*. It is possible to employ an *antivignetting filter,* darker at the center and clearer at the periphery, which can be partially effective in evening brightness across the photograph. Digital systems can also employ image processing algorithms, rather than physical filters, to compensate for vignetting.

The Shutter

The *shutter* controls the length of time that the film is exposed to light. The simplest shutters are often metal blades positioned between elements of the lens, forming "intralens,"

or "between-the-lens," shutters. An alternative form of shutter is the focal plane shutter, consisting of a metal or fabric curtain positioned just in front of the detector array, near the focal plane. The curtain is constructed with a number of slits; the choice of shutter speed by the operator selects the opening that produces the desired exposure. Although some analog aerial cameras once used focal plane shutters, the between-the-lens shutter is preferred for most aerial cameras. The between-the-lens shutter subjects the entire focal plane to illumination simultaneously and presents a clearly defined perspective that permits use of the image as the basis for precise measurements.

Image Motion Compensation

High-quality aerial cameras usually include a capability known as *image motion compensation* (or *forward motion compensation*) to acquire high-quality images. Depending on the sensitivity of the recording media (either analog or digital), the forward motion of the aircraft can subject the image to blur when the aircraft is operated at low altitudes and/or high speeds. In the context of analog cameras, image motion compensation is achieved by mechanically moving the film focal plane at a speed that compensates for the apparent motion of the image in the focal plane. In the context of digital systems, image motion compensation is achieved electronically. Use of image motion compensation widens the range of conditions (e.g., lower altitudes and faster flight speeds) that can be used, while preserving the detail and clarity of the image.

4.3 GEOMETRY OF THE VERTICAL AERIAL PHOTOGRAPH

This section presents the basic geometry of a vertical aerial photograph as acquired by a classic framing camera. Not all portions of this discussion apply directly to digital cameras, but the concepts and terminology presented here do apply to a broad range of optical systems used for remote sensing instruments described both in this chapter and in later sections of this book.

Aerial photographs can be classified according to the orientation of the camera in relation to the ground at the time of exposure (Figure 4.4). *Oblique* aerial photographs have been acquired by cameras oriented toward the side of the aircraft. *High oblique* photographs (Figure 4.4a and Figure 4.5, left) show the horizon; *low oblique* photographs (Figure 4.4b and Figure 4.5, right) are acquired with the camera aimed more directly toward the ground surface and do not show the horizon. Oblique photographs have the advantage of showing very large areas in a single image. Often those features in the foreground are easily recognized, as the view in an oblique photograph may resemble that from a tall building or mountain peak. However, oblique photographs are not widely used for analytic purposes, primarily because the drastic changes in scale that occur from foreground to background prevent convenient measurement of distances, areas, and elevations.

Vertical photographs are acquired by a camera aimed directly at the ground surface from above (Figures 4.4c and 4.6). Although objects and features are often difficult to recognize from their representations on vertical photographs, the map-like view of the Earth and the predictable geometric properties of vertical photographs provide practical advantages. It should be noted that few, if any, aerial photographs are truly vertical; most

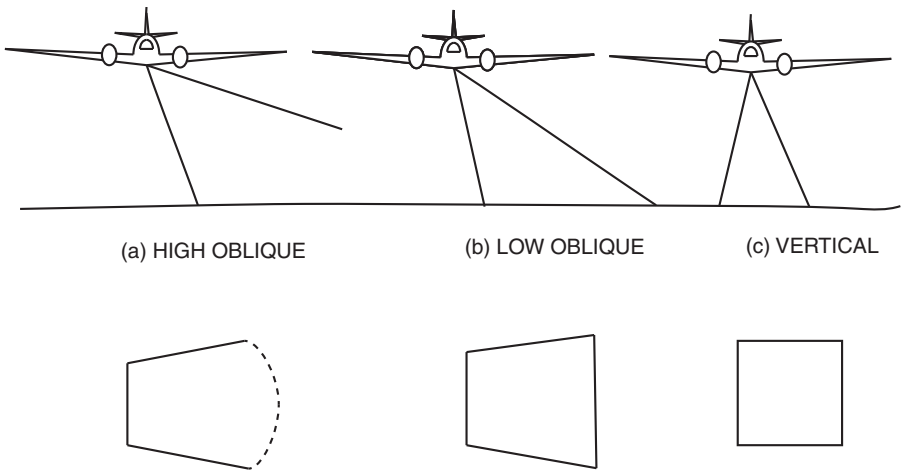


FIGURE 4.4 Oblique and vertical aerial photographs. Oblique orientations (a and b) provide intuitive perspectives but present large variations in image scale. Vertical photography (c) provides a more consistent scale but provides an unfamiliar view of the landscape. Image by Susmita Sen.

have some small degree of tilt due to aircraft motion and other factors. The term *vertical photograph* is commonly used to designate aerial photographs that are within a few degrees of a corresponding (hypothetical) truly vertical aerial photograph.

Because the geometric properties of vertical and nearly vertical aerial photographs are well understood and can be applied to many practical problems, they form the basis for making accurate measurements using aerial photographs. The science of making accurate measurements from aerial photographs (or from any photograph) is known as *photogrammetry*. The following paragraphs outline some of the most basic elements of introductory photogrammetry; the reader should consult a photogrammetry text (e.g., Wolf, 1983) for complete discussion of this subject.

FIGURE 4.5 Oblique aerial photographs, as acquired from a helicopter and light aircraft. Left: High oblique aerial photography; view to east from Denver, Colorado. Right: Low oblique aerial photograph, Blacksburg, Virginia.

FIGURE 4.6 Vertical aerial photography, Tampa, Florida, October 1987. From U.S. Geological Survey (USGS).

Analog aerial cameras are manufactured to include adjustable index marks attached rigidly to the camera so that the positions of the index marks are recorded on the photograph during exposure. These *fiducial marks* (usually four or eight in number) appear as silhouettes at the edges and/or corners of the photograph (**Figure 4.7**). Lines that connect opposite pairs of fiducial marks intersect to identify the *principal point,* defined as the intersection of the optical axis with the focal plane, which forms the optical center of the image.

The *ground nadir* is defined as the point on the ground vertically beneath the center of the camera lens at the time the photograph was taken (**Figure 4.8**). The *photographic nadir* is defined by the intersection with the photograph of the vertical line that intersects the ground nadir and the center of the lens (i.e., the image of the ground nadir). Accurate evaluation of these features depends on systematic and regular *calibration* of aerial cameras; the camera's internal optics and positioning of fiducial marks are assessed and adjusted to ensure the optical and positional accuracy of imagery for photogrammetric applications. Calibration can be achieved by using the cameras to photograph a standardized target designed to evaluate the quality of the imagery, as well as by internal measurements of the camera's internal geometry (Clarke and Fryer, 1998).

The *isocenter* can be defined informally as the focus of tilt. Imagine a truly vertical photograph that was taken at the same instant as the real, almost vertical, image. The almost vertical image would intersect with the (hypothetical) perfect image along a line

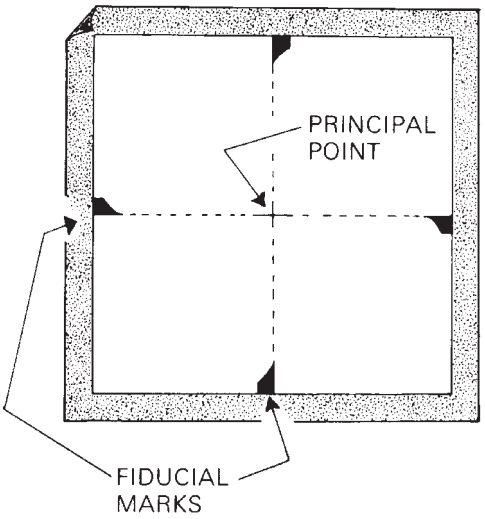


FIGURE 4.7 Fiducial marks and principal point, vertical aerial photography as acquired by a film camera.

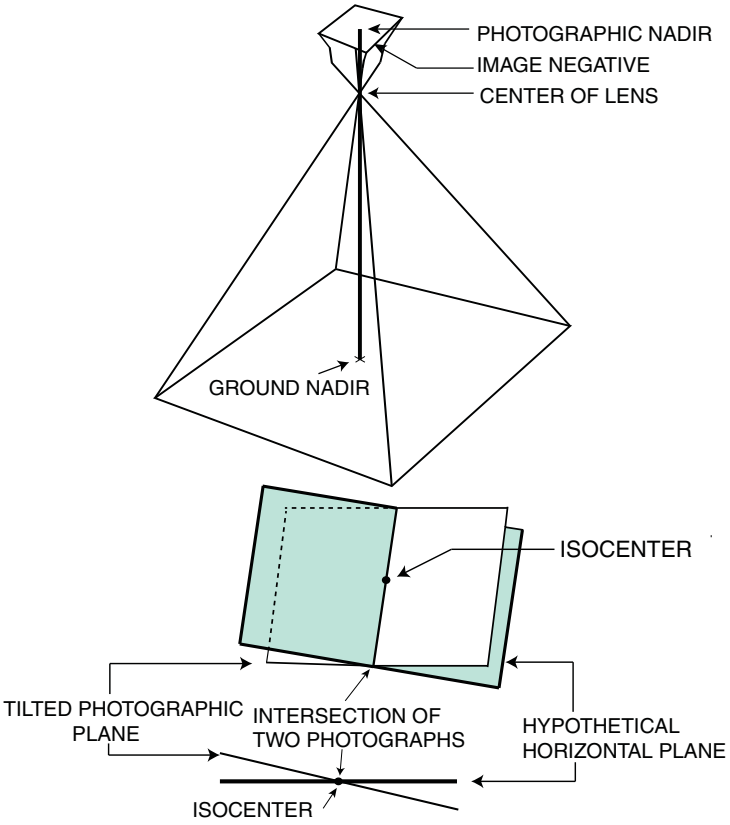


FIGURE 4.8 Schematic representation of key geometric features of a vertical aerial photograph as acquired by a framing camera. Image by Susmita Sen.

that would form a "hinge"; the isocenter is a point on this hinge. On a truly vertical photograph, the isocenter, the principal point, and the photographic nadir coincide. The most important positional errors in the vertical aerial photograph can be summarized as follows.

1. *Optical distortions* are errors caused by an inferior camera lens, camera malfunction, or similar problems. These distortions are probably of minor significance in most modern photography by professional aerial survey firms.

2. *Tilt* is caused by displacement of the focal plane from a truly horizontal position by aircraft motion (Figure 4.8). The focus of tilt, the isocenter, is located at or near the principal point. Image areas on the upper side of the tilt are displaced farther away from the ground than is the isocenter; these areas are therefore depicted at scales smaller than the nominal scale. Image areas on the lower side of the tilt are displaced down; these areas are depicted at scales larger than the nominal scale. Therefore, because all photographs have some degree of tilt, measurements confined to one portion of the image run the risk of including systematic error caused by tilt (i.e., measurements may be consistently too large or too small). To avoid this effect, it is a good practice to select distances used for scale measurements (Chapter 6) as lines that pass close to the principal point; then errors caused by the upward tilt compensate for errors caused by the downward tilt. The resulting value for image scale is not, of course, precisely accurate for either portion of the image, but it will not include the large errors that can arise in areas located farther from the principal point.

3. Because of routine use of high-quality cameras and careful inspection of photography to monitor image quality, today the most important source of positional error in vertical aerial photography is probably *relief displacement* (Figure 4.9). Objects positioned directly beneath the center of the camera lens will be photographed so that only the top of the object is visible (e.g., object A in Figure 4.9). All other objects are positioned such that both their tops and their sides are visible from the position of the lens. That is, these objects appear to lean outward from the central perspective of the camera lens. Correct planimetric positioning of these features would represent only the top view, yet the photograph shows both the top and sides of the object. For tall features, it is intuitively clear that the base and the top cannot both be in their correct planimetric positions.

This difference in apparent location is due to the height (*relief*) of the object and forms an important source of positional error in vertical aerial photographs. The direction of relief displacement is radial from the nadir; the amount of displacement depends on (1) the height of the object and (2) the distance of the object from the nadir. Relief displacement increases with increasing heights of features and with increasing distances from the nadir. (It also depends on focal length and flight altitude, but these may be regarded as constant for a selection of sequential photographs.) Relief displacement can form the basis of measurements of heights of objects, but its greatest significance is its role as a source of positional error. Uneven terrain can create significant relief displacement, so all measurements made directly from uncorrected aerial photographs are suspect. We should note that this is a source of positional error, but it is not the kind of error that can be corrected by selection of better equipment or more careful operation; it is an error that is caused by the central perspective of the lens and so is inherent to the choice of basic technology.

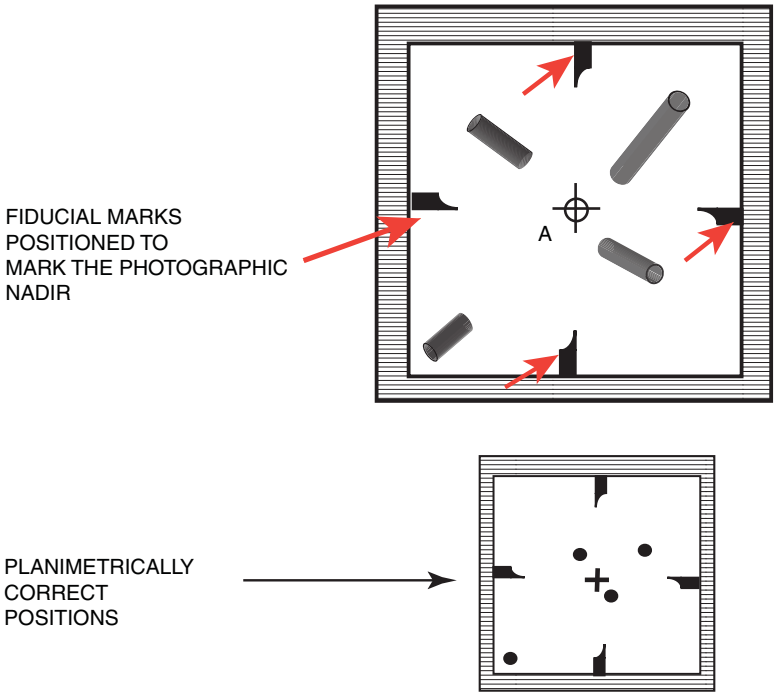


FIGURE 4.9 Relief displacement. The diagram displays a vertical aerial photograph of an idealized flat terrain with towers of equal height located at different positions with respect to the principal point. Images of the tops of towers are displaced away from the principal point along lines that radiate outward from the nadir. Image by Susmita Sen.

4.4 DIGITAL AERIAL CAMERAS

Digital imagery is acquired using a family of instruments that can systematically view portions of the Earth's surface, recording photons reflected or emitted from individual patches of ground, known as *pixels*, that together compose the array of discrete brightness values that form an image. Thus, a digital image is composed of a matrix of many thousands of pixels, each too small to be individually resolved by the human eye. Each pixel represents the brightness of a small region on the Earth's surface, recorded digitally as a numeric value, often with separate values for each of several regions of the electromagnetic spectrum (Figure 4.10).

Although the lens of any camera projects an image onto the focal plane, the mere formation of the image does not create a durable image that can be put to practical use. To record the image, it is necessary to position a light-sensitive material at the focal plane. Analog cameras record images using the photosensitive chemicals that coated the surfaces of photographic films, as previously described. In contrast, digital cameras use an array of detectors positioned at the focal plane to capture an electronic record of the image. Detectors are light-sensitive substances that generate minute electrical currents

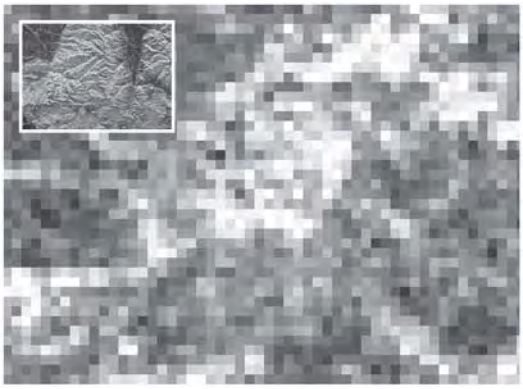


FIGURE 4.10 Pixels. A complete view of an image is represented in the inset; the larger image shows an enlargement of a small section to illustrate pixels that convey variations in brightness.

when they intercept photons from the lens, thereby creating an image from the matrix of brightness values that is proportional to the strengths of the electrical charges that reach the focal plane. Detectors in digital aerial cameras apply either of two alternative designs: charge-coupled devices (CCDs) or complementary metal oxide semiconductor (CMOS) chips (described below). Each strategy offers its own advantages and disadvantages.

A CCD (Figure 4.11) is formed from light-sensitive material embedded in a silicon chip. The *potential well* receives photons from the scene through an optical system designed to collect, filter, and focus radiation. The sensitive components of CCDs can be manufactured to be very small, perhaps as small as 1 μm in diameter, and sensitive to selected regions within the visible and near-infrared spectra. These elements can be connected to each other using microcircuitry to form *arrays*. Detectors arranged in a single line form a *linear array;* detectors arranged in multiple rows and columns form *two-dimensional arrays*. Individual detectors are so small that a linear array shorter than 2 cm in length might include several thousand separate detectors. Each detector within a CCD collects photons that strike its surface and accumulates a charge proportional to the intensity of the radiation it receives. At a specified interval, charges accumulated at each detector pass through a *transfer gate,* which controls the flow of data from the

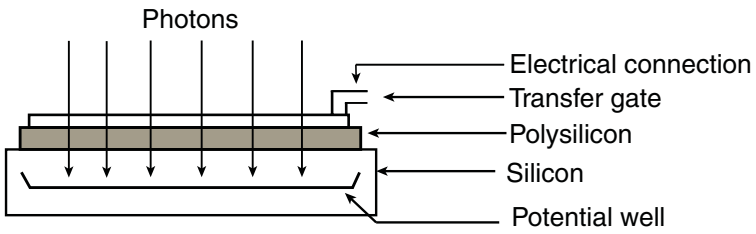


FIGURE 4.11 Schematic representation of a CCD. A CCD is a sensor for recording digital images, consisting of an integrated circuit containing an array of linked (coupled) capacitors serving as many small pixels. Light accumulating on each pixel is converted to charged pulse, which is then transmitted, measured by CCD electronics, and represented as a digital number. Image by Susmita Sen.

detectors. Microcircuits connect detectors within an array to form *shift registers*. Shift registers permit charges received at each detector to be passed to adjacent elements (in a manner analogous to a bucket brigade), temporarily recording the information until it is convenient to transfer it to another portion of the instrument. Through this process, information read from the shift register is read sequentially.

A CCD, therefore, scans electronically without the need for mechanical motion. Moreover, relative to other sensors, CCDs are compact and efficient in detecting photons (CCDs are especially effective when intensities are dim), and they respond linearly to brightness. As a result, CCD-based linear arrays have been used for remote sensing instruments that acquire imagery line by line as the motion of the aircraft or satellite carries the field of view forward along the flight track (**Figure 4.12**). As a result, over the past several decades, CCD technology has established a robust, reliable track record for scientific imaging. Disadvantages include (1) manufacturing expense, (2) high power consumption, and (3) analog-to-digital (A-to-D) conversion (Chapter 5), which requires an external design element.

An alternative imaging technology, CMOS, has a history of use for mobile phones and cameras and related consumer products. Early designs were characterized by low sensitivity and unfavorable signal-to-noise ratios. However, current CMOS-based instruments are recognized for their wide dynamic range, fast readout time, low power requirements, and ability to provide fine detail at low costs and at low power requirements. Some aerial cameras are now designed using CMOS technology (Neumann et al., 2016), a trend that will likely continue and increase.

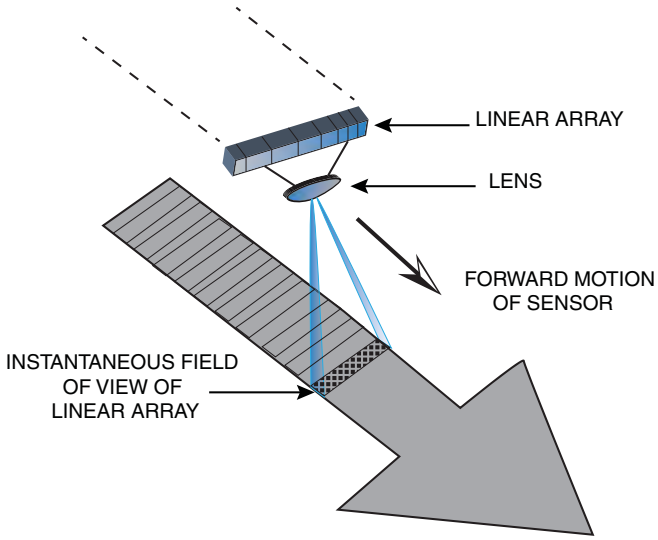


FIGURE 4.12 Schematic representation of a linear array. The satellite system uses an optical system that directs light to a linear array composed of CCDs (**Figure 4.10**), or similar devices, to capture light reflected from the Earth's surface. As the satellite's CCD scans the landscape along the orbital track, it forms a strip-like image of a region of the Earth (Chapter 7). Image by Susmita Sen.

Digital Camera Designs

In the digital realm, there are several alternative strategies for acquiring images, each representing a different strategy for forming digital images that are roughly equivalent to the 9 in. × 9 in. size of analog aerial photographs that became a commonly accepted standard in the United States after the 1930s. Although this physical size offers certain advantages with respect to convenience and standardization during the analog era, there is no technical reason to continue use of this format in the digital era. Indeed, some digital cameras use slightly different sizes. In due course, a new standard or set of standards may well develop as digital systems mature to establish their own conventions.

Practical constraints of forming the large arrays of detectors necessary to approximate this standard size have led to camera designs that differ significantly from those of analog cameras described earlier. Analog cameras captured images frame by frame, meaning that each image was acquired as a single image corresponding to the single image projected into the focal plane at the time the shutter closed. This area, known as the camera *format,* varied in size and shape depending on the design of the camera, although, as mentioned above, a common standard for mapping photography used the 9 in. × 9 in. standard, now defined by its metric equivalent, 230 mm × 230 mm. This photographic frame, acquired at a single instant, forms the fundamental unit for the image; every such image is a frame, a portion of a frame, or a composite of several frames. Such cameras are therefore designated as *framing cameras,* or as *frame array cameras,* which have formed the standard for analog aerial camera designs.

However, the framing camera design does not transfer cleanly into the digital domain. A principal reason for alternative designs for digital cameras is that the use of the traditional 230-mm × 230-mm film format for mapping cameras would require a nearly 660 megapixel array—a size that, currently, is much too large (i.e., too expensive) for most civilian applications. This situation requires some creative solutions for large-format digital cameras. One solution is to use multiple-area CCDs (and thus multiple lens systems) to acquire images of separate quadrants within the frame, and then to digitally stitch the four quadrants together to form a single image. Such composites provide an image that is visually equivalent to that of an analog mapping camera but that will have its own distinctive geometric properties. For example, such an image will have a nadir for each lens that might be used, and its brightness values will be altered when the images are processed to form the composite. Another design solution for a digital aerial camera is to employ linear rather than area arrays. One such design employs a camera with separate lens systems to view (1) the nadir, (2) the forward-looking, and (3) the aft-looking position. At any given instant, the camera views only a few lines at the nadir, at the forward-looking position, and at the aft-viewing position. However, as the aircraft moves forward along its flight track, each lens accumulates a separate set of imagery. These separate images can be digitally assembled to provide complete coverage from several perspectives in a single pass of the aircraft.

The following paragraphs highlight several state-of-the-art examples of digital camera designs, including the large-format frame-based digital modular camera (DMC) from Intergraph, the Vexcel UltraCamX, and the line sensor-based Hexagon/Leica Geosystems ADS100, to illustrate general concepts. By necessity, these descriptions can outline only basic design strategies; readers who require detailed specifications should refer to the manufacturers' complete design specifications to see both the full details of the many design variations and the up-to-date information describing the latest models, which offer many specific applications of the basic strategies outlined here.

Area Arrays: The Intergraph Digital Modular Camera

The Intergraph Digital Modular Camera (DMC; **Figure 4.13**) is a large-format-frame digital camera. It uses four high-resolution panchromatic camera heads (focal length = 120 mm) in the center and four multispectral camera heads (focal length 25 mm) on the periphery. The panchromatic CCD arrays are 7,000 × 4,000 pixels, resulting in a resolution of 3,824 pixels across track and 7,680 pixels along track (Boland et al., 2004). The multispectral arrays are 3,000 × 2,000 pixels, with wavelength ranges as follows: blue (0.40–0.58 µm), green (0.50–0.65 µm), red (0.59–0.675 µm), and near infrared (0.675–0.85 or 0.74–0.85 µm).

Images from the two camera modules are merged into a single frame, without the need for postprocessing, to provide a single-color-image, geometrically and radiometrically correct frame suitable for both photogrammetric and orthophoto applications.

Area Arrays: The Vexcel UltraCamX

The Vexcel UltraCamX employs multiple lens systems and CCDs positioned in the same plane, but with timing of the exposures to offset exposures slightly such that the scene is viewed from the same perspective center. Together, they form a system of eight CCDs: four panchromatic CCDs at fine resolution (which are combined to form a master image) and four multispectral CCDs at coarser resolution to image each frame. The master image provides the spatial framework for a mosaic of the coarser resolution multispectral images. The resulting single multispectral image is rectangular, with the long axis oriented in the across-track dimension. This is different from analog framing cameras, as the multiple lens systems and interpolation of the mosaic affects the optical and radiometric properties (**Figure 4.14**).

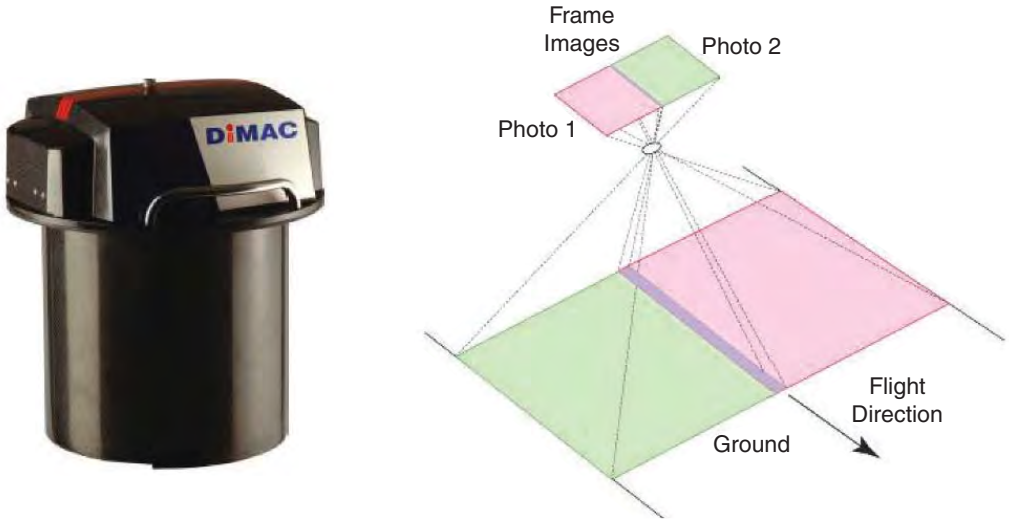


FIGURE 4.13 DiMAC area array. A single composite image is composed of two separate images acquired by independent lens systems with overlapping fields of view. From Intergraph. Used by permission of DiMAC.

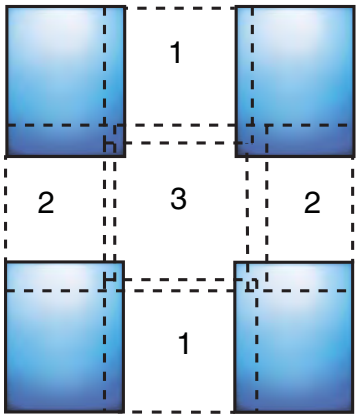


FIGURE 4.14 Schematic representation of a composite aerial image formed from images collected independently from separate areal arrays. Image by Susmita Sen.

Linear Arrays

The Leica ADS100 captures several linear arrays, each oriented to collect imagery line by line, from three orientations: forward-looking, nadir, and aft-looking. Multispectral arrays acquire data in the blue, green, red, and near-infrared regions. Thus, in one pass of the aircraft, using one instrument, it is possible to acquire, for a given region, multispectral imagery from several perspectives. One distinctive feature of this configuration is that the nadir for imagery collected by this system is, in effect, the line connecting the nadirs of each linear array, rather than the center of the image, as is the case for an image collected by a framing camera. Therefore, each image displays relief displacement along track as a function only of object height, whereas in the across-track dimension relief displacement resembles that of a framing camera (i.e., relief displacement is lateral from the nadir). This instrument, like many others, requires high-quality positioning data, and, to date, data from this instrument require processing by software provided by the manufacturer. We can summarize by saying that this linear array solution is elegant and robust and is now used by photogrammetric mapping organizations throughout the world.

4.5 DIGITAL SCANNING OF ANALOG IMAGES

The value of the digital format for digital analysis has led to the scanning of images originally acquired in analog form to create digital versions, which offer advantages for storage, transmission, and analysis. Although the usual scanners designed for office use provide, for casual use, reasonable positional accuracy and preserve much of the detail visible in the original, they are not satisfactory for scientific or photogrammetric applications. Such applications require scanning of original positives or transparencies to preserve the positional accuracy of images and to accurately capture colors and original spatial detail using specialized high-quality flatbed scanners, which provide large scanning surfaces, large CCD arrays, and sophisticated software to preserve information recorded in the original. Although there are obvious merits to scanning archived imagery, scanning of imagery for current applications must be regarded as an improvisation relative to original collection of data in a digital format.

4.6 SPECTRAL SENSITIVITY

Just as analog cameras used color films to capture the spectral character of a scene, so can detectors be configured to record separate regions of the spectrum as separate bands, or channels. CCDs and CMOS arrays have sensitivities determined by the physical properties of the materials used to construct sensor chips and the details of their manufacture. The usual digital sensors have spectral sensitivities that encompass the visible spectrum (with a maximum in the green region) and extend into the near infrared. Although arrays used for consumer electronics specifically filter to exclude the near-infrared region (NIR), aerial cameras can use this sensitivity to good advantage.

Color films use emulsions that are sensitive over a range of wavelengths, so even if their maximum sensitivity lies in the red, green, or blue regions, they are sensitive to radiation beyond the desired limits. In contrast, digital sensors can be designed to have spectral sensitivities cleanly focused in a narrow range of wavelengths and to provide high precision in measurement of color brightness. Therefore, digital sensors provide better records of the spectral characteristics of a scene—a quality that is highly valued by some users of aerial imagery. If the sensor chip is designed as separate arrays for each region of the spectrum, it acquires color images as separate planar arrays for each region of the spectrum. Although such designs would be desirable, they are not practical for current aerial cameras. Such large-array sizes are extremely expensive and difficult to manufacture to the required specifications, and they may require long readout times to retrieve the image data. In due course, these costs will decline, and the manufacturing and technical issues will improve. In the meantime, aerial cameras use alternative strategies for simulating the effect of planar color data.

One alternative strategy uses a single array to acquire data in the three primaries using a specialized filter, known as a *Bayer filter,* to select the wavelengths that reach each pixel. A Bayer filter is specifically designed to allocate 50% of the pixels in an array to receive the green primary and 25% each to the red and blue primaries (**Figure 4.15**). (The rationale is that the human visual system has higher sensitivity in the green region, and, as mentioned in Chapter 2, peak radiation in the visible region lies in the green region.) In effect, this pattern samples the distribution of colors within the image, and then the CCD chip processes pixel values to extrapolate, or interpolate, to estimate the missing values for the omitted pixels for each color. For example, the complete blue layer is formed by using the blue pixels to estimate the blue brightness values omitted in the array, and similarly for the red and green primaries. This basic strategy has been

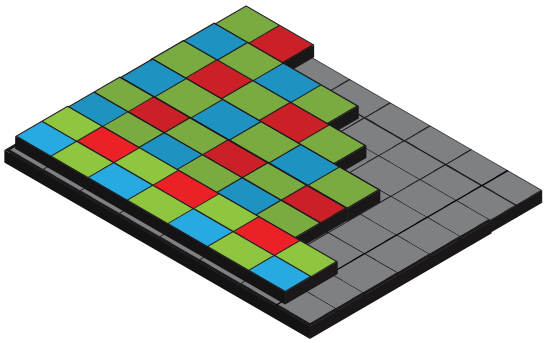


FIGURE 4.15 Bayer filter. Positioned at the focal plane of an imaging system, the blue, green, and red colors each signify pixels with blue, green, and red filters. Pixels for each primary are separately interpolated to produce individual layers for each primary, which then permit formation of digital color imagery. Image by Susmita Sen.

implemented in several variations of the Bayer filter that have been optimized for various applications, and variations on the approach have been used to design cameras sensitive to the near-infrared region.

This strategy, widely used in consumer electronics, produces an image that is satisfactory for visual examination because of the high density of detectors relative to the patterns recorded in most images and the short distances of the interpolation. However, this approach is less satisfactory for scientific applications and for aerial imagery—contexts in which the sharpness and integrity of each pixel may be paramount and artifacts of the interpolation process may be significant. Furthermore, the Bayer filter has the disadvantages that the color filters reduce the amount of energy reaching the sensor and that interpolation required to construct the several bands reduces image sharpness.

An alternative strategy—*Foveon technology*—avoids these difficulties by exploiting the differential ability of the sensor's silicon construction to absorb light. Foveon detectors (patented as the *X3 CMOS* design) are designed as three separate detector layers encased in silicon: blue-sensitive detectors at the surface, green-sensitive below, and red-sensitive detectors below the green. As light strikes the surface of the detector, blue light is absorbed near the chip's surface, green penetrates below the surface, and red radiation below the green. Thus, each pixel can be represented by a single point that portrays all three primaries without the use of filters. This design has been employed for consumer cameras and may well find a role in aerial systems. At present, however, there are concerns that colors captured deeper in the chip may receive weaker intensities of the radiation and may have higher noise levels.

4.7 BAND COMBINATIONS: OPTICAL IMAGERY

Effective display of an image is critical for effective practice of remote sensing. *Band combinations* is the term that remote sensing practitioners use to refer to the assignment of colors to represent brightness values in different regions of the spectrum. Although there are many ways to assign colors to represent different regions of the spectrum, experience shows that some are proven to be more useful than others. A key constraint for the display of any multispectral image is that human vision is sensitive only to the three additive primaries: blue, green, and red. Because our eyes can distinguish between brightness values in these spectral regions, we can distinguish not only between blue, green, and red surfaces but also between intermediate mixtures of the primaries, such as yellow, orange, and purple. Color films and digital displays portray the effect of color by displaying pixels that vary the mixtures of the blue, green, and red primaries. Although photographic films must employ a single, fixed strategy for portraying colors, image processing systems and digital displays offer the flexibility to use any of many alternative strategies for assigning colors to represent different regions of the spectrum. These alternative choices define the band selection task; that is, they show how to decide which primary colors to select to best portray on the display screen specific features represented on imagery.

If the imagery at hand is limited to three spectral regions (as is the case for normal everyday color imagery), then the band selection task is simple: that is, display radiation from blue objects in nature as blue on the screen, green as green, and red as red. However, once we have bands available from outside the visible spectrum, as is common for remotely sensed imagery, then the choice of color assignment must have an arbitrary dimension. For example, there can be no logical choice for the primary we might use to

display energy from the near-infrared region. The common choices for the band selection problem, then, are established in part by conventions that have been defined by accepted use over the decades and in part by practice that has demonstrated that certain combinations are effective for certain purposes.

In the following sections, we introduce the band combinations most common for optical aerial imagery.

Black-and-White Infrared Imagery

Imagery acquired in the near-infrared region, because it is largely free of the effects of atmospheric scattering, shows vegetated regions and land–water distinctions. It is one of the most valuable regions of the spectrum (**Figure 4.16**; see also **Figure 4.20**, later). An image representing the near infrared is formed using an optical sensor that has filtered the visible portion of the spectrum, so the image is prepared using only the brightness values of the near-infrared region (**Figure 4.16**). Examples are presented in **Figures 4.18** (right) and later in **Figure 4.20** (right).

Panchromatic Imagery

Panchromatic means "across the colors," indicating that the visible spectrum is represented as a single channel (without distinguishing between the three primary colors). A panchromatic view provides a black-and-white image that records the brightness values using radiation from the visible region but without separating the different colors (**Figure 4.17**; **Figure 4.18**, left). (This model is sometimes designated by the abbreviation PAN.) Digital remote sensing systems often employ a panchromatic band that substitutes spatial detail for a color representation; that is, the instrument is designed to capture a detailed version of the scene using the data capacity that might have been devoted to recording the three primaries. That is, a decision has been made that the added detail provides more valuable information than would a color representation.

Because the easily scattered blue radiation will degrade the quality of an aerial image, some instruments are designed to capture radiation across the green, red, and near-infrared regions of the spectrum, thereby providing a sharper, clearer image than would otherwise be the case. Therefore, even though the traditional definition of a panchromatic image is restricted to those based only on visible radiation, the term has a long history of use within the field of remote sensing to designate a broader region extending into the near infrared. If the single band encompasses the entire visible and the NIR, it

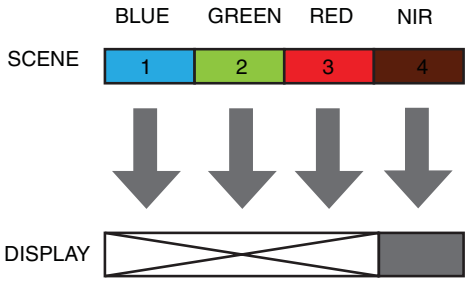


FIGURE 4.16 Diagram representing black-and-white infrared imagery. Visible radiation is filtered to isolate the near-infrared imagery use to form the image. Image by Susmita Sen.

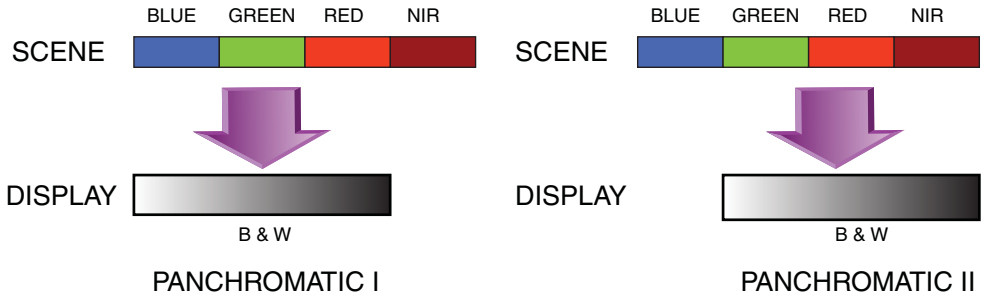


FIGURE 4.17 Two forms of panchromatic imagery that provide an image using a single band spanning the visible spectrum. Left: Use of full visible spectrum. Right: An alternative strategy using red, green, and NIR, omitting the blue region, thereby screening out effects of atmospheric scattering in the blue region. Image by Susmita Sen.

can sometimes be designated as VNIR, signifying the use of visible radiation and the NIR region together. Other versions of this approach use only the green, red, and NIR, as illustrated in Figure 4.17 (right). For many applications, panchromatic aerial imagery is completely satisfactory, especially for imagery of urban regions in which color information may not be essential and added spatial detail is especially valuable.

Natural-Color Model

In everyday experience, our visual system applies band combinations in what seems to be a totally obvious manner: we see blue as blue, green as green, and red as red. The usual color films, color displays, and television screens apply this same strategy for assigning colors, often known as the *natural-color* assignment model (Figure 4.19), or sometimes

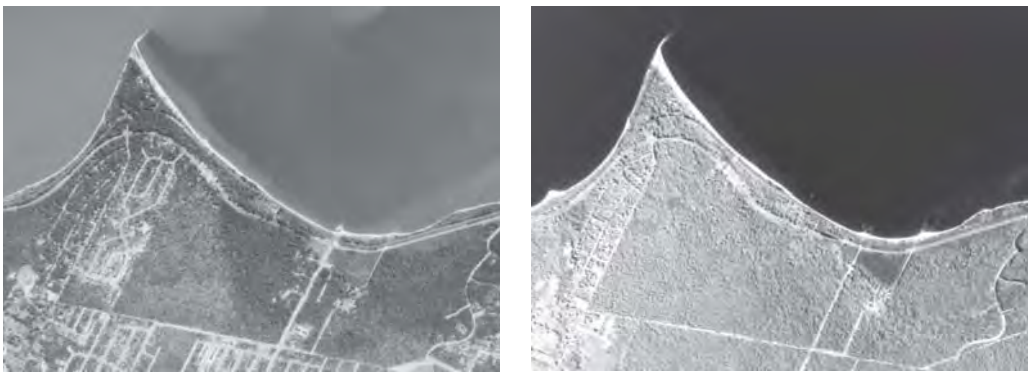


FIGURE 4.18 Panchromatic (left) and black-and-white infrared imagery (right). Notice the effect of black-and-white infrared imagery (right) in depicting the water surface as a jet-black surface, with a sharply defined edge at the shoreline, and its clear separation of deciduous and evergreen forest canopies. Panchromatic imagery reveals the sediment transport in coastal waters and is effective in delineating street patterns. From USGS, Earth Resources Observation and Science Center.

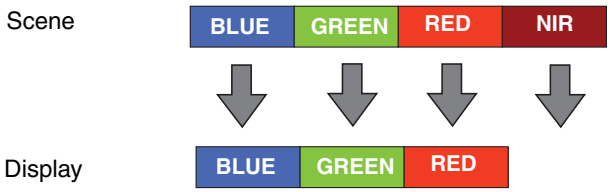


FIGURE 4.19 Natural-color model for color assignment. The camera captures the three primaries but does not record radiation outside the visible spectrum. Image by Susmita Sen.

as the RGB (i.e., red–green–blue) model. Although natural-color imagery has value for its familiar representation of a scene, it suffers from a disadvantage (outlined in Chapter 2) that the blue region of the spectrum is subject to atmospheric scattering, thereby limiting the utility of natural-color images acquired at high altitudes compared to other wavelengths (Figure 4.20).

Although remote sensing instruments collect radiation across many regions of the spectrum, outside the visible region we are limited by our visual system to perceive only the blue, green, and red primaries. Because our visual system is sensitive only in the visible region and can use only the three primaries in remote sensing, we must make color assignments that depart from the natural-color model. These create *false-color* images—false in the sense that the colors on the image do not match their true colors in nature. Analysts select specific combinations of three channels to represent those patterns on the imagery needed to attain specific objectives. When some students first encounter this concept, it often seems nonsensical to represent an object using any color other than its natural color. Because the field of remote sensing uses radiation outside the visible spectrum, use of the false-color model is a necessary task in displaying remote sensed imagery. The assignment of colors in this context is arbitrary, as there can be no correct way to represent the appearance of radiation outside the visible spectrum but simply a collection of practices that have proven to be effective for certain purposes.

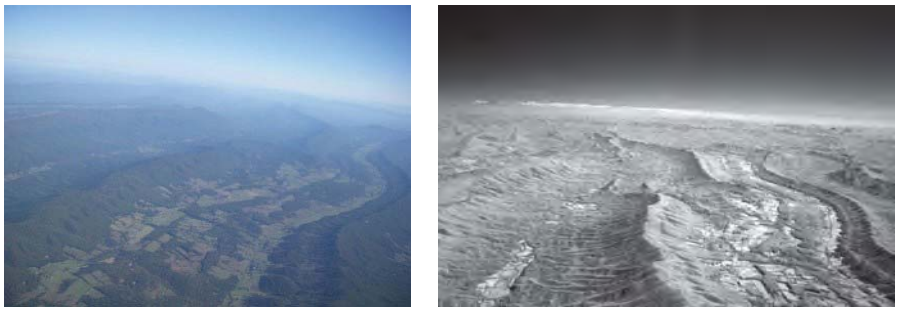


FIGURE 4.20 Aerial views—color (left) and infrared (right)—of landscapes. Left: Normal color aerial photograph of western Virginia ridges and valleys, with far horizon obscured by atmospheric haze. Right: The same region photographed, at the same time, as a black-and-white infrared image, with the far horizon viewed clearly, illustrating the ability of infrared radiation to pass through the atmosphere with minimal attenuation.

Color-Infrared Model

One of the most valuable regions of the spectrum is the NIR region, characterized by wavelengths that are just longer than the longest region of the visible spectrum. This region carries important information about vegetation and is not subject to atmospheric scattering, so it is a valuable adjunct to the visible region. Use of the NIR region adds a fourth spectral channel to the natural-color model. Because we can recognize only three primaries, adding an NIR channel requires omission of one of the visible bands.

The color-infrared (CIR) model (**Figures 4.21 and 4.22**) creates a three-band color image by discarding the blue band from the visible spectrum and adding a channel in the NIR. This widely used model was implemented in color-infrared films that were initially developed in World War II as *camouflage detection film* (i.e., designed to use NIR radiation to detect differences between actual vegetation and surfaces painted to resemble vegetation to the eye), and were later developed as *CIR film*, now commonly used for displays of digital imagery. It shows living vegetation and water bodies very clearly and greatly reduces atmospheric effects compared with the natural-color model. It is therefore very useful for high-altitude aerial photography, which otherwise is subject to atmospheric effects that degrade the image. This band combination is important for studies in agriculture, forestry, and water resources, to list only a few of many subjects.

Later chapters extend this discussion of band selection beyond those bands that apply primarily to aerial photography to include spectral channels acquired by other sensors.

4.8 COVERAGE BY MULTIPLE PHOTOGRAPHS

A flight plan usually calls for acquisition of vertical aerial photographs by flying a series of parallel flight lines that together build up complete coverage of a specific region. For framing cameras, each flight line consists of individual frames, usually numbered in sequence (**Figure 4.23**). The camera operator can view the area to be photographed through a viewfinder and can manually trigger the shutter as aircraft motion brings predesignated landmarks into the field of view, or the operator can set controls to automatically acquire photographs at intervals tailored to provide the desired coverage. Individual frames form

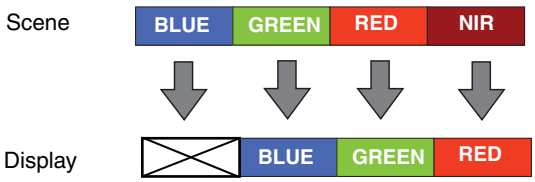


FIGURE 4.21 Color infrared (CIR) model for color assignment. Blue light (subject to high levels of scattering) is blocked from the image. Green and red radiation are recorded on the image but assigned to their adjacent colors (i.e., green radiation is displayed as blue; red light is displayed as green). NIR, outside the visible spectrum, is displayed as red. NIR is largely free of scattering and so is valued for its clarity, its sensitivity to living vegetation cover, and its ability to clearly define the edges of water bodies. See **Figure 4.22** for a comparison between a natural-color image and the same scene as a CIR image. Image by Susmita Sen.

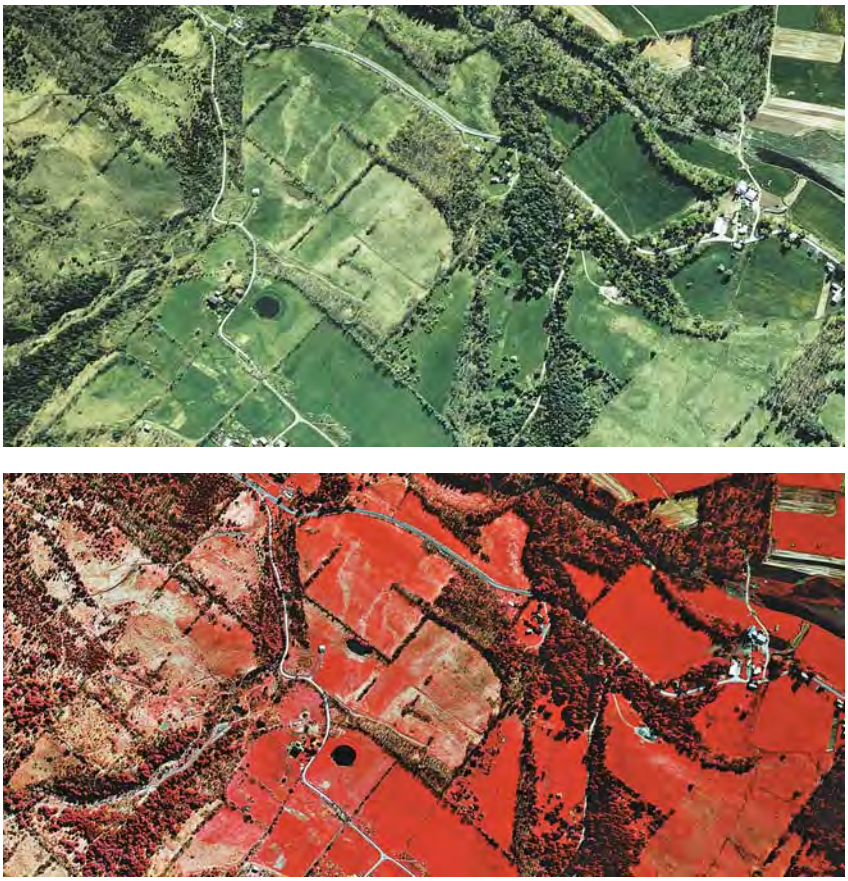


FIGURE 4.22 Color and CIR aerial photographs. Top: Torch Lake, Michigan, landscape imaged as a natural-color aerial photograph. Bottom: CIR aerial photograph of the same region. From U.S. Environmental Protection Agency.

ordered strips, as shown in Figure 4.23a. If the plane's course is deflected by a crosswind, the positions of ground areas shown by successive photographs form the pattern shown in Figure 4.23b, known as *drift*. *Crab* (Figure 4.23c) is caused by correction of the flight path to compensate for drift without a change in the orientation of the camera.

Usually, flight plans call for a certain amount of *forward overlap* (Figure 4.24) in order to duplicate coverage by successive frames in a flight line, usually by about 50–60% of each frame. If forward overlap is 50% or more, then the image of the principal point of one photograph is visible on the next photograph in the flight line. These are known as *conjugate principal points* (Figure 4.24). When it is necessary to photograph large areas, coverage is built up by means of several parallel strips of photography; each strip is called a *flight line*. Sidelap between adjacent flight lines may vary from about 5 to 15%, in an effort to prevent gaps in coverage of adjacent flight lines.

However, as pilots collect complete photographic coverage of a region, there may still be gaps (known as *holidays*) in coverage due to equipment malfunction, navigation errors, or cloud cover. Sometimes photography acquired later to cover holidays differs noticeably from adjacent images with respect to Sun angle, vegetative cover, and other

qualities. For planning flight lines, the number of photographs required for each line can be estimated using the relationship:

$$\text{Number of photos} = \frac{\text{Length of flight line}}{(gd \text{ of photo}) \times (1 - \text{overlap})} \quad \text{(EQ. 4.2)}$$

where *gd* is the ground distance represented on a single frame, measured in the same units as the length of the planned flight line. For example, if a flight line is planned to be 33 mi in length, if each photograph is planned to represent 3.4 mi on a side, and if forward overlap is to be 0.60, then 33/ [3.4 × (1 – .60)] = 33/(1.36) = 24.26; about 25 photographs are required. (Chapter 6 shows how to calculate the coverage of a photograph for a given negative size, focal length, and flying altitude.)

Stereoscopic Parallax

If we have two photographs of the same area taken from different perspectives (i.e., from different camera positions), we observe a displacement of images of objects from one image to the other (to be discussed further in Chapter 6). The reader can observe this effect now by simple observation of nearby objects. Look up from this book at nearby objects. Close one eye, then open it and close the other. As you do this, you observe a change in the appearance of objects from one eye to the next. Nearby objects are slightly different in appearance because one eye tends to see, for example, only the front of an

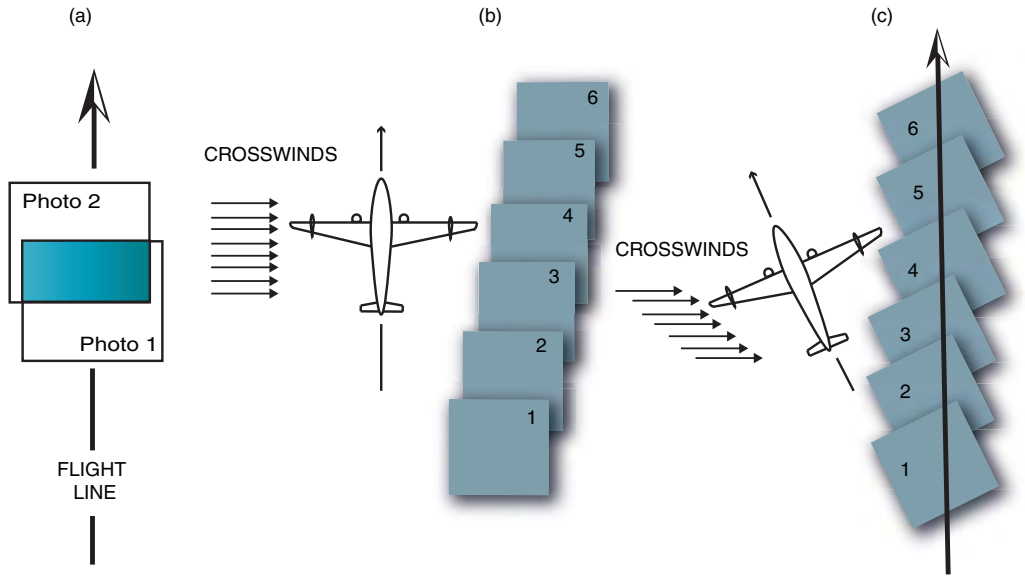


FIGURE 4.23 Aerial photographic coverage for framing cameras. (a) *Forward overlap* (offset to illustrate overlap), successive frames deliberately offset to ensure continuity, or to acquire stereo images (see Chapter 6). (b) *Drift*, aircraft flight path altered by transverse wind without correction by pilot). (c) *Crab,* indicating that the pilot has adjusted the aircraft flight path for a crosswind, but the camera operator has not adjusted the orientation of the camera to compensate for the change in orientation. Image by Susmita Sen.

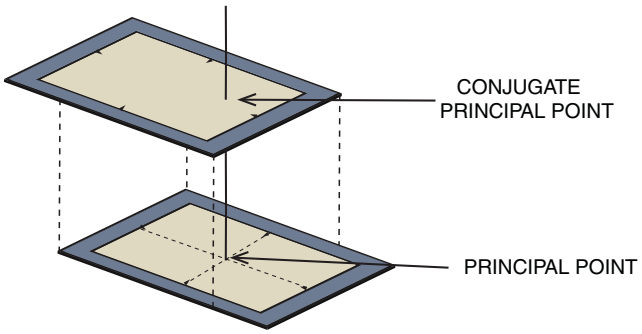


FIGURE 4.24 Forward overlap and conjugate principal points. *Forward overlap* means that a photograph is acquired to duplicate coverage of the previous photograph, usually by about 60%. Overlap ensures that there are no gaps between adjacent frames and provides the capability for stereoscopic analysis. A *conjugate principal point* is the plot of a previous principal point to the adjacent image. Conjugate principal points facilitate the alignment of photos in a flight line and detection of errors in planned flight lines. Image by Susmita Sen.

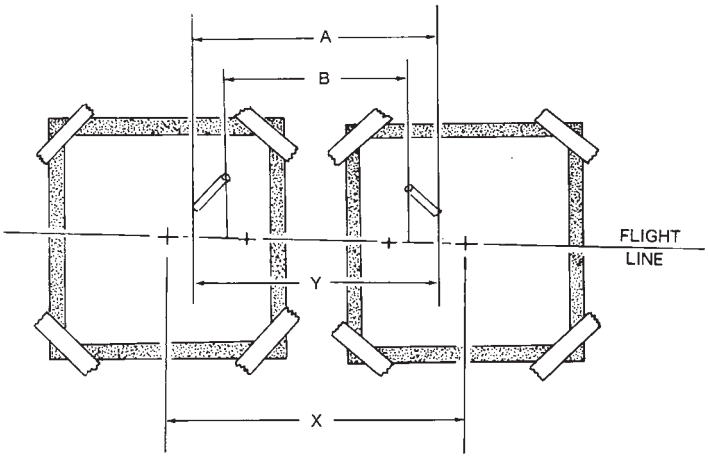
object, whereas the other, because of its position about 2.5 in. from the other, sees the front and some of the side of the same object. This difference in the appearance of objects due to change in perspective is known as *stereoscopic parallax*. The amount of parallax decreases as objects increase in distance from the observer (Figure 4.25). If you repeat the experiment looking out the window at a landscape, you can confirm this effect by noting that distant objects display little or no observable parallax.

Stereoscopic parallax can therefore be used as a basis for measuring distance or height. Overlapping aerial photographs record parallax due to the shift in position of the camera as aircraft motion carries the camera forward between successive exposures. If forward overlap is 50% or more, then the entire ground area shown on a given frame can be viewed in stereo using three adjacent frames (a *stereo triplet*). Forward overlap of 50–60% is common. This amount of overlap doubles the number of photographs required but ensures that the entire area can be viewed in stereo because each point on the ground will appear on two successive photographs in a flight line.

Displacement due to stereo parallax is always parallel to the flight line. Tops of tall objects nearer to the camera show more displacement than do shorter objects, which are more distant from the camera. Measurement of parallax therefore provides a means of estimating the heights of objects. Manual measurement of parallax can be accomplished as follows. Tape photographs of a stereo pair to a work table so that the axis of the flight line is oriented from right to left (Figure 4.26). For demonstration purposes, distances can be measured with an engineer's scale.

1. Measure the distance between two principal points (X).
2. Measure the distance between separate images of the base of the object as represented on the two images (Y). Subtract this distance from that found in (1) to get P.
3. Measure top-to-top distances (B) and base-to-base distances (A), then subtract to find dp.

FIGURE 4.25 Stereoscopic parallax. These two photographs of the same scene were taken from slightly different positions. Note the difference in the appearance of objects due to differences caused by differences in perspectives. Note also that differences are greatest for objects nearest to the camera and least for the object in the distance.



$$h = \frac{H \times dp}{P + dp}$$

H = FLYING HEIGHT OF AIRCRAFT

dp = DIFFERENTIAL PARALLAX = A - B

P = X - Y

h = HEIGHT OF THE OBJECT

FIGURE 4.26 Measurement of stereoscopic parallax. Distance "X" is the distance between principal points of two overlapping vertical aerial photos. "A" and "B" measure distances between bases and the tops (of an idealized tower).

In practice, parallax measurements can be made more conveniently using devices that permit accurate measurement of small amounts of parallax.

Orthophotos and Orthophotomaps

Aerial photographs are not planimetric maps because they have geometric errors, most notably the effects of tilts and relief displacement, in the representations of the features they show. That is, objects are not represented in their correct planimetric positions, and as a result the images cannot be used as the basis for accurate measurements.

Stereoscopic photographs and terrain data can be used to generate a corrected form of an aerial photograph known as an *orthophoto* that shows photographic detail without the errors caused by tilt and relief displacement. During the 1970s, an optical–mechanical instrument known as an *orthophotoscope* was developed to optically project a corrected version of a very small portion of an aerial photograph. An orthophotoscope, instead of exposing an entire image from a central perspective (i.e., through a single lens), systematically exposes a small section of an image individually in a manner that corrects for the elevation of that small section. The result is an image that has orthographic properties rather than those of the central perspective of the original aerial photograph. Digital versions of the orthophotoscope, developed in the mid-1980s, are capable of scanning an entire image piece by piece to generate a corrected version of that image. The result is an image that shows the same detail as the original aerial photograph but without the geometric errors introduced by tilt and relief displacement. Orthophotos form the basis for *orthophotomaps,* which show the image in its correct planimetric form, together with place names, symbols, and geographic coordinates. Thus, they form digital map products that can be used in GIS as well as traditional maps because they show correct planimetric position and preserve consistent scale throughout the image.

Orthophotomaps are valuable because they show the fine detail of an aerial photograph without the geometric errors that are normally present and because they can be compiled much more quickly and cheaply than the usual topographic maps. Therefore, they can be useful as map substitutes in instances in which topographic maps are not available or as map supplements when maps are available, but the analyst requires the finer detail and more recent information provided by an image. Because of their digital format, fine detail, and adherence to national map accuracy standards, orthophotomaps are routinely used in GIS.

Digital Orthophoto Quadrangles

Digital orthophoto quadrangles (DOQs) are orthophotos prepared in a digital format designed to correspond to the 7.5-minute quadrangles of the U.S. Geological Survey (USGS). DOQs are presented either as black-and-white or color images that have been processed to attain the geometric properties of a planimetric map.

DOQs are prepared from National Aerial Photography Program (NAPP) photography (high-altitude photography described in Section 4.10) at scales of 1:40,000, supplemented by other aerial photography as needed. The rectification process is based on the use of digital elevation models (DEMs) to represent variations in terrain elevation. The final product is presented (as either panchromatic or CIR imagery) to correspond to the matching USGS 7.5-minute quadrangle, with a supplementary border of imagery representing 50–300 m beyond the limits of the quadrangle, to facilitate matching

and mosaicking with adjacent sheets. A related product, the *digital orthophoto quarter-quadrangle* (DOQQ), formatted to provide a more convenient unit, represents one-fourth of the area of a DOQ at a finer level of detail and is available for some areas (**Figure 4.27**). DOQs provide image detail equivalent to 2 m or so when presented in the quadrangle format and finer detail (about 1 m) for DOQQs. The USGS has responsibility for leading the U.S. federal government's effort to prepare and disseminate digital cartographic data. The USGS has a program to prepare DOQs for many regions of the United States, especially urbanized regions, and the U.S. Department of Agriculture supports preparation of DOQs for agricultural regions (Section 4.10).

For more information on DOQs, visit the USGS website at:

www.usgs.gov/centers/eros

4.9 PHOTOGRAMMETRY

Photogrammetry is the science of making accurate measurements from photographs. Photogrammetry applies the principles of optics and knowledge of the interior geometry of the camera and its orientation to reconstruct dimensions and positions of objects represented in photographs. Therefore, its practice requires detailed knowledge of specific cameras and the circumstances under which they were used, as well as accurate measurements of features in photographs. Photographs used for analog photogrammetry have traditionally been prepared on glass plates or other dimensionally stable materials (i.e., materials that do not change in size as temperature and humidity change).

Photogrammetry can be applied to any photograph, provided the necessary supporting information is at hand to reconstruct the optical geometry of the image. However, by far the most frequent application of photogrammetry is the analysis of stereo aerial photography to derive estimates of topographic elevation for topographic mapping. With the aid of accurate locational information describing key features within a scene (*ground control*), photogrammetrists estimate topographic relief using stereo parallax for an array of points within a region. Although stereo parallax can be mea-

FIGURE 4.27 Digital orthophoto quarter quad, Platte River, Nebraska. From USGS.

sured manually, it is far more practical to employ specialized instruments designed for stereoscopic analysis.

Initially, such instruments, known as *analytical stereoplotters,* first designed in the 1920s, reconstruct the orientations of photographs at the time they were taken using optical and mechanical instruments to reconstruct the geometry of the images at the time they were acquired (see **Figure 1.8** for an example of an optical–mechanical photogrammetric instrument). Operators could then view the image in stereo; by maintaining constant parallax visually, they could trace lines of uniform elevation. The quality of information derived from such instruments depends on the quality of the photography, the accuracy of the data, and the operator's skill in setting up the stereo model and tracing lines of uniform parallax. As the design of instruments improved, it eventually became possible to automatically match corresponding points on stereo pairs and thereby identify lines of uniform parallax, with limited assistance from the operator.

With further advances in instrumentation, it became possible to extend automation of the photogrammetric process to conduct the stereo analysis completely within the digital domain. With the use of GPS (*airborne global positioning systems [AGPS]*) to acquire accurate, real-time positional information and the use of data recorded from the aircraft's navigational system (*inertial navigational systems [INS]*) to record orientations of photographs, it then became feasible to reconstruct the geometry of the image using precise positional and orientation data gathered as the image was acquired. This process forms the basis for *softcopy photogrammetry,* so named because it does not require the physical (hardcopy) form of the photograph necessary for traditional photogrammetry. Instead, the digital (softcopy) version of the image is used as input for a series of mathematical models that reconstruct the orientation of each image to create planimetrically correct representations. This process requires specialized computer software installed in workstations (see **Figure 6.19**) that analyzes digital data specifically acquired for the purpose of photogrammetric analysis. Softcopy photogrammetry, now the standard for photogrammetric production, offers the advantages of speed and accuracy and generates output data that are easily integrated into other production and analytical systems, including GIS.

The application of photogrammetric principles to imagery collected by digital cameras described above differs from that tailored for the traditional analog framing camera. Because each manufacturer has specific designs, each applying a different strategy for collecting and processing imagery, the current photogrammetric analyses are matched to the differing cameras. One characteristic common to many of these imaging systems is the considerable redundancy within imagery they collect. That is, each pixel on the ground can be viewed many times, each from a separate perspective. Because these systems each collect so many independent views of the same features (due to the use of several lenses, or several linear arrays, as outlined previously), it is possible to apply *multiray photogrammetry,* which can exploit these redundancies to extract highly detailed positional and elevation data beyond that which was possible using analog photography. Because, in the digital domain, these additional views do not incur significant additional costs, photogrammetric firms can provide high detail and a wide range of image products without the increased costs of acquiring additional data.

4.10 SOURCES OF AERIAL PHOTOGRAPHY

Aerial photography can be (1) acquired by the user or (2) purchased from organizations that serve as repositories for imagery flown by others (*archival imagery*). In the first instance, aerial photography can be acquired by contract with firms that specialize in high-quality aerial photography. Such firms are listed in the business sections of most metropolitan phone directories. Customers may be individuals, governmental agencies, or other businesses that use aerial photography. Such photography is, of course, customized to meet the specific needs of customers with respect to date, scale, film, and coverage. As a result, costs may be prohibitive for many noncommercial uses.

Thus, for pragmatic reasons, many users of aerial photography turn to archival photography to acquire the images they need. Although such photographs may not exactly match users' specifications with respect to scale or date, the inexpensive costs and ease of access may compensate for any shortcomings. For some tasks that require reconstruction of conditions at earlier dates (such as the Environmental Protection Agency's search for abandoned toxic waste dumps), the archival images may form the only source of information (e.g., Erb et al., 1981; Lyon, 1987).

It is feasible to take "do-it-yourself" aerial photographs. Many handheld cameras are suitable for aerial photography. Often, the costs of local air charter services for an hour or so of flight time are relatively low. Small-format cameras, such as the usual 35-mm cameras, can be used for aerial photography if the photographer avoids the effects of aircraft vibration. (Do not rest the camera against the aircraft!) A high-wing aircraft offers the photographer a clear view of the landscape, although some low-wing aircraft are satisfactory. The most favorable lighting occurs when the camera is aimed away from the Sun. Photographs acquired in this manner (e.g., Figure 4.5) may be useful for illustrative purposes, although for scientific or professional work the large-format, high-quality work of a specialist or an aerial survey firm may be required.

EROS Data Center

The EROS Data Center (EDC) in Sioux Falls, South Dakota, is operated by the USGS as a repository for aerial photographs and satellite images acquired by NASA, the USGS, and many other federal agencies. A computerized database at EDC provides an indexing system for information pertaining to aerial photographs and satellite images. For more information contact:

Customer Services
U.S. Geological Survey
Earth Resources Observation and Science (EROS)
47914 252nd Street
Sioux Falls, SD 57198–0001
Tel: 800–252–4547
E-mail: *custserv@usgs.gov*
Website: *www.usgs.gov/centers/eros*

Earth Science Information Centers

The Earth Science Information Centers (ESIC) are operated by the USGS as a central source for information pertaining to maps and aerial photographs. ESIC has a special interest in information pertaining to federal programs and agencies but also collects data pertaining to maps and photographs held by state and local governments. The ESIC headquarters is located at Reston, Virginia, but ESIC also maintains seven other offices throughout the United States, and other federal agencies have affiliated offices. ESIC can provide information to the public concerning the availability of maps and remotely sensed images. The following sections describe two programs administered by ESIC that can provide access to archival aerial photography.

National Aerial Photography Program

NAPP acquires aerial photography for the coterminous United States, according to a systematic plan that ensures uniform standards. This program was initiated in 1987 by the USGS as a replacement for the National High-Altitude Aerial Photography Program (NHAP), which began in 1980 to consolidate the many federal programs that use aerial photography. The USGS manages NAPP, but it is funded by the federal agencies that are the primary users of its photography. Program oversight is provided by a committee of representatives from the USGS, the Bureau of Land Management, the National Agricultural Statistics Service, the National Resources Conservation Service (NRCS; previously known as the Soil Conservation Service), the Farm Services Agency (previously known as the Agricultural Stabilization and Conservation Service), the U.S. Forest Service, and the Tennessee Valley Authority. Light (1993) and Plasker and TeSelle (1988) provide further details.

Under NHAP, photography was acquired under a plan first to obtain complete coverage of the coterminous 48 states and then to update coverage as necessary to keep pace with requirements for current photography. Current plans call for updates at intervals of 5 years, although the actual schedules are determined in coordination with budgetary constraints. NHAP flight lines were oriented north–south, centered on each of four quadrants systematically positioned within USGS 7.5-minute quadrangles, with full stereoscopic coverage at 60% forward overlap and sidelap of at least 27%. Two camera systems were used to acquire simultaneous coverage: black-and-white coverage was acquired at scales of about 1:80,000, using cameras with focal lengths of 6 in. Color infrared coverage was acquired at 1:58,000, using a focal length of 8.25 in. **Figure 4.27** shows a high-altitude CIR image illustrating the broad-scale coverage provided by this format.

Dates of NHAP photography varied according to geographic region. Flights were timed to provide optimum atmospheric conditions for photography and to meet specifications for Sun angle, snow cover, and shadowing, with preference for autumn and winter seasons to provide images that show the landscape without the cover of deciduous vegetation.

Specifications for NAPP photographs differ from those of NHAP. NAPP photographs are acquired at 20,000-ft altitude using a 6-in. focal length lens. Flight lines are centered on quarter quads (1:24,000-scale USGS quadrangles). NAPP photographs are planned for 1:40,000, black-and-white or color infrared film, depending on specific requirements for each area.

Photographs are available to all who may have an interest in their use. Their detail and quality permit use for land cover surveys and assessment of agricultural, mineral,

and forest resources, as well as examination of patterns of soil erosion and water quality. Further information is available at:

www.usgs.gov/centers/eros/science/usgs-eros-archive-aerial-photography-national-aerial-photography-program-napp

National Agricultural Imagery Program

The National Agriculture Imagery Program (NAIP) acquires aerial imagery during the agricultural growing seasons in the continental United States. The NAIP program focuses on providing digital orthophotography freely to governmental agencies and the public, usually as color or CIR imagery at about 1-m resolution. The DOQQ format means that the images are provided in a ready-to-use format (i.e., digital and georeferenced). An important difference between NAIP imagery and other programs (such as NHAP) is that NAIP imagery is acquired during the growing season (i.e., "leaf-on"), so it forms a valuable resource not only for agricultural applications but also for broader planning and resources assessment efforts. Further information is available at:

www.fsa.usda.gov/programs-and-services/aerial-photography/imagery-programs/naip-imagery

Two other important sources of archival aerial photography include the U.S. Department of Agriculture (USDA) Aerial Photography Field Office:

www.fsa.usda.gov/programs-and-services/aerial-photography

and the U.S. National Archives and Records Administration:

www.archives.gov

4.11 SUMMARY

Aerial photography offers a simple, reliable, flexible, and inexpensive means of acquiring remotely sensed images. The transition from the analog systems that formed the foundation for aerial survey in the 20th century to digital systems is now basically complete, although the nature of the digital systems that will form the basis for the field in the 21st century is not yet clear. The migration to digital formats has reconstituted, even rejuvenated aerial imagery's role in providing imagery for state and local applications. Although aerial photography is useful mainly in the visible and near-infrared portions of the spectrum, it applies optical and photogrammetric principles that are important throughout the field of remote sensing.

Aerial photographs form the primary source of information for compilation of large-scale maps, especially large-scale topographic maps. Vertical aerial photographs are valuable as map substitutes or as map supplements. Geometric errors in the representation of location prevent direct use of aerial photographs as the basis for measurement of distance or area. But as these errors are known and are well understood, it is possible for photogrammetrists to use photographs as the basis for reconstruction of correct positional rela-

tionships and the derivation of accurate measurements. Aerial photographs record complex detail of the varied patterns that constitute any landscape. Each image interpreter must develop the skills and knowledge necessary to resolve these patterns by disciplined examination of aerial images.

SOME TEACHING AND LEARNING RESOURCES

- Additive Color vs Subtractive Color
  www.youtube.com/watch?v=ygUchcpRNyk&feature=related
- What Are CMYK and RGB Color Modes?
  www.youtube.com/watch?v=0K8fqf2XBaY&feature=related
- Evolution of Analog-to-Digital Mapping
  www.youtube.com/watch?v=4jABMysbNbc
- Aerial Survey Photography Loch Ness Scotland G-BKVT
  www.youtube.com/watch?v=-YsDflbXMHk
- Video of the Day; Aerial Photography
  www.youtube.com/watch?v=VwtSTvF_Q2Q&NR=1
- How a Pixel Gets Its Color; Bayer Sensor; Digital Image
  www.youtube.com/watch?v=2-stCNB8jT8
- Photography Equipment and Info: Explanation of Camera Lens Magnification
  www.youtube.com/watch?v=YEG93Hp3y4w&feature=fvst
- Digital Camera Tips: How a Compact Digital Camera Works
  www.youtube.com/watch?v=eyyMu8UEAVc&NR=1
- Aero Triangulation
  www.youtube.com/watch?v=88KFAU6I_jg

REVIEW QUESTIONS

1. List several reasons why time of day might be very important in flight planning for aerial imagery.
2. Outline the advantages and disadvantages of high-altitude photography. Explain why routine high-altitude aerial photography was not practical before infrared imagery was available.
3. List several problems that you would encounter in acquiring and interpreting large-scale aerial imagery of a mountainous region.
4. Speculate on the likely progress of aerial photography since 1890 if George Eastman had not been successful in popularizing the practice of photography among the general public.
5. Should an aerial photograph be considered a “map”? Explain.
6. Assume you have recently accepted a position as an employee of an aerial survey company; your responsibilities include preparation of flight plans for the company’s customers. What factors must you consider as you plan each mission?

7. List some of the factors you would consider in selection of band combinations described in this chapter.
8. Suggest circumstances in which oblique aerial photography might be more useful than vertical photographs.
9. It might seem that large-scale aerial images might always be more useful than small-scale aerial photographs, yet larger scale images are not always the most useful. What are the disadvantages of using large-scale images?
10. A particular object will not always appear the same when imaged by an aerial camera. List some of the factors that can cause the appearance of an object to change from one photograph to the next.

REFERENCES

Aber, J. S., S. W. Aber, and F. Pavri. 2002. Unmanned Small-Format Aerial Photography from Kites for Acquiring Large-Scale, High-Resolution Multiview-Angle Imagery. *Pecora 15/ Land Satellite Information IV/ISPRS Commission I/FIEOS 2002 Conference Proceedings*. Bethesda, MD: American Society of Photogrammetry and Remote Sensing.

Aber, J. S., I. Marzolff, and J. Ries. 2010. *Small-Format Aerial Photography: Principles, Techniques, and Geosciences Applications*. Amsterdam: Elsevier, 268 pp.

Boland, J., T. Ager, E. Edwards, E. Frey, P. Jones, R. K. Jungquiet, . . . and H. Zuegge. 2004. Cameras and Sensing Systems. Chapter 8 in the *Manual of Photogrammetry* (J. C. McGlone, E. M. Mikhail, J. Bethel, and R. Mullen, eds.). Bethesda, MD: American Society for Photogrammetry and Remote Sensing, pp. 581–676.

Boreman, G. D. 1998. *Basic Electro-Optics for Electrical Engineers*. SPIE Tutorial Texts in Optical Engineering, Vol. TT32. Bellingham, WA: Society of Photo-optical Instrumentation Engineers, 97 pp.

British Columbia Ministry of Sustainable Resource Management. 2003. *Specifications for Scanning Aerial Photographic Imagery*. Victoria: Base Mapping and Geomatic Services Branch, 26 pp.

Clarke, T. A., and J. G. Fryer. 1998. The Development of Camera Calibration Methods and Models. *Photogrammetric Record*, Vol. 91, pp. 51–66.

Eller, R. 2000. *Secrets of Successful Aerial Photography*. Buffalo, NY: Amherst Media, 104 pp.

Erb, T. L., W. R., Philipson, W. T. Tang, and T. Liang. 1981. Analysis of Landfills with Historic Airphotos. *Photogrammetric Engineering and Remote Sensing*, Vol. 47, pp. 1363–1369.

Graham, R., and A. Koh. 2002. *Digital Aerial Survey: Theory and Practice*. Boca Raton, FL: CRC Press, 247 pp.

Li, R., and C. Liu. 2010. Photogrammetry for Remote Sensing. Chapter 16 in *Manual of Geospatial Science and Technology* (J. D. Bossler, ed.). Boca Raton, FL: CRC Press, pp. 285–302.

Light, D. L. 1993. The National Aerial Photography as a Geographic Information System Resource. *Photogrammetric Engineering and Remote Sensing*, Vol. 59, pp. 61–65.

Linder, W. 2006. *Digital Photogrammetry: A Practical Course* (2nd ed.). Berlin: Springer.

Lyon, J. G. 1987. Use of Maps, Aerial Photographs, and Other Remote Sensor Data for Practical Evaluations of Hazardous Waste Sites. *Photogrammetric Engineering and Remote Sensing*, Vol. 53, pp. 515–519.

Neumann, K., M. Welzenbach, and M. Timm. 2016. CMOS Imaging Sensor Technology for Aerial Mapping Cameras. *ISPRS International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, Vol. XLI-B1, pp. 69–72.

Petrie, G. 2007. Airborne Digital Imaging Technology: A New Overview. *Photogrammetric Record,* Vol. 22, No. 119, pp. 203–225.
Petrie, G. 2009. Systematic Oblique Aerial Photography Using Multiple Frame Cameras. *Photogrammetric Engineering and Remote Sensing,* Vol. 75, pp. 102–107.
Plasker, J. R., and G. W. TeSelle. 1998. Present Status and Future Applications of the National Aerial Photography Program. In Proceedings of the ACSM/ASPRS Convention. Bethesda, MD: American Society for Photogrammetry and Remote Sensing, pp. 86–92.
Sandau, R., B. Braunecker, H. Driescher, A. Eckart, S. Hilbert, J. Hutton, . . . and S. Wicki. 2000. Design Principles of the LH Systems ADS40 Airborne Digital Sensor. *International Archives of Photogrammetry and Remote Sensing,* Vol. 33, Part B1, pp. 258–265.
Stimson, A. 1974. *Photometry and Radiometry for Engineers*. New York: John Wiley, 446 pp.
Stow, D. A., L. L. Coulter, and C. A. Benkleman. 2009. Airborne Digital Multispectral Imaging. Chapter 11 in *The Sage Handbook of Remote Sensing* (T. A. Warner, M. Duane Nellis, and G. M. Foody, eds.). London: Sage, pp. 131–165.
Wolf, P. R. 1983. *Elements of Photogrammetry, with Air Photo Interpretation and Remote Sensing*. New York: McGraw-Hill, 628 pp.

APPENDIX 4.1 YOUR OWN INFRARED PHOTOGRAPHS

Anyone with even modest experience with amateur photography can practice infrared photography, given the necessary materials (see Figure 4.28). Although 35-mm film cameras, the necessary filters, and infrared-sensitive films are still available for the dedicated amateur, many will prefer to use digital cameras that have been specially modified to acquire only radiation in the near-infrared region. Infrared films are essentially similar to the usual films, but they should be refrigerated prior to use and exposed promptly, as the emulsions deteriorate much more rapidly than do those of normal films. Black-and-white infrared films should be used with a deep red filter to exclude most of the visible spectrum. Black-and-white infrared film can be developed using normal processing for black-and-white emulsions, as specified by the manufacturer. Digital cameras that have been modified for infrared photography do not require use of an external filter.

CIR films are also available in 35-mm format. They should be used with a yellow filter, as specified by the manufacturer. Processing of CIR film will require the services of a photographic laboratory that specializes in customized work, rather than the laboratories that handle only the more usual films. Before purchasing the film, it is best to inquire concerning the availability and costs of processing. There are few digital cameras currently available that have been modified for color infrared photography. Models formerly available may be available in the used camera market, although expense may be high even for secondhand cameras.

Results are usually best with bright illumination. For most scenes, the photographer should take special care to face away from the Sun while taking photographs. Because of differences in the reflectances of objects in the visible and the NIR spectra, the photographer should anticipate the nature of the scene as it will appear in the infrared region of the spectrum. (Artistic photographers have sometimes used these differences to create special effects.) The camera lens will bring infrared radiation to a focal point that differs from that for visible radiation, so infrared images may be slightly out of focus if the normal focus is used. Some lenses have special markings to show the correct focus for infrared films; most digital cameras modified for infrared photography have also been modified to provide the correct focus.

FIGURE 4.28 Black-and-white infrared photograph (top) and a black-and-white photograph of the same scene (bottom).

APPENDIX 4.2 YOUR OWN 3D PHOTOGRAPHS

You can take your own stereo photographs using a handheld camera simply by taking a pair of overlapping photographs. Two photographs of the same scene, taken from slightly different positions, create a stereo effect in the same manner in which overlapping aerial photographs provide a three-dimensional view of the terrain.

This effect can be accomplished by aiming the camera to frame the desired scene, taking the first photograph, moving the camera laterally a short distance, and then taking a second photograph that overlaps the field of view of the first. The lateral displacement need only be a few inches (equivalent to the distance between the pupils of a person's eyes), but a displacement of a few feet will often provide a modest exaggeration of depth that can be useful in distinguishing depth (Figure 4.25). However, if the displacement is too great, the eye cannot fuse the two images to simulate the effect of depth.

Prints of the two photographs can then be mounted side by side to form a stereo pair that can be viewed with a stereoscope, just as a pair of aerial photos can be viewed in stereo. Stereo images can provide three-dimensional ground views that illustrate conditions encountered within different regions delineated on aerial photographs. Chapter 6 provides more information about viewing stereo photographs.

No 3D Glasses Required—Amazing 3D Stereoscopic Images:
www.youtube.com/watch?v=Eq3MyjDS1co&feature=related

APPENDIX 4.3 YOUR OWN KITE PHOTOGRAPHY

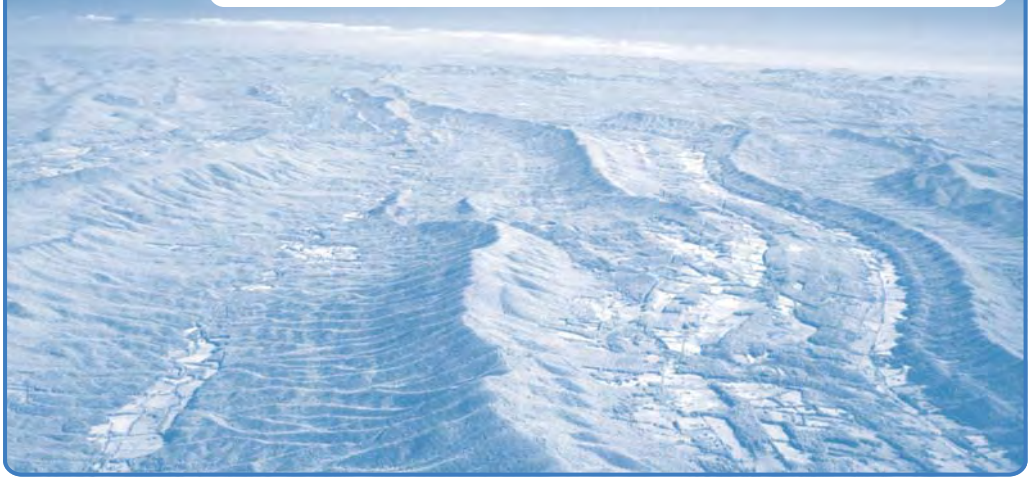
Although success requires persistence and attention, do-it-yourself kite photography is within the reach of nearly anyone who has the interest. The main prerequisites are access to a small digital camera, a reasonably robust kite, and the skill to fabricate a homemade mount for the camera. Aside from experience, the main obstacle for most beginners will be devising the mount to permit the camera's field of view to face the ground at the desired orientation. An abundance of books and websites are available that can provide design and instructions. The motion of the kite will cause the camera to swing from side to side, thereby producing a number of unsatisfactory photographs that must be screened to find those that are most suitable. These effects can be minimized by use of more elaborate mounts for cameras and possibly by attention to the choice of kite.

http://kap.ced.berkeley.edu

Make Podcast: Weekend Projects—Make a Kite Aerial Photograph:
www.youtube.com/watch?v=kEprozoxnLY&feature=fvw

Maker Workshop—Kite Aerial Photography on MAKE:television:
www.youtube.com/watch?v=swqFA9Mvq5M

5 Digital Imagery



MAJOR TOPICS TO UNDERSTAND

- Electronic Imagery
- Spectral Sensitivity
- Digital Data
- Data Formats
- Band Combinations: Multispectral Imagery
- Image Enhancement
- Image Display
- Image Processing Software

5.1 INTRODUCTION

For much of the history of aerial survey and remote sensing, images were recorded as photographs or photograph-like images. A photographic image forms a physical record: pieces of paper or film with chemical coatings that record the patterns of the images. Such images are referred to as *analog* images because the brightness values within a photograph are proportional (i.e., analogous) to the brightness values within a scene. Although

photographic media have enduring value for recording images, in the context of remote sensing, their disadvantages (including difficulties of storage, transmission, searching, and analysis) form liabilities.

In contrast, digital image formats represent images as arrays of many individual values, known as *pixels,* that together form an image. When an image is represented as discrete numbers, it acquires qualities that offer many advantages over earlier analog formats. Digital values can be added, subtracted, multiplied, and, in general, subjected to statistical manipulation that is not feasible if images are presented only in analog format. Digital images are also easy to store in compact formats and easily transmitted, and storage and retrieval is inexpensive and effective. Thus, the digital format greatly increases our ability to display, examine, and analyze remotely sensed data. However, we should note that digital formats have their own limitations, which are not always recognized. Imagery can be only as secure as the media on which they are stored, so just as analog images are subject to deterioration by aging, mishandling, and wear of physical media, so digital data are subject to corruption, damage to disk drives, magnetic fields, and deterioration of the physical media. Equally significant are changes in the formats of digital storage media, which can render digital copies inaccessible because of obsolescence of the hardware necessary to read the digital media.

This chapter introduces some of the fundamental concepts underlying applications of digital data for remote sensing and expands on some of the concepts first introduced in Chapter 4. It addressees the collection of digital data, representation of digital values, alternative formats for storing digital data, display of digital data, and image processing software systems.

5.2 ELECTRONIC IMAGERY

Digital data can be created by a family of instruments that can systematically scan portions of the Earth's surface, recording photons reflected or emitted from individual patches of ground, known as pixels. A digital image is composed of many thousands of pixels, each of which is usually too small to be individually resolved by the human eye, each representing the brightness of a small region on the Earth's surface, recorded digitally as a numeric value, usually with separate values for each of several regions of the electromagnetic spectrum. Color images are composed of several such arrays of the same ground area, each representing brightness values in a separate region of the spectrum.

Digital images can be generated by several kinds of instruments. Chapter 4 has already introduced some of the most important technologies for digital imaging: CCDs and CMOS. Another technology, *optical–mechanical scanning,* is older but has proven to be reliable and is still important in several realms of remote sensing practice (**Figure 5.1**). Here, we focus on a specific form of the optical–mechanical scanner that is designed to acquire imagery in several spectral regions—the *multispectral scanner,* which has formed an enduring technology for the practice of remote sensing. Other forms of optical–mechanical scanners have been used for collecting thermal imagery (thermal scanners; Chapter 10) and hyperspectral imagery (hyperspectral scanners; Chapter 14).

As noted in Chapter 4, CCDs can be positioned in the focal plane of a sensor such that they view a thin rectangular strip oriented at right angles to the flight path (**Figure 5.1a**). The forward motion of the aircraft or satellite moves the field of view forward along the flight path, building up coverage. This process is known as *pushbroom scan-*

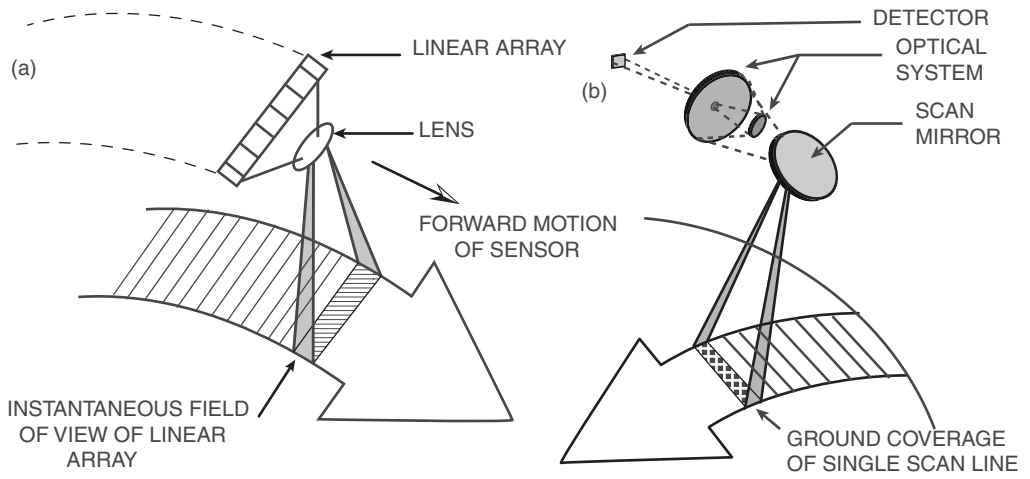


FIGURE 5.1 Optical–mechanical scanner. Whereas a linear array (a) acquires imagery line by line as its field of view slides forward along the ground track, the scan mirror of an optical–mechanical scanner (b) oscillates from side to side to build coverage pixel by pixel as the field of view progresses forward. Image by Susmita Sen.

ning—the linear array of pixels slides forward along the flight path in a manner analogous to the motion of a janitor's pushbroom along a floor. In contrast, mechanical scanning can be visualized by analogy to a whiskbroom, in which the side-to-side motion of the scanner constructs the lateral dimension of the image (**Figure 5.1b**), as the forward motion of the aircraft or satellite creates its longitudinal dimension.

Optical–mechanical scanners physically move mirrors, or prisms, to systematically aim the field of view over the Earth's surface. The scanning mirror scans across the field of view at a right angle to the flight path of the sensor, directing radiation from the Earth's surface to a secondary optical system and eventually to detectors that generate an electrical current that varies in intensity as the land surface varies in brightness (**Figure 5.1b**). Filters or diffraction gratings (discussed in the next section) split the radiation into several segments to define separate spectral channels, so the instrument generates several signals, each carrying information about the brightness in a separate region of the spectrum. The electrical current provides an electronic version of the brightness of the terrain but is still in analog form; it provides a continuous record of brightness values observed by the sensor's optics. To create a digital version, the electrical signal must be subdivided into distinct units to create the discrete values necessary for digital analysis. This conversion from the continuously varying analog signal to the discrete values is accomplished by sampling the current at a uniform interval, a process known as analog-to-digital, or A-to-D, conversion (**Figure 5.2**). Because the values within this interval are represented as a single average, all variation within this interval is lost. The process of subdivision and averaging the continuous signal corresponds to sampling the terrain at a set spatial interval, so the choice of sampling interval establishes the spatial detail recorded by the image.

The *instantaneous field of view* (IFOV) of an optical–mechanical scanner refers to the area viewed by the instrument if it were possible to suspend the motion of the aircraft

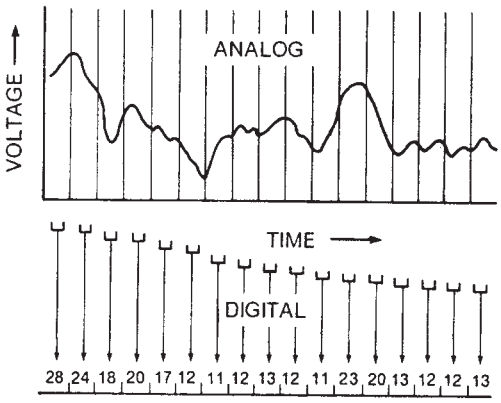


FIGURE 5.2 Analog-to-digital conversion.

and the scanning of the sensor for an instant (Figure 5.3). The IFOV therefore defines the smallest area viewed by the scanner and establishes a limit for the level of spatial detail that can be represented in a digital image. Although data in the final image can be aggregated so that an image pixel represents a ground area larger than the IFOV, it is not possible for pixels to carry information about ground areas smaller than the IFOV. More generally, the concept of the *ground resolved distance* (GRD) specifies the estimated dimension of the size of the smallest feature that can be reliably resolved by an imaging system.

Electronic sensors must be operated within the limits of their design capabilities. Altitudes and speeds of aircraft and satellites must be selected to match the sensitivities of the sensors, so that detectors view a given ground area (pixel) long enough to accumulate enough photons to generate reliable signals (this interval is known as *dwell time*). If designed and operated effectively, the imaging system should provide a linear response to scene brightness, such that the values within an image will display consistent, predictable relationships with brightness on the ground. Although most sensors have good performance under normal operating conditions, they will be subject to failures under extreme conditions.

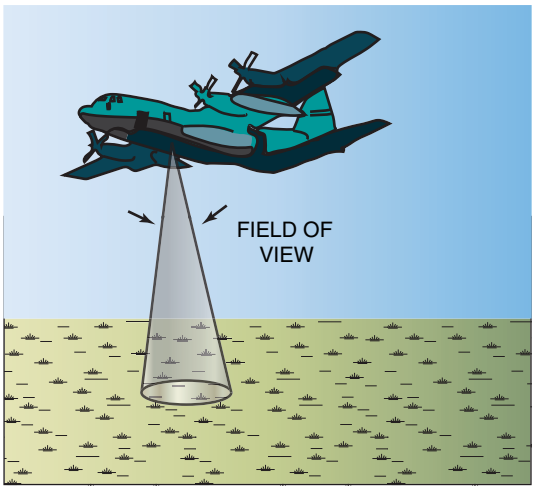


FIGURE 5.3 Instantaneous field of view. Image by Susmita Sen.

The lower end of an instrument's sensitivity is subject to the *dark current* signal (or *dark current noise*) (Figure 5.4). A CCD can record low levels of brightness even when there is none in the scene, due to energy within the CCD's structure that is captured by the potential well and presented as brightness. Thus, very dark features will not be represented at their correct brightness. Likewise, at an instrument's upper threshold of sensitivity, bright targets *saturate* the sensor's response—the instrument fails to record the full magnitude of the target's brightness. As an example in the context of remote sensing, saturation might be encountered in images representing glaciers or snowfields, which may exceed a sensor's ability to record the full range of brightness values in the optical region of the spectrum. For instruments using CCDs, saturation can sometimes manifest itself as *streaking* or *blooming*, as the excess charges that accumulate at a specific pixel site spill over to influence the charges at adjacent locations, creating bright streaks or patches unrelated to the actual features in the scene.

Between these limits, sensors are designed to generate signals that have predictable relationships with scene brightness; these relationships are established by careful design, manufacture, and calibration of each instrument. These characteristics define the upper and lower limits of the system's sensitivity to brightness and the range of brightness values over which a system can generate measurements with consistent relationships to scene brightness values.

The range of brightnesses that can be accurately recorded is known as the sensor's *dynamic range*. The lower limit of an instrument's dynamic range is set during calibration at a level above the minimum illustrated in Figure 5.4, known as the *offset*. In general, electronic sensors have large dynamic ranges compared with those of photographic

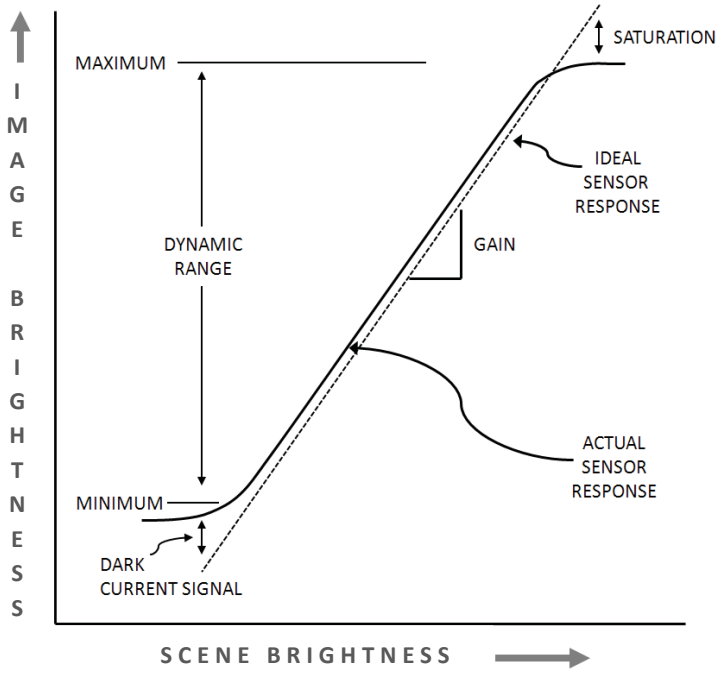


FIGURE 5.4 Dark current, saturation, and dynamic range.

films, computer displays, or the human visual system. Therefore, photographic representations of electronic imagery tend to lose information at the upper and/or lower ranges of brightnesses. Because visual interpretation forms such an important dimension of our understanding of images, the way that image displays and image-enhancement methods (discussed subsequently) handle this problem forms an important dimension of the field of image analysis.

The slope of the line depicted in Figure 5.4 defines the *gain* of a sensor, expressing the relationship between the brightness values in the original scene and its representation in the image. Figure 5.4 represents the behavior of an instrument that portrays a range of brightness values that are approximately proportional to the range of brightness in the original scene (i.e., the slope of the line representing the relationship between scene brightness and image brightness is oriented at approximately 45°). A slope of 1 means that a given range of brightness values in the scene is assigned the same range in the image, whereas a steeper slope (*high gain*) indicates that a given range of brightness values in the scene is expanded to assume a larger range in the scene. In contrast, Figure 5.5 shows two hypothetical instruments, one with *high gain* (i.e., it portrays a given range of brightness values in the scene as a larger range in the image) and another with *low gain* (i.e., the instrument creates an image with a narrower range of brightness values than is observed in the scene). The gain for a sensor is usually fixed by the design of an instrument, although some may have alternative settings (*high gain* or *low gain*) to accommodate varied scenes or operational conditions.

Each sensor creates responses unrelated to target brightness, that is, *noise*, created in part by accumulated electronic errors from various components of the sensor. (In this context, "noise" refers specifically to noise generated by the sensor, although the noise that the analyst receives originates not only in the sensor but also in the atmosphere, the interpretation process, and so on.) For effective use, instruments must be designed so that their noise levels are small relative to the signal (brightness of the target). This is

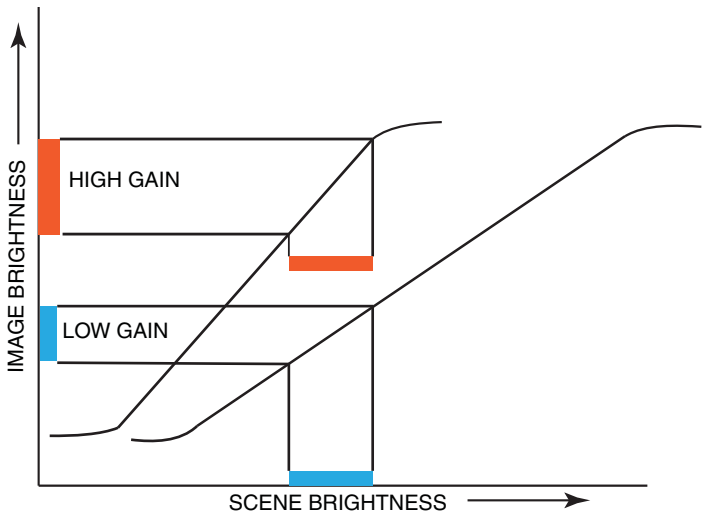


FIGURE 5.5 Examples of sensors characterized by high and low gain. Image by Susmita Sen.

measured as the signal-to-noise ratio (S/N or SNR) (Figure 5.6). Analysts desire signals to be large relative to noise, so the SNR should be large not only for bright targets when the signal is large but also over the entire dynamic range of the instrument, especially at the lower levels of sensitivity when the signal is small relative to noise. Engineers who design sensors must balance the radiometric sensitivity of the instrument with pixel size, dynamic range, operational altitude, and other factors to maintain acceptable SNRs.

5.3 SPECTRAL SENSITIVITY

Optical sensors often use prisms and filters to separate light into spectral regions. Filters are pieces of specialized glass that selectively pass certain wavelengths and block or absorb those that the designer desired to exclude. The most precise (and therefore most expensive) filters are manufactured by adding dyes to glass during manufacture. Less precise, and less durable, filters are manufactured by coating the surface of glass with a film that absorbs the desired wavelengths. Usually, filters are manufactured by firms that produce a suite of filters, each with its own system for defining and designating filters, specifically tailored for the needs of certain communities of customers with specialized needs.

Because of the scattering of shorter wavelengths, filters are often used when recording visible radiation to screen out ultraviolet or blue light (Figure 5.7a). Such a filter creates an image within the visible region that excludes the shorter wavelengths that can degrade the visual quality of the image. Often it is desirable to exclude all visible radiation, to create an image that is based entirely on near-infrared radiation (Figure 4.18, right; Figure 4.20, right). A deep red filter (Figure 5.7b) blocks visible radiation but allows infrared radiation to pass. An image recorded in the near-infrared region is quite different from its representation in the visible spectrum (Figure 4.20). For example, living vegetation is many times brighter in the near-infrared portion of the spectrum than

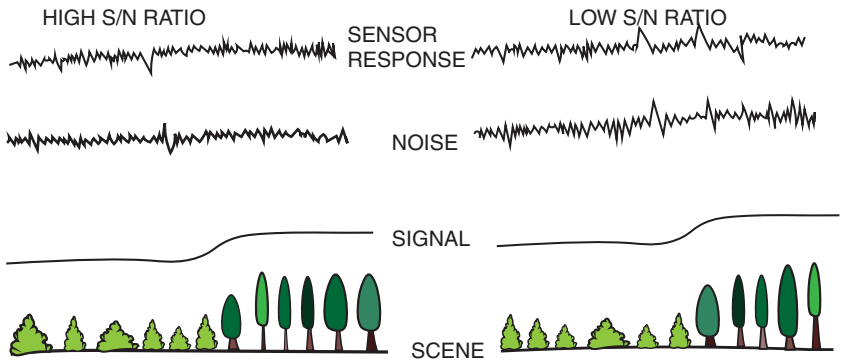


FIGURE 5.6 Signal-to-noise (S/N) ratio. At the bottom, a hypothetical scene is composed of two cover types. The signal records this region, with only a small difference in brightness between the two classes. Atmospheric effects, sensor error, and other factors contribute to noise, which is added to the signal. The sensor then records a combination of signal and noise. When noise is small relative to the signal (left: high S/N ratio), the sensor conveys the difference between the two regions. When the signal is small relative to noise (right: low S/N ratio), the sensor cannot portray the difference in brightness between the two regions. Image by Susmita Sen.

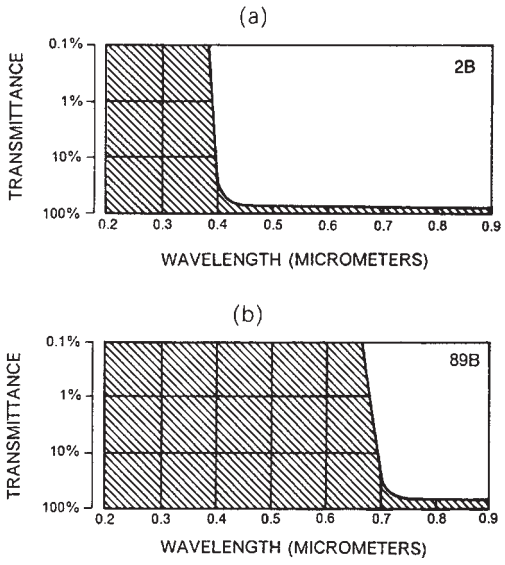


FIGURE 5.7 Transmission curves for two filters. (a) Pale yellow filter (Kodak filter 2B) to prevent ultraviolet light from reaching the focal plane; it is frequently used to acquire panchromatic images. (b) Kodak 89B filter used to exclude visible light, used for infrared images. (Shaded portions of the diagrams signify that the filter is blocking transmission of radiation at specified wavelengths.) From KODAK Photographic Filters Handbook (Code B-3). Copyright © Eastman Kodak Company. Used by permission.

it is in the visible portion, so vegetated areas appear bright white on the black-and-white infrared image.

Although filters can be used in the collection of digital imagery, electronic sensors often use *diffraction gratings,* considered more efficient because of their effectiveness, small size, and light weight. Diffraction gratings are closely spaced transmitting slits cut into a flat surface (a *transmission grating*) or grooves cut into a polished surface (a *reflection grating*). Effective transmission gratings must be accurately and consistently spaced and must have very sharp edges. Light from a scene is passed through a *collimating lens*, designed to produce a beam of parallel rays of light that is oriented to strike the diffraction grating at an angle (**Figure 5.8**).

Light striking a diffraction grating experiences both destructive and constructive interference as wavefronts interact with the grating. *Destructive interference* causes some wavelengths to be suppressed, whereas *constructive interference* causes others to be reinforced. Because the grating is oriented at an angle with respect to the beam of light, different wavelengths are diffracted at different angles, and the radiation can be separated spectrally. This light then illuminates detectors to achieve the desired spectral sensitivity.

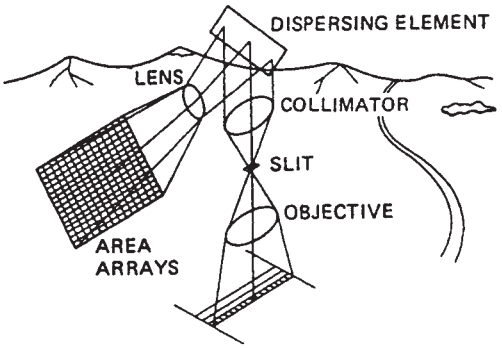


FIGURE 5.8 Diffraction grating and collimating lens. From NASA.

Because the various filters and diffraction gratings that instruments use to define the spectral limits (i.e., the "colors" that they record) do not define discrete limits, spectral sensitivity varies across a specific defined interval. For example, an instrument designed to record radiation in the green region of the spectrum will not exhibit equal sensitivity across the green region but will exhibit greater sensitivity near the center of the region than at the transitions to the red and blue regions on either side (Figure 5.9). Defining the spectral sensitivity to be the extreme limits of the energy received would not be satisfactory because it is clear that the energy received at the extremes is so low that the effective sensitivity of the instrument is defined by a much narrower wavelength interval.

As a result, the spectral sensitivity of an instrument is often specified using the definition of *full width, half maximum* (*FWHM*)—the spectral interval measured at the level at which the instrument's response reaches one-half of its maximum value (Figure 5.9). Thus FWHM forms a definition of *spectral resolution*, the narrowest spectral interval that can be resolved by an instrument. (Even though the instrument is sensitive to radiation at the extreme limits, beyond the limits of FWHM, the response is so weak and unreliable at these limits that FWHM forms a measure of functional sensitivity.) Figure 5.9 also illustrates the definition of the *spectral sampling interval* (known also as *spectral bandwidth*), which specifies the spectral interval used to record brightness in relation to wavelength.

5.4 DIGITAL DATA

Output from electronic sensors reaches the analyst as a set of numeric values. Each digital value is recorded as a series of binary values known as *bits*. Each bit records an exponent of a power of 2, with the value of the exponent determined by the position of the bit in the sequence. As an example, consider a system designed to record 7 bits for each digital

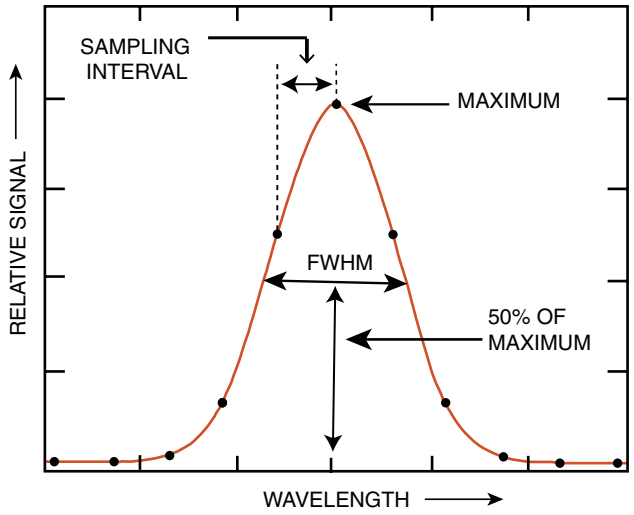


FIGURE 5.9 Full width, half maximum (FWHM). Image by Susmita Sen.

value. This means (for unsigned integers) that seven binary places are available to record the brightness sensed for each band of the sensor. The seven values record, in sequence, successive powers of 2. A "1" signifies that a specific power of 2 (determined by its position within the sequence) is to be evoked; a "0" indicates a value of zero for that position. Thus, the 7-bit binary number "1111111" signifies $2^6 + 2^5 + 2^4 + 2^3 + 2^2 + 2^1 + 2^0 = 64 + 32 + 16 + 8 + 4 + 2 + 1 = 127$. And "1001011" records $2^6 + 0^5 + 0^4 + 2^3 + 0^2 + 2^1 + 2^0 = 64 + 0 + 0 + 8 + 0 + 2 + 1 = 75$. Figure 5.10 shows two different examples. Eight bits constitute a *byte*, intended to store a single character. Larger amounts of memory can be indicated in terms of *kilobytes* (KB), 1,024 (2^{10}) bytes; *megabytes* (MB), 1,048,576 (2^{20}) bytes; and *gigabytes* (GB), 1,073,741,824 (2^{30}) bytes (Table 5.1).

In this manner, discrete digital values for each pixel are recorded in a form suitable for storage on disks and for analysis. These values are popularly known as digital numbers (DNs), brightness values (BVs), or digital counts, in part as a means of signifying that these values do not record true brightness (known as *radiance*) from the scene but rather are scaled values that represent relative brightness within each scene. The number of brightness values within a digital image is determined by the number of bits available. The 7-bit example given above permits a maximum range of 128 possible values (0–127) for each pixel. A decrease to 6 bits would decrease the range of brightness values to 64 (0–63); an increase to 8 bits would extend the range to 256 (0–255). Thus, given a constant noise level, the number of bits minus a reserved sign bit, if used, determines the radiometric resolution of a digital image. The number of bits available is determined

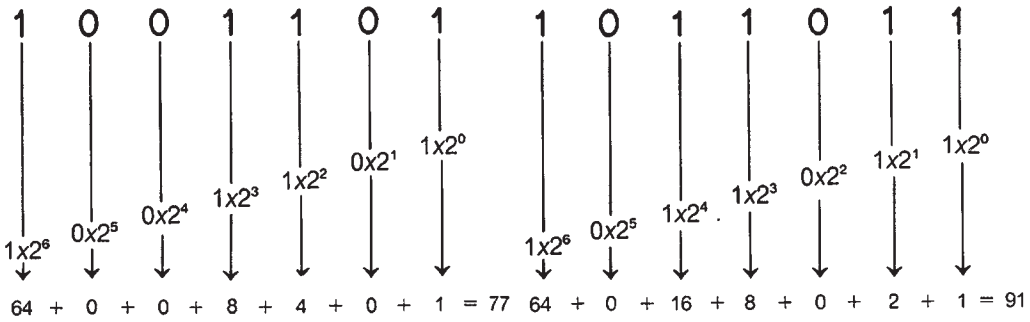


FIGURE 5.10 Digital representation of values in 7 bits.

TABLE 5.1 Terminology for Computer Storage

	Bit	A binary digit (0 or 1)	
	Byte	8 bits, 1 character	
	Kilobyte (K or KB)	1,024 bytes	(2^{10} bytes)
	Megabyte (MB)	1,048,576 bytes	(2^{20} bytes)
	Gigabyte (GB)	1,073,741,824 bytes	(2^{30} bytes)
	Terabyte (TB)	1,099,511,627,776 bytes	(2^{40} bytes)

by the design of the system, especially the sensitivity of the sensor and its capabilities for recording and transmitting data (each added bit increases transmission requirements). If we assume that transmission and storage resources are fixed, then increasing the number of bits for each pixel means that we will have fewer pixels per image and that pixels would each represent a larger ground area. Thus, technical specifications for remote sensing systems require trade-offs between image coverage and radiometric, spectral, and spatial resolutions.

Radiances

The brightness of radiation reflected from the Earth's surface is measured as brightness (watts) per wavelength interval (micrometer) per angular unit (steradian) per square meter from which it was reflected. Thus, the measured brightness is defined with respect to wavelength (i.e., "color"), spatial area (angle), intensity (brightness), and area. Radiance is a record of actual brightness, measured in physical units and represented as real values (i.e., to include decimal fractions). Use of DNs facilitates the design of instruments, data communications, and the visual display of image data. For visual comparison of different scenes, or analyses that examine relative brightness, use of DNs is satisfactory. However, because a DN from one scene does not represent the same brightness as the same DN from another scene, DNs are not comparable from scene to scene if an analysis must examine actual scene brightness for purposes that require use of original physical units. Such applications include comparisons of scenes of the same area acquired at different times, or matching adjacent scenes to make a mosaic.

For such purposes, it is necessary to convert the DNs to the original radiance or to use reflectance (Chapters 2 and 11), which are comparable from scene to scene and from one instrument to another. Calculation of radiance and reflectance from DNs requires knowledge of calibration data specific to each instrument. To ensure that a given sensor provides an accurate measure of brightness, it must be calibrated against targets of known brightness. The sensitivities of electronic sensors tend to drift over time, so to maintain accuracy, they must be recalibrated on a systematic schedule. Although those sensors used in aircraft can be recalibrated periodically, those used in satellites are not available after launch for the same kind of recalibration. Typically, such sensors are designed so that they can observe calibration targets onboard the satellite, or they are calibrated by viewing landscapes of uniform brightness (e.g., the moon or desert regions). Nonetheless, calibration errors, such as those described in Chapter 11, sometimes remain.

5.5 DATA FORMATS

Digital image analysis is usually conducted using *raster* data structures in which each image is treated as an array of values. Additional spectral channels form additional arrays that register to one another. Each pixel is treated as a separate unit, which can always be located within the image by its row and column coordinates. In most remote sensing analysis, coordinates originate in the upper left-hand corner of an image and are referred to as *rows* and *columns,* or as *lines* and *pixels,* to measure position down and to the right, respectively.

Raster data structures offer advantages for manipulation of pixel values by image processing systems, as it is easy to find and locate pixels and their values. The disad-

vantages are usually apparent only when we need to represent not the individual pixels, but areas of pixels, as discrete patches or regions. Then the alternative structure, *vector format,* becomes more attractive. The vector format uses polygonal patches and their boundaries as the fundamental units for analysis and manipulation. The vector format is not appropriate for digital analysis of remotely sensed data, although sometimes we may wish to display the results of our analysis using a vector format. Almost always, equipment and software for digital processing of remotely sensed data must be tailored for a raster format.

Digital remote sensing data are typically organized according to one of three alternative strategies for storing images. Consider an image consisting of four spectral channels, which together can be visualized as four superimposed images, with corresponding pixels in one band registering exactly to those in the other bands.

One of the earliest formats for digital data was *band interleaved by pixel* (BIP). Data are organized in sequence values for line 1, pixel 1, band 1; then for line 1, pixel 1, band 2; then for line 1, pixel 1, band 3; and finally for line 1, pixel 1, band 4. Next are the four bands for line 1, pixel 2, and so on (Figure 5.11). Thus, values for all four bands are written before values for the next pixel are represented. Any given pixel, once located within the data, is found with values for all four bands written in sequence one directly after the other. This arrangement is advantageous for many analyses in which the brightness value (or digital number) vector is queried or used to calculate another quantity. However, it is an unwieldy format for image display.

The *band interleaved by line* (BIL) format treats each line of data as a separate unit (Figure 5.12). In sequence, the analyst encounters line 1 for band 1, line 1 for band 2, line 1 for band 3, line 1 for band 4, line 2 for band 1, line 2 for band 2, and so on. Each line is represented in all four bands before the next line is encountered. A common variation on the BIL format is to group lines in sets of 3 or 7, for example, rather than to consider each single line as the unit.

A third convention for recording remotely sensed data is the *band sequential* (BSQ) format (Figure 5.13). All data for band 1 are written in sequence, followed by all data for band 2, and so on. Each band is treated as a separate unit. For many applications, this

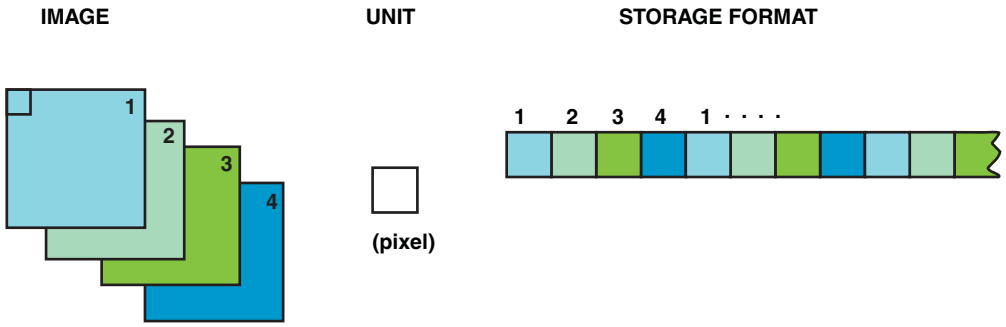


FIGURE 5.11 Band interleaved by pixel format. In effect, each band is subdivided such that data from pixels in the same location are collected from each sequential band and written to digital storage neighboring positions. Pixels from each band are intermingled as illustrated. Image by Susmita Sen.

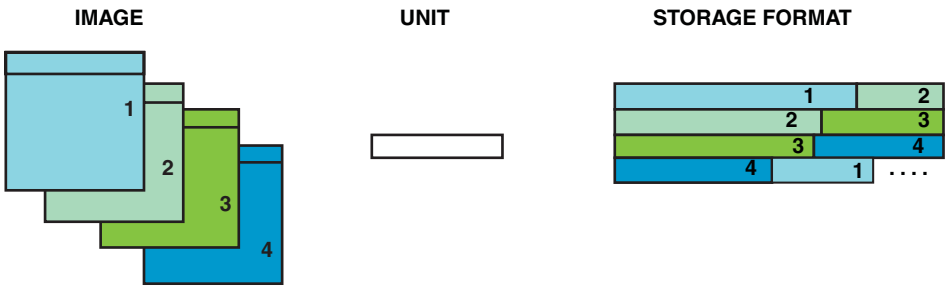


FIGURE 5.12 Band interleaved by line format. Lines of pixels from each band are selected and then written to digital storage such that lines for separate bands are positioned in sequence. Lines from each band are intermingled as illustrated. Image by Susmita Sen.

format is the most practical, as it presents data in the format that most closely resembles the data structure used for display and analysis. However, if areas smaller than the entire scene are to be examined, the analyst must read all four images before the subarea can be identified and extracted.

Actual data formats used to distribute digital remote sensing data are usually variations on these basic alternatives. Exact details of data formats are specific to particular organizations and to particular forms of data, so whenever an analyst acquires data,

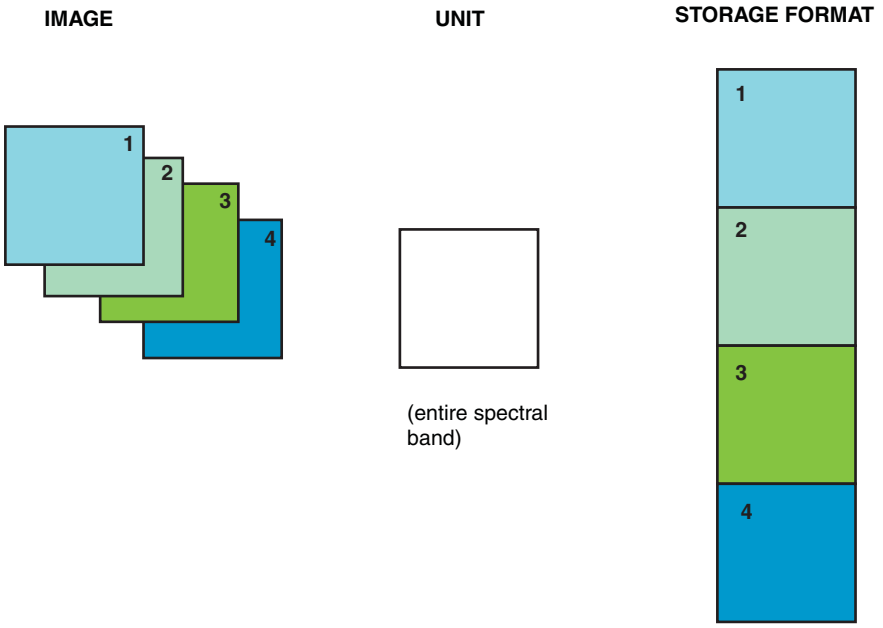


FIGURE 5.13 Band sequential format. The structure of each band is retained in digital storage; all pixels for each band are written in their entirety before the next band is written. There is no intermingling of pixels from separate bands. Image by Susmita Sen.

he or she must make sure to acquire detailed information regarding the data format. Although organizations attempt to standardize formats for specific kinds of data, it is also true that data formats change as new mass storage media come into widespread use and as user communities employ new kinds of hardware or software.

The "best" data format depends on immediate context and often on the specific software and equipment available. If all bands for an entire image must be used, then the BSQ and BIL formats are useful because they are convenient for reconstructing the entire scene in all four bands. If the analyst knows beforehand the exact position on the image of the subarea that is to be studied, then the BIP format is useful because values for all bands are found together and it is not necessary to read through the entire data set to find a specific region. In general, however, the analyst must be prepared to read the data in the format in which they are received and to convert them into the format most convenient for use at a specific laboratory.

Other formats are less common in everyday applications but are important for applications requiring use of long sequences of multispectral images. *Hierarchical data format* (HDF) is a specialized data structure developed and promoted by the National Center for Supercomputing Applications (*www.ncsa.illinois.edu/enabling/software*) and designed specifically to promote effective management of scientific data. Whereas the formats discussed thus far organize data conveyed by a specific image, HDF and related structures provide frameworks for organizing collections of images. For example, conventional data formats become awkward when it is necessary to portray three-dimensional data structures as they might vary over time. Although such structures might typically portray complex atmospheric data as it varies hourly, daily, seasonally, or yearly, they also lend themselves to recording large sequences of multispectral images. HDF therefore enables effective analysis and visualization of such large, multifaceted data structures.

A related but distinctly different format, *network common data form* (NetCDF), which also provides structures tailored for handling dynamic, array-oriented data, is specifically designed to be compatible with a wide variety of computer platforms, so that it can facilitate sharing of data over the World Wide Web. NetCDF was designed specifically for the Unidata system (*www.unidata.ucar.edu/software/netcdf*), which allows rapid transmission of meteorological data to a wide range of users.

Although HDF and NetCDF structures are unlikely to be encountered in usual remote sensing applications, they are becoming more common in advanced applications requiring the handling of very large sequences of images, such as those encountered in geophysics, meteorology, and environmental modeling—applications that often include remotely sensed data.

Data compression reduces the amount of digital data required to store or transmit information by exploiting the redundancies within a data set. If data arrays contain values that are repeated in sequence, then compression algorithms can exploit that repetition to reduce the size of the array, while retaining the ability to restore the array to its original form. When the complete array is needed for analysis, then the original version can be restored by *decompression*. Because remotely sensed images require large amounts of storage and usually are characterized by modest levels of redundancies, data compression is an important tool for effective storage and transmission of digital remote sensing data. Compression and decompression are accomplished, for example, by executing computer programs that receive compressed data as input and produce a decompressed version as output.

The *compression ratio* compares the size of the original image with the size of the compressed image. A ratio of 2:1 indicates that the compressed image is one-half the size of the original. *Lossless compression* techniques restore compressed data to their exact original form; *lossy* techniques degrade the reconstructed image, although in some applications the visual impact of a lossy technique may be imperceptible. For digital satellite data, lossless compression techniques can achieve ratios from 1.04:1 to 1.9 to 1. For digitized cartographic data, ratios of 24:1 using lossy techniques have been reported to exhibit good quality.

It is beyond the scope of this discussion to describe the numerous techniques and algorithms available for image compression. Probably the most well-known compression standard is the JPEG (Joint Photographic Experts Group) format, a lossy technique that applies the discrete cosine transform (DCT) as a compression–decompression algorithm. Although the JPEG algorithm has been widely accepted as a useful technique for compression of continuous-tone photographs, it is not likely to be ideal for either remotely sensed images or geospatial data in general. A modification of the basic JPEG format, JPEG2000 (*www.jpeg.org*) has been recognized as providing a high compression rate with high fidelity (Liu, Wu, and Shih, 2005). Depending on its application, JPEG2000 can be either lossy or nonlossy. Generally stated, lossy compression techniques should not be applied to data intended for analysis or as archival copies. Lossy compression may be appropriate for images used to present the visual records of results of the analytical process, provided they do not form input to other analyses.

5.6 BAND COMBINATIONS: MULTISPECTRAL IMAGERY

Effective display of an image is critical for effective practice of remote sensing. *Band combinations* is the term that remote sensing practitioners use to refer to the assignment of colors to represent brightness in different regions of the spectrum. Although there are many ways to assign colors to represent different regions of the spectrum, experience shows that some have proven to be more useful than others. A key constraint for the display of any multispectral image is that human vision portrays differences in the colors of surfaces through our eyes' ability to detect differences in brightness in the three additive primaries—blue, green, and red. Because our eyes can distinguish between brightness in these spectral regions, we can distinguish not only between blue, green, and red surfaces but also between intermediate mixtures of the primaries, such as yellow, orange, and purple.

Color films, digital displays, and the like portray the effect of color by varying the mixtures of the blue, green, and red primaries. Although films must employ a single strategy for portraying colors, image processing systems and digital displays offer the flexibility to use any of many alternative strategies for assigning colors to represent different regions of the spectrum. These alternative choices then define the band selection task, that is, how to decide which primary colors to select to portray on the display screen specific radiation collected by remote sensing systems.

If imagery at hand is limited to three spectral regions (as is the case with normal everyday color imagery), then the band selection task is simple: display radiation from blue in nature as blue on the screen, green as green, red as red. However, once we have more than three channels at hand, as is common for remotely sensed imagery, then the

choice of assignment can have only arbitrary answers because, for example, there can be no logical selection of colors portrayed outside the three primaries. The common choices for the band selection problem, then, are established in part on the basis of conventions that have been accepted by use over the decades and in part by practice that has demonstrated the effectiveness of certain combinations for certain purposes. An important theme for band combinations is that bands that are close to one another tend to replicate information in their adjacent regions of the spectrum. Therefore, the most effective band combinations are often (but not always!) formed from spectral regions that have different locations on the spectrum because they tend to provide independent representations of the same landscape.

Other color assignment models are used often but do not have widely accepted names. Here we discuss a few that are designated by the band numbers used for the Landsat Thematic Mapper, or Landsat TM (discussed in more detail in Chapter 7).

742

The 742 combination uses one region from the visible spectrum, one from the near infrared, and one from the mid-infrared region (Figure 5.14). It portrays landscapes using "false" colors but in a manner that resembles their natural appearance. Living, healthy vegetation appears in bright greens, barren soil as pink, dry vegetation and sparsely vegetated areas as oranges and browns, and open water as blue. This combination is often employed for geologic analyses, especially in desert landscapes, as differing mineralogies of surface soils appear as distinctive colors. Applications also include agriculture, wetlands, and, in forestry, fire management for postfire analysis of burned and unburned areas.

451

A 451 combination uses blue and mid-infrared radiation, together with a band from the near-infrared region (Figure 5.15). Deep, clear water bodies will appear very dark using this choice of bands—shallow or turbid water appears as shades of lighter blues. Healthy vegetation is represented in reds, browns, and oranges. Greens and browns often represent bare soils; white, cyan, and gray colors often represent urban features.

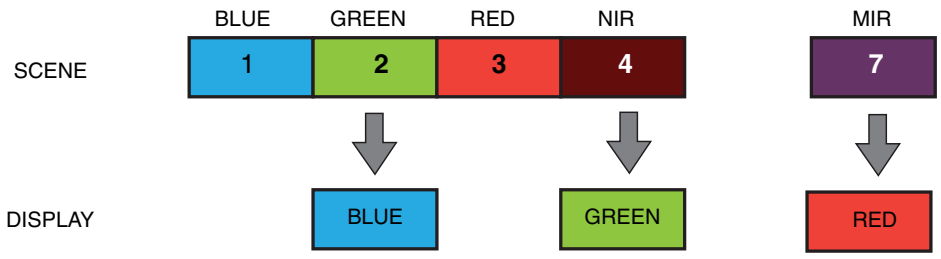


FIGURE 5.14 742 band combination, based on Landsat TM band designations (comparable Landsat 8 designations are 753). Image by Susmita Sen.

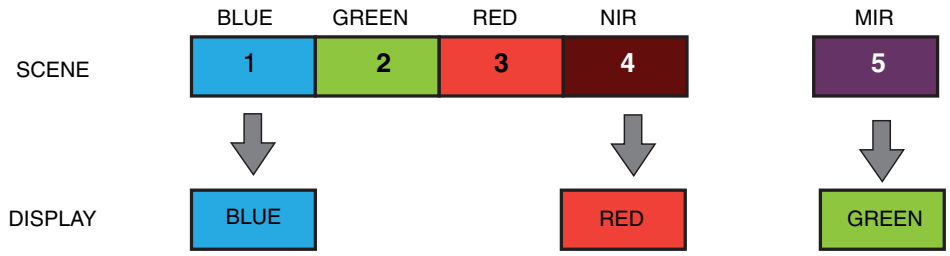


FIGURE 5.15 451 band combination, based on Landsat TM band designations (comparable Landsat 8 designations are 562). Image by Susmita Sen.

754

The 754 combination uses three bands from outside the visible region (Figure 5.16), which is often employed for geological analysis. Because it uses longer wavelengths, it is free of the effects of atmospheric scattering. Coastlines are clearly and sharply defined. Textural and moisture characteristics of soils can often be discerned.

543

A 543 combination uses the near-infrared, mid-infrared, and red regions (Figure 5.17). Edges of water bodies are sharply defined. It is effective in displaying variations in vegetation type and status as browns, greens, and oranges. It is sensitive to variations in soil moisture and useful for analysis of soil and vegetation conditions. Wetter surface soils appear in darker tones.

5.7 IMAGE ENHANCEMENT

Image enhancement is the process of improving the visual appearance of digital images. Image enhancement has increasing significance in remote sensing because of the growing importance of digital analyses. Although some aspects of digital analysis may seem

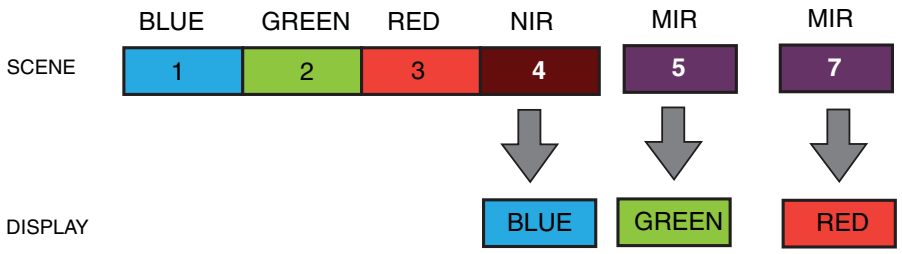


FIGURE 5.16 754 band combination, based on Landsat TM band designations (comparable Landsat 8 designations are 765). Image by Susmita Sen.

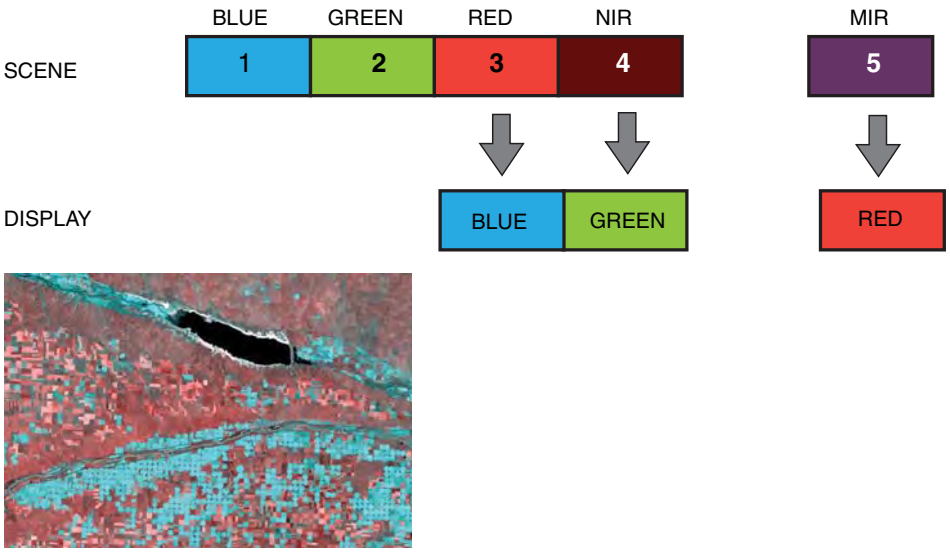


FIGURE 5.17 543 band combinations, based on Landsat TM band designations (top). Comparable Landsat 8 designations are 654. Note the separation of urban and rural areas, and clear identification of land-water boundaries (bottom). Top image by Susmita Sen.

to reduce or replace traditional image interpretation, many of these procedures require analysts to examine images on computer displays, performing tasks that require many of the skills outlined in earlier sections of this chapter.

Most image-enhancement techniques are designed to improve the visual appearance of an image, often as evaluated by narrowly defined criteria. Therefore, it is important to remember that enhancement is often an arbitrary exercise; what is successful for one purpose may be unsuitable for another image or for another purpose. In addition, image enhancement is conducted without regard for the integrity of the original data. The original brightness values will be altered in the process of improving their visual qualities, and they will lose their relationships to the original brightness on the ground. Therefore, enhanced images should not be used as input for additional analytical techniques; rather, any further analysis should use the original values as input.

Contrast Enhancement

Contrast refers to the range of brightness values present on an image. Contrast enhancement is required because sensors often generate brightness ranges that do not match the capabilities of the human visual system. Therefore, for analysts to view the full range of information conveyed by digital images, it is usually necessary to rescale image brightness to ranges that can be accommodated by human vision, photographic films, and computer displays. For example, if the maximum possible range of values is 0–255 (i.e., 8 bits) but the display can show only the range from 0 to 63 (6 bits), then the image will have poor contrast, and important detail may be lost in the values that cannot be shown on the display (Figure 5.18a). Contrast enhancement alters each pixel value in the old

image to produce a new set of values that exploits the full range of 256 brightness values (Figure 5.18b).

Figure 5.19 illustrates the practical effect of image enhancement. Before enhancement (left), detail is lost in the darker regions of the image. After enhancement has stretched the histogram of brightness values to take advantage of the capabilities of the display system, the detail is more clearly visible to the eye.

Many alternative approaches have been proposed to improve the quality of the displayed image. The appropriate choice of technique depends on the image, the previous experience of the user, and the specific problem at hand. The following paragraphs illustrate a few of the simpler and more widely used techniques.

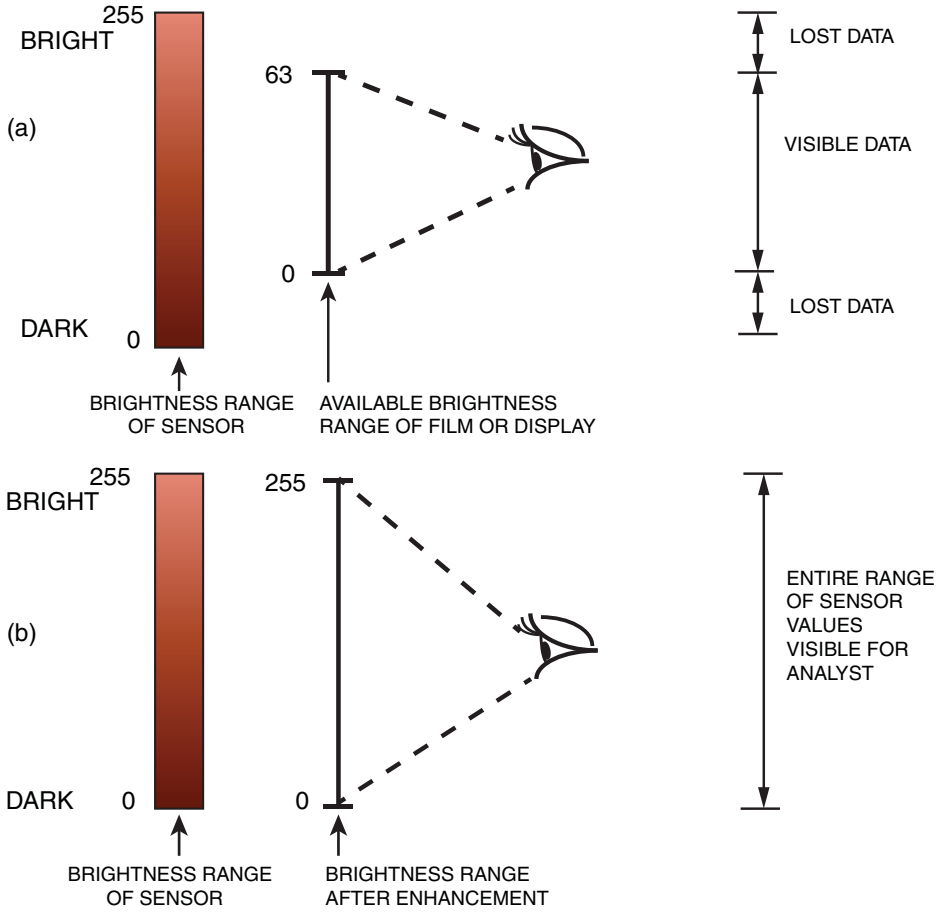


FIGURE 5.18 Schematic representation of the loss of visual information in display of digital imagery. (a) Often, the brightness range of digital imagery exceeds the ability of the image display to represent it to the human visual system. (b) Image enhancement rescales the digital values to more nearly match the capabilities of the display system. Image by Susmita Sen.

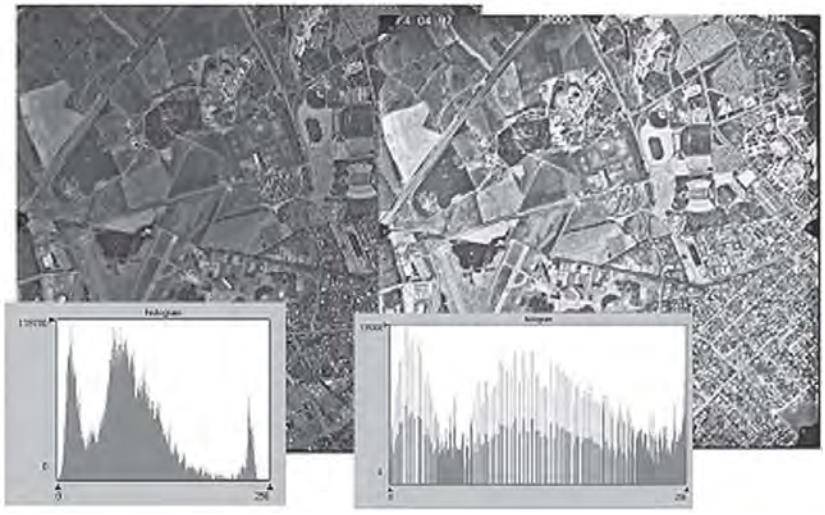


FIGURE 5.19 Pair of images illustrating the effect of image enhancement. By altering the distribution of brightness values, the analyst is able to view detail formerly hidden by the ineffective distribution of image brightness.

Linear Stretch

Linear stretch converts the original digital values into a new distribution, using new minimum and maximum values specified, often plus or minus two standard deviations from the mean. The algorithm then matches the old minimum to the new minimum and the old maximum to the new maximum. All of the old intermediate values are scaled proportionately between the new minimum and maximum values (Figure 5.20). *Piecewise linear stretch* means that the original brightness range was divided into segments before

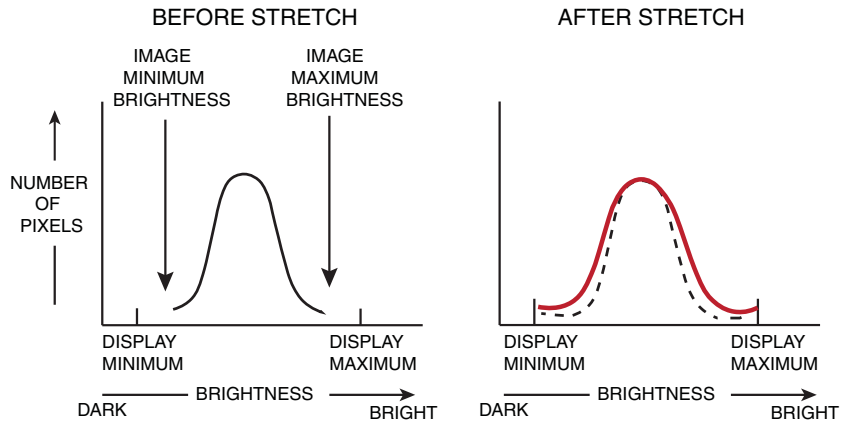


FIGURE 5.20 *Linear stretch.* Brightness values are spread over a broader range, allowing the eye to see detail formerly concealed in extremely dark or bright tones. Image by Susmita Sen.

each segment was stretched individually. This variation permits the analyst to emphasize certain segments of the brightness range that might have more significance for a specific application.

Histogram Equalization

Histogram equalization reassigns digital values in the original image such that brightness in the output image is equally distributed among the range of output values (Figure 5.21). Unlike contrast stretching, histogram equalization is achieved by applying a nonlinear function to reassign the brightness in the input image such that the output image approximates a uniform distribution of intensities. The histogram peaks are broadened, and the valleys are made shallower. Histogram equalization has been widely used for image comparison processes (because it is effective in enhancing image detail) and for adjustment of artifacts introduced by digitizers or other instruments.

Density Slicing

Density slicing is accomplished by arbitrarily dividing the range of brightness in a single band into intervals, then assigning each interval a color (Figure 5.22). Density slicing may have the effect of emphasizing certain features that may be represented in vivid colors, but, of course, it does not convey any more information than does the single image used as the source.

Edge Enhancement

Edge enhancement is an effort to reinforce the visual transitions between regions of contrasting brightness. Typically, the human interpreter prefers sharp edges between adjacent parcels, whereas the presence of noise, coarse resolution, and other factors often tend to blur or weaken the distinctiveness of these transitions. Edge enhancement in

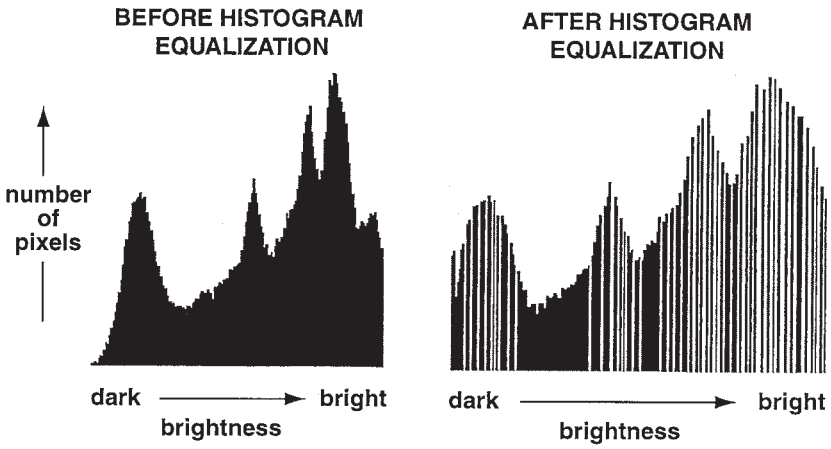


FIGURE 5.21 Histogram equalization. Equalization spreads the range of brightness values but preserves peaks and valleys in the histogram.

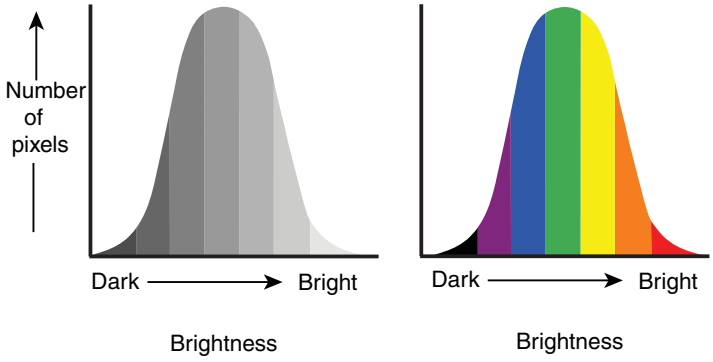


FIGURE 5.22 Density slicing. Colors are assigned to specific brightness values. Image by Susmita Sen.

effect magnifies local contrast—enhancement of contrast within a local region. A typical edge-enhancement algorithm consists of a usually square window that is systematically moved through the image, centered successively on each pixel. There are many edge-enhancement filters, but one of the most common is a variant of the Laplacian that works as follows for a 3 × 3 window: (1) the brightness value of each input pixel under the moving window, except for the center pixel, is multiplied by –1; (2) the center pixel is multiplied by 8; and (3) the center pixel in the output image is then given the value of the sum of all nine products resulting from (1) and (2).

Rohde, Lo, and Pohl (1978) describe an edge-enhancement procedure that can illustrate some of the specifics of this approach. A new (*output*) digital value is calculated using the original *(input*) value and the local average of five adjacent pixels. A constant can be applied to alter the effect of the enhancement as necessary in specific situations. The output value is the difference between twice the input value and the local average, thereby increasing the brightness of those pixels that are already brighter than the local average and decreasing the brightness of pixels that are already darker than the local average. Thus, the effect is to accentuate differences in brightness, especially at places ("edges") at which a given value differs greatly from the local average (Figure 5.23).

5.8 IMAGE DISPLAY

For remote sensing analysis, the image display is especially important because the analyst must be able to examine images and to inspect the results of analyses, which often are themselves images. At the simplest level, an image display can be thought of as a high-quality television screen, although those tailored specifically for image processing have image-display processors, which are special computers designed to rapidly receive digital data from the main computer, then display them as brightness on the screen. The capabilities of an image display are determined by several factors. First is the *size* of the image it can display, usually specified by the number of rows and columns it can show at any one time. Second, a display has a given *radiometric resolution;* that is, for each pixel, it has a capability to show a range of brightness. One-bit resolution would give the

capability to represent either black or white—certainly not enough detail to be useful for most purposes. In practice, most modern displays use 256 brightness levels for each of the primary colors of light (red, green, and blue).

A third factor controls the *rendition of color* in the displayed image. The method of depicting color is closely related to the design of the image display and the display processor. Image-display data are held in the *frame buffer,* a large segment of computer memory dedicated to handling data for display. The frame buffer provides one or more bits to record the brightness of each pixel to be shown on the screen (the "bit plane"); thus, the displayed image is generated, bit by bit, in the frame buffer. The more bits that have been designed in the frame buffer for each pixel, the greater the range of brightness that can be shown for that pixel, as explained earlier. For actual display on the screen, the digital value for each pixel is converted into an electrical signal that controls the brightness of the pixel on the screen. This requires a digital-to-analog (D-to-A) converter that translates discrete digital values into continuous electrical signals (the opposite function of the A-to-D converter mentioned previously).

Three strategies have been used for designing image displays: cathode ray tubes, liquid crystals, and plasma displays, each outlined here in abbreviated form.

The *cathode ray tube* (CRT) dates from the early 1940s, when it formed the basis for the first television displays. A CRT is formed from a large glass tube, wide at one end (the "screen") and narrow at the other. The inside of the wide end is coated with phosphor atoms. An electron gun positioned at the narrow end directs a stream of electrons against the inside of the wide end of the tube. As the electrons strike the phosphor coating, it glows, creating an image as the intensity of the electron beam varies according to the strength of the video signal. Electromagnets positioned on four sides of the narrow

FIGURE 5.23 Edge enhancement and image sharpening. A sample image is shown with and without enhancement. In this example, the effect of sharpening is especially noticeable at the edges of some of the larger shadows.

portion of the tube control the scan of the electron stream across the face of the tube, left to right, top to bottom.

As each small region of the screen is illuminated from the inside by the stream of electrons, it glows. Because the gun directs the stream of electrons systematically and very rapidly, it creates the images we can see on a computer display or television screen. Because the electron gun scans very rapidly (30–70 times each second), it can return to refresh, or update, the brightness at each pixel before the phosphor coating fades. In this manner, the image appears to the eye as a continuous image.

CRTs produce very clear images, but because an increase in the size of the screen requires a commensurate increase in the depth of the tube (so that the gun can illuminate the entire width of the screen), the size and weight of a CRT display present a major inconvenience. Thus, today CRTs have been supplanted by alternative technologies.

An alternative display technology was developed in the early 1970s. *Liquid crystal displays* (LCDs) depend on *liquid crystals,* substances that are intermediate between solid and liquid phases. The state assumed at a specific time depends on the temperature. An electrical current can change the orientation of the molecules within a liquid crystal and thereby block or transmit light at each pixel.

LCD displays use two sheets of polarizing materials that enclose a liquid crystal solution between them. When the video signal sends an electrical current to the display, the crystals align to block the passage of light between them. In effect, the liquid crystal at each pixel acts like a shutter, either blocking or passing the light. Color LCD displays use either of two alternative strategies; the best quality is provided by *active matrix displays,* also known as thin-film transistors (TFTs), which permit rapid refresh of the image. LCDs are used not only in watches, alarm clocks, and similar consumer products, but also in the flat-panel displays in portable computers and the compact displays now used for desktop computers.

A third display technology, *plasma display,* represents each pixel using three tiny fluorescent lights. These lights are small, sealed glass tubes containing an internal phosphor coating, an inert gas (mercury), and two electrodes. As an electrical current flows across the electrodes, it vaporizes some of the mercury. The electrical current also raises the energy levels of some of the mercury atoms; when they return to their original state, they emit photons in the ultraviolet portion of the spectrum. The ultraviolet light strikes the phosphor coating on the tube, creating visible light used to make an image. Variations in the coatings can create different colors. The positions of the tiny fluorescent light can be referenced as intersections in a raster grid, so the tube required for the CRT is not necessary for a plasma display, and the screen can be much more compact. Plasma displays are suitable for large, relatively compact image displays, but they are expensive, so they are not now preferred for analytical use.

Advanced Image Display

Remote sensing typically generates very large images portraying fine levels of detail. Conventional systems permit users to examine regions in fine detail only by zooming in to display the region of interest at the cost of losing the broader context. Or users can discard the finer detail and examine broad regions at coarse detail. Although this trade-off sometimes causes little or no inconvenience, in other situations the sacrifice of one quality for the other means that some of the most valuable qualities of the data are discarded.

As a result, there are incentives to design display systems that can simultaneously represent fine detail and large image size. Two alternative strategies have each found roles for viewing remotely sensed imagery.

"*Fisheye,*" or "*focus + context,*" displays enable the analyst to simultaneously view selected detail without discarding the surrounding context. Fisheye displays use existing display hardware, but with a simulated magnifier that can roam over the image to selectively enlarge specific regions within the context of the coarser-resolution display. Software "lenses" locally magnify a subset while maintaining a continuous visual connection to the remainder of the unmagnified image. Such capabilities can be linked to existing analytical software to enable the analyst to annotate, measure, and delineate regions to improve the functionality of existing software. The same effects can be achieved in a different manner by linking two windows within the display—one for detail, one for the broader context—and providing the analyst with the capability to alter the size of the window. This approach is known as "multiple linked views" or the "overview + detail" display.

Multiple-monitor systems (sometimes referred to as *tiled displays*) are formed as arrays of flat-panel monitors (or rear-projection displays) that can display very large images at high levels of spatial detail (**Figure 5.24**). The highest quality tiled displays can project between 250 and 300 million pixels; in the near future, gigapixel displays (1 billion pixels) will likely be attempted. Multiple-monitor systems enable the computer's operating system to use the display areas from two or more display devices to create a single display. The rear-projection systems (sometimes referred to as "power walls") do not have the seams between tiles that characterize the LCD tiled systems, so they have greater visual continuity, but the tiled LCDs have better visual quality and are cheaper.

Multiple-monitor systems became practical as the costs of random access memory (RAM) and LCD displays decreased to enable economical development of larger displays. Tiled displays are formed as a mosaic of screens, supported by operating systems configured to support multiple displays. Innovations in titled displays became possible when

FIGURE 5.24 Example of a tiled image display. Such displays are used for handling images that are too large to be read completely into memory (for example, satellite images, images of broad-scale weather systems, or detailed street grids). Photograph by Chris North, Virginia Tech Center for Human-Computer Interaction. Used by permission.

LCD technology permitted the assembly of multiple flat-panel displays into tiled displays. Tiled displays assembled from several LCDs use special mounts to hold multiple displays provided by commercial vendors and specialized software systems to permit integrated use of the several monitors. Analysts can display images across the seams formed by the edges of the displays and can zoom, create multiple windows, run multiple applications, and arrange windows as most effective for specific tasks.

Obvious applications include remotely sensed and GIS images that require the analyst to simultaneously exploit fine detail and broad areal coverage. Analysts sometimes desire to run two or more analytical programs simultaneously, with independent displays of the images generated by each analysis. Both fisheye and multiple-monitor systems are emerging technologies in the sense that they are both proven to be successful but are still under investigation to explore how they can be most effectively used in specific applications. Multiple-monitor systems have value for emergency management systems, which benefit from having the capability for many people to simultaneously view large images representing states, counties, or similar regions as single images, and also having the ability to manipulate the display as needed to assist in evaluation of complex situations.

5.9 IMAGE PROCESSING SOFTWARE

Digital remote sensing data can be interpreted by computer programs that manipulate the data recorded in pixels to yield information about specific subjects, as described in subsequent chapters. This kind of analysis is known as *image processing*, a term that encompasses a wide range of techniques. Image processing requires a system of specialized computer programs tailored to the manipulation of digital image data. Although such programs vary greatly in purpose, sophistication, and detail, there are also significant commonalities among most image processing systems.

A separate specific portion of the system is designed to read image data and to reorganize the data into the form to be used by the program. For example, many image processing programs manipulate the data in BSQ format. Thus, the first step may be to read BIL or BIP data and then reformat the data into the BSQ format required for the analytical components of the system. Another portion of the system may permit the analyst to subdivide the image into subimages; to merge, superimpose, or mosaic separate images; and in general to prepare the data for analysis, as described in Chapter 11. The heart of the system consists of a suite of programs that analyze, classify (Chapter 12), and manipulate data to produce output images and the statistics and data that may accompany them. Finally, a section of the image processing system must prepare data for display and output, either to the display processor or to the line printer. In addition, the program requires "housekeeping" subprograms that monitor movement and labeling of files from one portion of the program to another, generate error messages, and provide online documentation and assistance to the analysts.

Widely used image processing systems run on personal computers (PCs), Macs, or workstations. More elaborate systems can be supported by peripheral equipment, including extra mass storage, digitizers, scanners, color printers, disk drives, and related equipment. Almost all such systems are directed by menus and graphic user interfaces that permit the analyst to select options from a list on the screen.

Although many good image processing systems are available, some of the most commonly used are:

- Hexagon—ERDAS ER Mapper
  https://download.hexagongeospatial.com/en/downloads/imagine/erdas-er-mapper-2020
- PCI Geomatics—Geomatica
  www.pcigeomatics.com
- L3Harris Geospatial—ENVI
  www.harrisgeospatial.com/Software-Technology/ENVI
- Hexagon—ERDAS Imagine
  www.erdas.com
- GRASS GIS (Open Source)
  https://grass.osgeo.org
- Clark Labs—TerrSet (IDRISI)
  https://clarklabs.org

Lemmens (2004) provides a point-by-point comparison of image processing systems designed for remote sensing applications; Voss et al. (2010) list image processing systems, with point-by-point comparisons of their capabilities. The specific systems listed here can be considered general-purpose image processing systems; others have been designed specifically to address requirements for specific kinds of analysis (e.g., geology, hydrology), and some of the general-purpose systems have added optional modules that focus on more specific topics. Further details of image analysis systems are given by user manuals or help files for specific systems. Authors of image processing systems typically upgrade their systems to add new or improved capabilities, accommodate new equipment, or address additional application areas.

Several image processing systems are available to the public either without cost or at minimal cost. For students, some of these systems offer respectable capabilities to illustrate basics of image processing. A partial list includes:

ISIS: *http://isis.astrogeology.usgs.gov/index.html*
MultiSpec: *https://engineering.purdue.edu/~biehl/MultiSpec*
TNTlite: *www.microimages.com*
GRASS GIS: *https://grass.osgeo.org*
QGIS: *https://www.qgis.org/en/site*

Image Viewers, Online Digital Image Archives, and Cloud-Based Image Analysis

Image viewers (or, sometimes, *map viewers*) are programs designed to provide basic capabilities to view and navigate through digital maps and images. Some image viewers are available commercially; others are available online at minimal cost or as freeware. They provide a convenient means of examining digital maps, GIS data, and aerial imagery. Although most image viewers do not offer analytical capabilities, they do permit users to examine a wide range of spatial data by searching, roaming, magnifying, and applying a variety of projection and coordinate systems. Image viewers are closely connected to the idea of *digital imagery archives* or *libraries*, which provide collections of digital imagery

in standardized formats, such that viewers can easily retrieve and navigate through the collection. *Google Earth* (*http://earth.google.com*) is a well-known example of an online imagery archive that provides comprehensive coverage of the Earth's land areas, with the ability to roam and change scale, orientation, and detail.

Cloud-based image analysis is a rapidly evolving area for remote sensing and other disciplines that utilize very large data sets that are difficult to download and transfer. This allows researchers to apply their algorithms to data sets hosted on the cloud, using high-performance computing. We are now in the era of big data analytics and image fusion, and there are numerous examples of this technology in use or under development. *Google Earth Engine* provides a powerful cloud-based image analysis interface for scientific research through the *Earth Engine API,* allowing global-scale analysis of satellite data hosted in Google's archive *(https://earthengine.google.com)*. The *Planet* platform is another cloud-based application programming interface (API) designed specifically for remote sensing imagery *(www.planet.com/products/platform)*. Other large, well-known companies such as Amazon (e.g., Amazon Web Services: *https://aws.amazon.com)* and Microsoft (e.g., Microsoft Azure: *https://azure.microsoft.com/en-us)* have cloud-based computing platforms that provide services on a pay-by-use basis and are widely used for a variety of applications.

5.10 SUMMARY

Although digital data provide multiple advantages for practitioners of remote sensing, these advantages can be exploited only if the analyst has mastered the underlying concepts and how they influence applications of remote sensing to specific problems. The information presented here resurfaces in later chapters, as the fundamental nature of digital data underlie the basic design and application of remote sensing systems. Each instrument must be operated within its design constraints and applied to an appropriate task. Although this chapter cannot provide the details necessary to make such assessments, it can equip the reader with the perspective to seek the specifics that will permit a sound assessment of the question at hand.

SOME TEACHING AND LEARNING RESOURCES

- Diffraction Gratings
  www.youtube.com/watch?v=5D8EVNZdyy0
- Navigating a 13.3 Gigapixel Image on a 22 Megapixel Display Wall
  www.youtube.com/watch?v=8bHWuvzBtJo
- DOG (Difference of Gaussians)
  www.youtube.com/watch?v=Fe-pubQw5Xc

REVIEW QUESTIONS

1. It may be useful to practice conversion of some values from digital to binary form as confirmation that you understand the concepts. Convert the following digital numbers to 8-bit binary values:

   a. 100 c. 24 e. 2 g. 256

   b. 15 d. 31 f. 111 h. 123

2. Convert the following values from binary to digital form:

   a. 10110 c. 10111 e. 0011011

   b. 11100 d. 1110111 f. 1101101

3. Consider the implications of selecting the appropriate number of bits for recording remotely sensed data. One might be tempted to say, "Use a large number of bits to be sure that all values are recorded precisely." What would be the disadvantage of using, for example, 7 bits to record data that are accurate only to 5 bits?

4. Describe in a flow chart or diagram steps required to read data in a BIP format, and then organize them in a BSQ structure.

5. What is the minimum number of bits required to represent the following values precisely?

   a. 1,786 c. 689 e. 17 g. 29

   b. 32 d. 32,000 f. 3

6. Why are enhanced images usually not used as input for other analyses?

7. Density slicing produces an image that uses a range of contrasting colors (see **Figure 5.22**). Prepare a list of advantages and disadvantages of this form of image enhancement.

8. Do you expect that it is possible to estimate a sensor's SNR by visual examination of an image? How?

9. The digital format is now becoming the de facto standard for recording and storing aerial imagery. Discuss some of the advantages and disadvantages that accompany this change.

10. Some students find it highly illogical to use band combinations in which, for example, radiation in the red region of the spectrum is displayed using another color. Explain in a few concise sentences why such band combinations are useful.

REFERENCES

Ball, R., and C. North. 2005. Effects of Tiled High-Resolution Display on Basic Visualization and Navigation Tasks. In *Proceedings, Extended Abstracts of ACM Conference on Human Factors in Computing Systems* (CHI 2005). Portland, OR: Association for Computing Machinery, pp. 1196–1199.

Jensen, J. R. 2004. *Introductory Digital Image Processing, A Remote Sensing Perspective* (3rd ed.). Upper Saddle River, NJ: Prentice-Hall, 544 pp.

Jensen, J. R., and R. R. Jensen. 2002. Remote Sensing Digital Image Processing System Hardware and Software Considerations. *Manual of Geospatial Science and Technology* (J. Bossler, ed.). Boca Raton, FL: CRC Press, pp. 325–348.

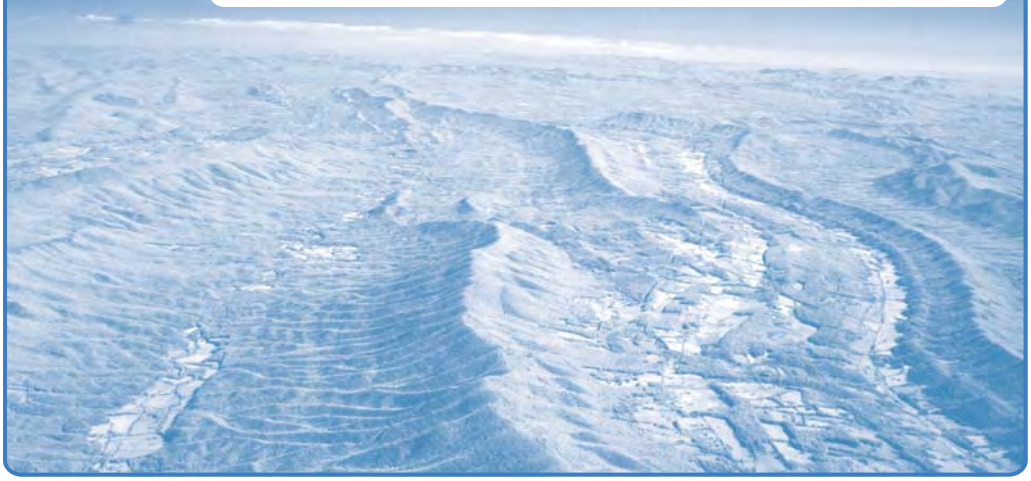
Lemmens, M. 2004. Remote Sensing Processing Software. *GIM International,* Vol. 18, pp. 53–57.

Liu, J.-K., H. Wu, and T. Shih. 2005. Effects of JPEG2000 on the Information and Geometry Content of Aerial Photo Compression. *Photogrammetric Engineering and Remote Sensing,* Vol. 71, pp. 157–167.

Rohde, W. G., J. K. Lo, and R. A. Pohl. 1978. EROS Data Center Landsat Digital Enhancement Techniques and Imagery Availability, 1977. *Canadian Journal of Remote Sensing,* Vol. 4, pp. 63–76.

Voss, M., R. Sugumaran, and D. Ershov. 2010. Software for Processing Remotely Sensed Data. Chapter 20 in *Manual of Geospatial Science and Technology* (J. D. Bossler, ed.). Boca Raton, FL: CRC Press, pp. 391–402.

6 Image Interpretation



MAJOR TOPICS TO UNDERSTAND

- The Context for Image Interpretation
- Image Interpretation Tasks
- Elements of Image Interpretation
- Collateral Information
- Imagery Interpretability Rating Scales
- Image Interpretation Keys
- Interpretive Overlays
- The Significance of Context
- Stereovision
- Digital Photointerpretation
- Image Scale Calculations

6.1 INTRODUCTION

Earlier chapters have defined our interest in remote sensing as focused primarily on images of the Earth's surface—map-like representations of the Earth's surface based on

the reflection of electromagnetic energy from vegetation, soil, water, rocks, and human-made structures. From such images we learn much that cannot be derived from other sources. Yet such information is not presented to us directly: the information we seek is encoded in the varied tones and textures we see on each image. To translate images into information, we must apply a specialized knowledge—knowledge that forms the field of *image interpretation,* which we can apply to derive useful information from the raw uninterpreted images we receive from remote sensing systems.

Photographic interpretation has been defined as "the act of examining photographic images for the purpose of identifying objects and judging their significance" (Colwell, 1960). Aerial photointerpretation dates from the early days of aerial photography (Campbell, 2008), which initially focused on military operations and later was applied to civil applications.

World War I formed the context for matching aviation with the camera. Specialized cameras were developed to enhance aerial observation, and instructional materials were prepared for training British and French aerial observers, then later for U.S. aviators. These materials were preserved as instructors' notes at the direction of Col. Edwin Gorrell (1891–1945), a U.S. aviator assigned to organize historical materials documenting U.S. aviation in the war. (This extensive work, *History of the Air Service AEF,* was never published as intended but is retained in the U.S. National Archives in a microfilm version available to the public.)

During the interwar years, publications regarding aerial photography focused mainly on the practicalities of aviation equipment. By the late 1930s, texts on applications of aerial photography were developed for geological mapping. Progress in development of aerial cameras, aircraft, and supporting services during the interwar decades increased interest in applications of aerial photography. By 1939 (e.g., Melton, 1939), geologists recognized the value of aerial observation for teaching earth sciences and supporting instruction to teach the basics of earth science (including streams, wind, waves, and glaciers).

World War II belligerents developed training materials, systemized as instructional materials, handbooks, and reference works, for instruction in quantitative measurements and recognition of enemy vehicles, weapons, and infrastructure. In the United States, especially in the aftermath of the war, experienced photoanalysts were in a position to apply their experience to civil enterprises and (in some instances) to take advantage of experienced pilots and sales of surplus materials and equipment to launch enterprises in aerial survey. Some of these efforts were described in early publications that expanded the audience for remote sensing image interpretation tasks to a broader public (e.g., Smith, 1942, 1943).

Aerial photo interpretation is an important skill, although specifics have changed over the decades as technology and institutional needs have evolved (**Figure 6.1**), while basics have remained central to the interpretation of digital imagery from any platform. As noted, its value was recognized initially in World War I and World War II and later in the cold war years. Its value has been recorded in the works of Babington-Smith (1957), Brugioni (1991), Brugioni and Doyle (1997), and O'Conner (2015), which provide first-person accounts of the role of image interpretation.

Proficiency in image interpretation is formed from three separate kinds of knowledge: the subject, the geographic region, and the remote sensing system. Only the remote sensing system falls within the scope of this book.

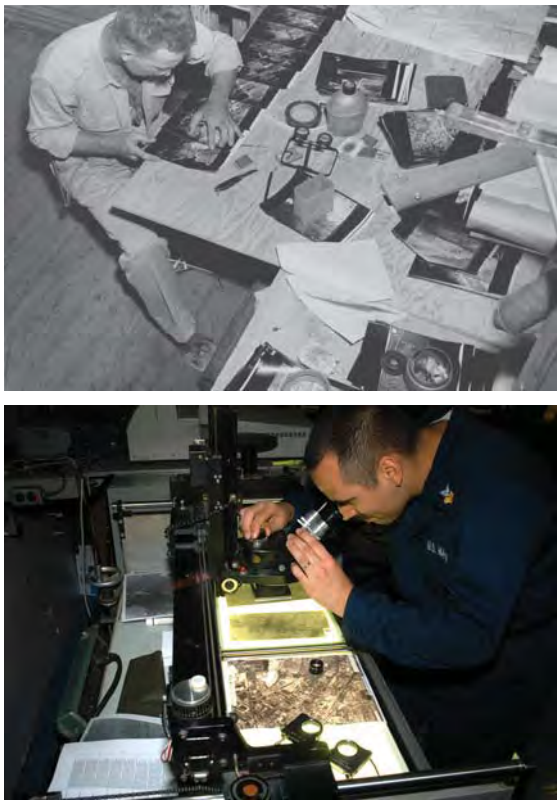


FIGURE 6.1 Changes in aerial photointerpretation, 1943–2004. From U.S. Army (top), U.S. Navy (bottom).

Subject

Knowledge of the *subject* of our interpretation—the kind of information that motivates us to examine the image—is the heart of the interpretation. Accurate interpretation requires familiarity with the subject of the interpretation. For example, interpretation of geologic information requires education and experience in the field of geology. Yet narrow specializations are a handicap because each image records a complex mixture of many kinds of information, requiring application of broad knowledge that crosses traditional boundaries between disciplines. For example, accurate interpretation of geologic information may require knowledge of botany and the plant sciences as a means of understanding how vegetation patterns on an image reflect geologic patterns that may not be directly visible. As a result, image interpreters should be equipped with a broad range of knowledge pertaining to the subjects at hand and their interrelationships.

Geographic Region

Knowledge of the specific geographic *region* depicted on an image can be equally significant. Every locality has unique characteristics that influence the patterns recorded on an

image. Often the interpreter may have direct experience within the area depicted on the image that can be applied to the interpretation. In unfamiliar regions, the interpreter may find it necessary to make a field reconnaissance or to use maps and books that describe analogous regions with similar climate, topography, or land use.

Remote Sensing System

Knowledge of the *remote sensing system* is obviously essential. The interpreter must understand how each image is formed and how each sensor portrays landscape features. Different instruments use separate portions of the electromagnetic spectrum, operate at different resolutions, and use different methods of recording images. The image interpreter must know how each of these variables influences the image to be interpreted and how to evaluate their effects on their ability to derive useful information from the imagery. This chapter outlines how the image interpreter derives useful information from the complex patterns of tone and texture on each image.

6.2 THE CONTEXT FOR IMAGE INTERPRETATION

Human beings are well prepared to examine images, as our visual system and experience equip us to discern subtle distinctions in brightness and darkness, to distinguish among various image textures, to perceive depth, and to recognize complex shapes and features. Even in early childhood we apply such skills routinely in everyday experience so that few of us encounter difficulties as we examine, for example, family snapshots or photographs in newspapers. Yet image analysis requires a conscious, explicit effort not only to learn about the subject matter, geographic setting, and imaging systems in unfamiliar contexts but also to develop our innate abilities for image analysis.

Three issues distinguish interpretation of remotely sensed imagery from interpretation conducted in everyday experience. First, remotely sensed images usually portray an *overhead view*—an unfamiliar perspective. Training, study, and experience are required to develop the ability to recognize objects and features from this perspective. Second, many remote sensing images use *radiation outside the visible portion* of the spectrum. In fact, use of such radiation is an important advantage that we exploit as often as possible. Even the most familiar features may appear quite different in nonvisible portions of the spectrum than they do in the familiar world of visible radiation. Third, remote sensing images often portray the Earth's surface at *unfamiliar scales and resolutions*. Commonplace objects and features may assume strange shapes and appearances as scale and resolution change from those to which we are accustomed.

This chapter outlines the art of image interpretation as applied to aerial photography. Students cannot expect to become proficient in image analysis simply by reading about image interpretation. Experience forms the only sure preparation for skillful interpretation. Nonetheless, this chapter can highlight some of the issues that form the foundations for proficiency in image analysis.

In order to discuss this subject at an early point in this book, we must confine the discussion to interpretation of aerial photography, the only form of remote sensing imagery discussed thus far. But the principles, procedures, and equipment described here are equally applicable to other kinds of imagery acquired by the sensors described in later chapters. Manual image interpretation is discussed in detail by Paine and Kiser (2003),

Avery and Berlin (2003), Philipson (1996), and Campbell (2005); older references that may also be useful are the text by Lueder (1959) and the *Manual of Photographic Interpretation* (Colwell, 1960).

6.3 IMAGE INTERPRETATION TASKS

The image interpreter must routinely conduct several kinds of tasks, many of which may be completed together in an integrated process. Nonetheless, for purposes of clarification, it is important to distinguish among these separate functions (Figure 6.2).

Classification

Classification is the assignment of objects, features, or areas to classes based on their appearance on the imagery. Often a distinction is made among three levels of confidence and precision. *Detection* is the determination of the presence or absence of a feature. *Recognition* implies a higher level of knowledge about a feature or object, such that the object can be assigned an identity in a general class or category. Finally, *identification* means that the identity of an object or feature can be specified with enough confidence and detail to place it in a specific class. Often an interpreter may qualify his or her confidence in an interpretation by specifying the identification as "possible" or "probable."

Enumeration

Enumeration is the task of listing or counting discrete items visible on an image. For example, housing units can be classified as "detached single-family home," "multifam-

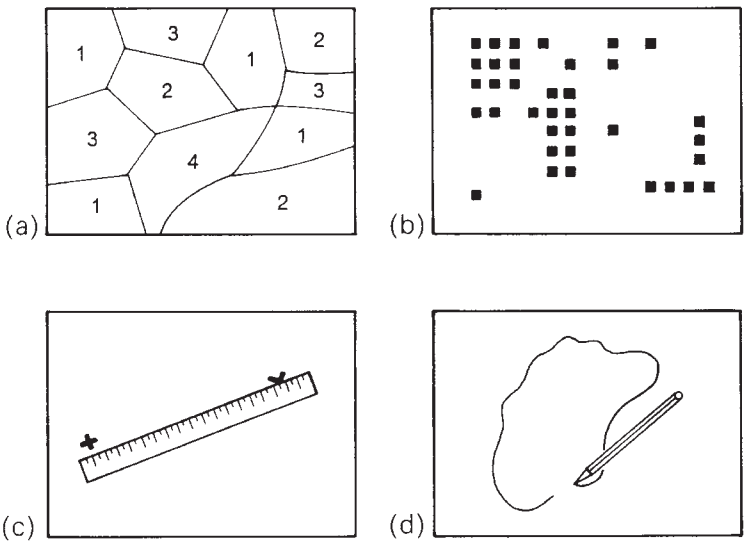


FIGURE 6.2 Image interpretation tasks: (a) classification, (b) enumeration, (c) mensuration, (d) delineation.

ily complex," "mobile home," and "multistory residential," and then reported as numbers present within a defined area. Clearly, the ability to conduct such an enumeration depends on an ability to accurately identify and classify items as discussed above.

Measurement

Measurement, or *mensuration,* is an important function in many image interpretation problems. Two kinds of measurement are important. First is the measurement of distance and height and, by extension, the measurement of volumes and areas as well. The practice of making such measurements forms the subject of *photogrammetry* (Chapter 4), which applies knowledge of image geometry to the derivation of accurate distances. Although, strictly speaking, photogrammetry applies only to measurements from photographs, by extension it has analogs for the derivation of measurements from other kinds of remotely sensed images.

A second form of measurement is quantitative assessment of image brightness. The science of *photometry* is devoted to measurement of the intensity of light and includes estimation of scene brightness by examination of image tone, using special instruments known as *densitometers*. If the measured radiation extends outside the visible spectrum, the term *radiometry* applies. Both photometry and radiometry apply similar instruments and principles, so they are closely related.

Delineation

Finally, the interpreter must often delineate, or outline, regions as they are observed on remotely sensed images. The interpreter must be able to separate distinct areal units that are characterized by specific tones and textures and to identify edges or boundaries between separate areas. Typical examples include delineation of separate classes of forest or of land use—both of which occur only as areal entities (rather than as discrete objects). Typical problems include: (1) selection of appropriate levels of generalization (e.g., when boundaries are intricate, or when many tiny but distinct parcels are present); and (2) placement of boundaries when there is a gradation (rather than a sharp edge) between two units.

The image analyst may simultaneously apply several of these skills in examining an image. Recognition, delineation, and mensuration may all be required as the interpreter examines an image. Yet specific interpretation problems may emphasize specialized skills. Military photointerpretation often depends on accurate recognition and enumeration of specific items of equipment, whereas land-use inventory emphasizes delineation, although other skills are obviously important. Image analysts therefore need to develop proficiency in all of these skills.

6.4 ELEMENTS OF IMAGE INTERPRETATION

By tradition, image interpreters are said to employ some combination of the eight *elements of image interpretation*: image tone, texture, shadow, pattern, association, shape, size, and site. These elements describe the characteristics of objects and features as they appear on remotely sensed images. Image interpreters quite clearly use these characteristics together in complex, but poorly understood, processes as they examine images.

Nonetheless, it is convenient to list them separately as a way of emphasizing their significance. Olson (1960) and Colwell (1960) have discussed the application of the elements of image interpretation to aerial photography and to other forms of aerial imagery.

Image Tone

Image tone denotes the lightness or darkness of a region within an image (Figure 6.3). For black-and-white images, tone may be characterized as "light," "medium gray," "dark gray," "dark," and so on, as the image assumes varied shades of white, gray, or black. For color or CIR imagery, image tone refers simply to "color," described informally perhaps in such terms as "dark green," "light blue," or "pale pink."

Image tone can also be influenced by the intensity and angle of illumination and by the processing of the film. Within a single aerial photograph, vignetting (Section 4.2) may create noticeable differences in image tone due solely to the position of an area within a frame of photography: The image becomes darker near the edges. Thus, the interpreter must employ caution in relying solely on image tone for an interpretation, as it can be influenced by factors other than the absolute brightness of the Earth's surface. Also, nonphotographic sensors may record such a wide range of brightness values that they cannot all be accurately represented on photographic film. In such instances, digital analyses (Chapter 5) may be more accurate.

Experiments have shown that interpreters tend to be consistent in interpretation of tones on black-and-white imagery but less so in interpretation of color imagery (Cihlar and Protz, 1972). Interpreters' assessment of image tone is much less sensitive to subtle differences in tone than are measurements by instruments (as might be expected). For the range of tones used in the experiments, human interpreters' assessment of tone expressed a linear relationship with corresponding measurements made by instruments. Cihlar and Protz's results imply that a human interpreter can provide reliable estimates of relative differences in tone but not be capable of accurate description of absolute image brightness.

Image Texture

Image texture refers to the apparent roughness or smoothness of an image region. Usually, texture is caused by the pattern of highlighted and shadowed areas created when an irregular surface is illuminated from an oblique angle. Contrasting examples (Figure 6.4) include the rough textures of a mature forest and the smooth textures of a mature wheat field. The human interpreter is very good at distinguishing subtle differences in image

FIGURE 6.3 Varied image tones, dark to light (left to right). From the U.S. Department of Agriculture (USDA).

FIGURE 6.4 Varied image textures, with descriptive terms. From USDA.

texture, so it is a valuable aid to interpretation—certainly equal in importance to image tone in many circumstances.

Image texture depends not only on the surface itself but also on the angle of illumination, so it can vary as lighting varies. Also, good rendition of texture depends on favorable image contrast, so images of poor or marginal quality may lack the distinct textural differences that are so valuable to the interpreter.

Shadow

Shadow is an especially important clue in the interpretation of objects. A building or vehicle, illuminated at an angle, casts a shadow that may reveal characteristics of its size or shape that would not be obvious from the overhead view alone (Figure 6.5). Because military photointerpreters often are primarily interested in identification of individual items of equipment, they have developed methods to use shadows to distinguish subtle differences that might not otherwise be visible. By extension, we can emphasize this role of shadow in interpretation of any human-made landscape in which identification of separate kinds of structures or objects is significant.

Shadow is also of great significance in interpretation of natural phenomena, even though its role may not be as obvious. For example, Figure 6.6 depicts an open field in which scattered shrubs and bushes are separated by areas of open land. Without shadows, the individual plants might be too small (as seen from above) and too nearly similar

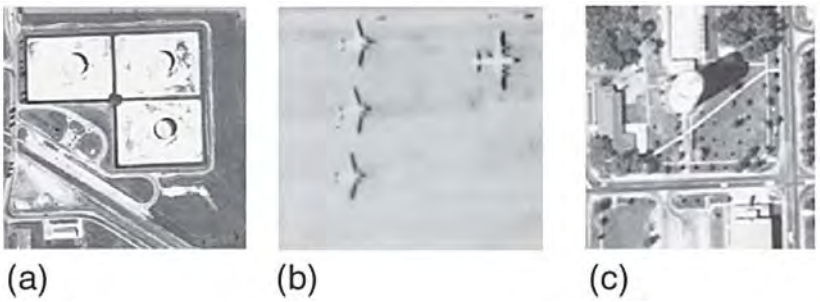


FIGURE 6.5 Examples of significance of shadow for image interpretation, as illustrated by (a) fuel storage tanks, (b) military aircraft on a runway, and (c) a water tower. From USDA.

in tone to their background to be visible. Yet their shadows are large enough and dark enough to create the streaked pattern on the imagery typical of this kind of land. A second example is also visible in Figure 6.6: at the edges between the trees in the hedgerows and the adjacent open land, trees cast shadows that form a dark strip that enhances the boundary between the two zones, as seen on the imagery.

Pattern

Pattern refers to the arrangement of individual objects into distinctive recurring forms that facilitate their recognition on aerial imagery (Figure 6.7). Pattern on an image usually follows from a functional relationship among the individual features that compose the pattern. Thus, the buildings in an industrial plant may have a distinctive pattern due to their organization to permit economical flow of materials through the plant, from receiving raw material to shipping of the finished product. The distinctive spacing of trees in an orchard arises from careful planting of trees at intervals that prevent competition between individual trees and permit convenient movement of equipment through the orchard.

Association

Association specifies the occurrence of certain objects or features, usually without the strict spatial arrangement implied by pattern. In the context of military photointerpretation, the association of specific items has great significance, as, for example, when the identification of a specific class of equipment implies that other, more important items are likely to be found nearby.

Shape

The *shapes* of features are obvious clues to their identities (Figure 6.8). For example, individual structures and vehicles have characteristic shapes that, if visible in sufficient detail, provide the basis for identification. Features in nature often have such distinctive

FIGURE 6.6 Significance of shadow for image interpretation, as illustrated by the characteristic pattern caused by shadows of shrubs cast on open field. Shadows at the edge of a forest enhance the boundary between the two different land covers. From USDA.

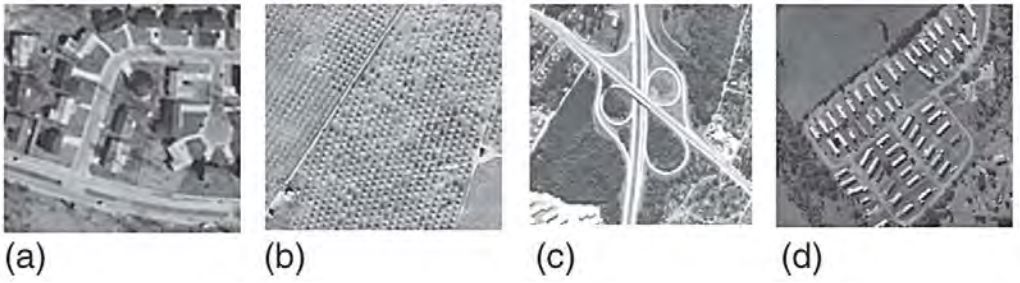


FIGURE 6.7 Significance of distinctive image patterns, as illustrated by (a) structures in a suburban residential neighborhood, (b) an orchard, (c) a highway interchange, and (d) a rural trailer park. From Virginia Department of Transportation (a, d), USDA (b), USGS (c).

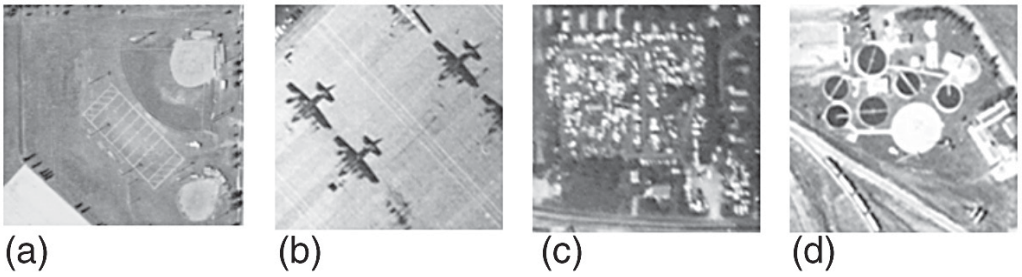


FIGURE 6.8 Significance of shape for image interpretation, as illustrated by (a) athletic fields; (b) aircraft parked on a runway, (c) automobiles in a salvage yard, and (d) a water treatment plant. From USDA.

shapes that shape alone might be sufficient to provide clear identification. For example, ponds, lakes, and rivers occur in specific shapes unlike others found in nature. Often, specific agricultural crops tend to be planted in fields that have characteristic shapes (perhaps related to the constraints of equipment used or the kind of irrigation that the farmer employs).

Size

Size is important in two ways. First, the relative size of an object or feature in relation to other objects on the image provides the interpreter with an intuitive notion of its scale and resolution, even though no measurements or calculations may have been made. This intuition is achieved via recognition of familiar objects (dwellings, highways, rivers, etc.), followed by extrapolation to use the sizes of these known features in order to estimate the sizes and identities of those objects that might not be easily identified. This is probably the most direct and important function of size.

Second, absolute measurements can be equally valuable as interpretation aids. Measurements of the size of an object can confirm its identification based on other factors, especially if its dimensions are so distinctive that they form definitive criteria for specific items or classes of items. Furthermore, absolute measurements permit derivation of quan-

titative information, including lengths, volumes, or (sometimes) even rates of movement (e.g., of vehicles or ocean waves as they are shown in successive photographs).

Site

Site refers to topographic position. For example, sewage treatment facilities are positioned at low topographic sites near streams or rivers to collect waste flowing through the system from higher locations. Orchards may be positioned at characteristic topographic sites—often on hillsides (to avoid cold air drainage to low-lying areas) or near large water bodies (to exploit cooler spring temperatures near large lakes to prevent early blossoming).

6.5 COLLATERAL INFORMATION

Collateral, or ancillary, *information* refers to non-image information used to assist in the interpretation of an image. Actually, all image interpretations use collateral information in the form of the implicit, often intuitive, knowledge that every interpreter brings to an interpretation in the form of everyday experience and formal training. In its narrower meaning, it refers to the explicit, conscious effort to employ maps, statistics, and similar material to aid in analysis of an image. In the context of image interpretation, use of collateral information is permissible, and certainly desirable, provided two conditions are satisfied. First, the use of such information is to be explicitly acknowledged in the written report, and second, the information must not be focused on a single portion of the image or map to the extent that it produces uneven detail or accuracy in the final map. For example, it would be inappropriate for an interpreter to focus on acquiring detailed knowledge of tobacco farming in an area of mixed agriculture if he or she then produced highly detailed, accurate delineations of tobacco fields but mapped other fields at lesser detail or accuracy.

Collateral information can consist of information from books, maps, statistical tables, field observations, or other sources. Written material may pertain to the specific geographic area under examination, or, if such material is unavailable, it may be appropriate to search for information pertaining to analogous areas—similar geographic regions (possibly quite distant from the area of interest) characterized by comparable ecology, soils, landforms, climate, or vegetation.

6.6 IMAGERY INTERPRETABILITY RATING SCALES

Remote sensing imagery can vary greatly in quality due to both environmental and technical conditions influencing acquisition of the data. In the United States, some governmental agencies use rating scales to evaluate the suitability of imagery for specific purposes. The National Imagery Interpretability Rating Scale (NIIRS) has been developed for single-channel and panchromatic imagery, and the Multispectral Imagery Interpretability Rating Scale (MSIIRS; Erdman et al., 1994) has been developed for multispectral imagery. Such scales are based on evaluations using a large number of experienced interpreters to independently evaluate images of varied natural and human-made features, as recorded by images of varying characteristics. They provide a guide for evaluation of whether a specific form of imagery is likely to be satisfactory for specific purposes.

6.7 IMAGE INTERPRETATION KEYS

Image interpretation keys are valuable aids for summarizing complex information portrayed as images. They have been widely used for image interpretation (e.g., Coiner and Morain, 1972). Such keys serve either or both of two purposes: (1) they are a means of training inexperienced personnel in interpreting complex or unfamiliar topics, and (2) they are a reference aid for experienced interpreters to organize information and examples pertaining to specific topics.

An *image interpretation key* is simply reference material designed to permit rapid and accurate identification of objects or features represented on aerial images. A key usually consists of two parts: (1) a collection of annotated or captioned images or stereograms, and (2) a graphic or word description, possibly including sketches or diagrams. These materials are organized in a systematic manner that permits retrieval of desired images by, for example, date, season, region, or subject.

Keys of various forms have been used for many years in the biological sciences, especially botany and zoology. These disciplines rely on complex taxonomic systems that are so extensive that even experts cannot master the entire body of knowledge. The key, therefore, is a means of organizing the essential characteristics of a topic in an orderly manner. It must be noted that scientific keys of all forms require a basic familiarity with the subject matter. A key is not a substitute for experience and knowledge but a means of systematically ordering information so that an informed user can learn it quickly.

Keys were first routinely applied to aerial images during World War II, when it was necessary to train large numbers of inexperienced photointerpreters in the identification of equipment of foreign manufacture and in the analysis of regions far removed from the experience of most interpreters. The interpretation key formed an effective way of organizing and presenting the expert knowledge of a few individuals. After the war ended, interpretation keys were applied to many other subjects, including agriculture, forestry, soils, and landforms. Their use has been extended from aerial photography to other forms of remotely sensed imagery. Today interpretation keys are still used for instruction and training, but they may have somewhat wider use as reference aids. Also, it is true that construction of a key tends to sharpen one's interpretation skills and encourages the interpreter to think more clearly about the interpretation process.

Keys designed solely for use by experts are referred to as *technical keys*. *Nontechnical keys* are those designed for use by those with a lower level of expertise. Often it is more useful to classify keys by their formats and organizations. *Essay keys* consist of extensive written descriptions, usually with annotated images as illustrations. A *file key* is essentially a personal image file with notes; its completeness reflects the interests and knowledge of the compiler. Its content and organization suit the needs of the compiler, so it may not be organized in a manner suitable for use by others.

6.8 INTERPRETIVE OVERLAYS

Often in resource-oriented interpretations, it is necessary to search for complex associations of several related factors that together define the distribution or pattern of interest. For example, soil patterns may be revealed by distinctive relationships between separate patterns of vegetation, slope, and drainage. The *interpretive overlays* approach to image interpretation is a way of deriving information from complex interrelationships between

separate distributions recorded on remotely sensed images. The correspondence between several separate patterns may reveal other patterns not directly visible on the image (Figure 6.9).

The method is applied by means of a series of individual overlays for each image to be examined. The first overlay might show the major classes of vegetation, perhaps consisting of dense forest, open forest, grassland, and wetlands. A second overlay maps slope classes, including perhaps level, gently sloping, and steep slopes. Another overlay shows the drainage pattern, and still others might show land use and geology. Thus, for each image, the interpreter may have as many as five or six overlays, each depicting a separate pattern. By superimposing these overlays, the interpreter can derive information presented by the coincidence of several patterns. From his or her knowledge of the local terrain, the interpreter may know that certain soil conditions can be expected where the steep slopes and the dense forest are found together and that others are expected where the dense forest matches to the gentle slopes. From the information presented by several patterns, the interpreter can resolve information not conveyed by any single pattern.

6.9 THE SIGNIFICANCE OF CONTEXT

It is important to consider the broader context and perspective during image interpretation, as a purely visual understanding of an image does not necessarily lead to an understanding of its underlying meaning. This topic deserves further exploration in the context of image interpretation.

Most of us are familiar with the kind of visual illusion illustrated in Figure 6.10, the *Rubin illusion*, in which the viewer sees either a white vase against a black background or two faces in silhouette facing each other against a white background. The success of the illusion depends on its ability to confuse the viewer's capacity to assess the *figure–ground relationship*. To make visual sense of an image, our visual system must decide which part

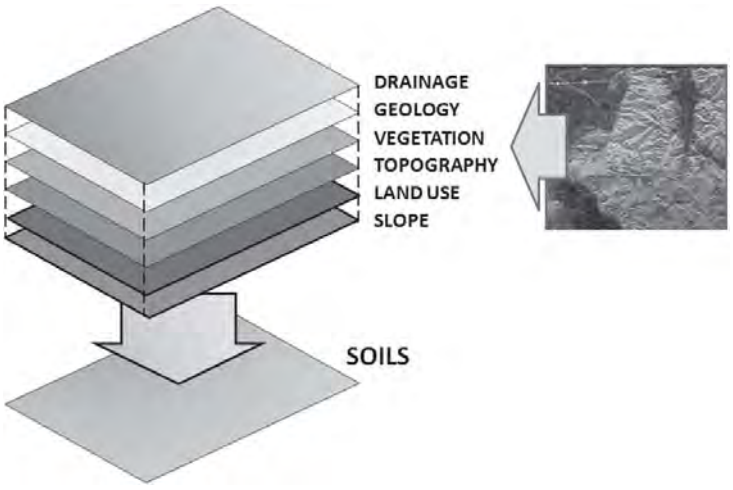


FIGURE 6.9 Interpretive overlays. Image interpretation produces several separate overlays that can combine to permit interpretation of another feature that is not directly visible on the image.

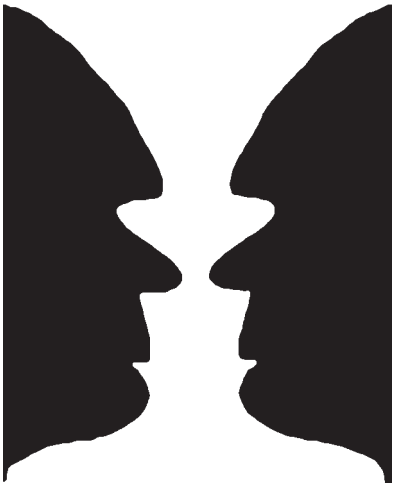


FIGURE 6.10 Rubin face/vase illusion. Because this figure (one of many such examples) is contrived to have equal balance between the white and black, the human visual system cannot decisively determine if it is viewing two black outlines of faces displayed against a white background or a white vase displayed against a black background. This effect, the *figure–ground relationship,* is an important part of how humans exact information from an image.

of a scene is the *figure* (the feature of interest) and which is the *ground* (the background that simply outlines the figure).

Normally, our visual system expects the background to constitute the larger proportion of a scene. The Rubin illusion, like most visual illusions, is effective because it is contrived to isolate the viewer's perception of the scene—in this instance, by designing the illustration so that figure and ground constitute equal proportions of the scene. The viewer's visual system cannot resolve the ambiguity, so the viewer experiences difficulty in interpreting the meaning of the scene.

Although such contrived images are not encountered in day-to-day practice, the principles that they illustrate apply to situations that are frequently encountered. For example, *relief inversion* occurs when aerial images of shadowed terrain are oriented in a manner that confuses our intuitive expectations. Normally, we expect to see terrain illuminated from the upper right (**Figure 6.11**, left); most observers see such images in their correct relief. If the image is oriented so that the illumination appears to originate from

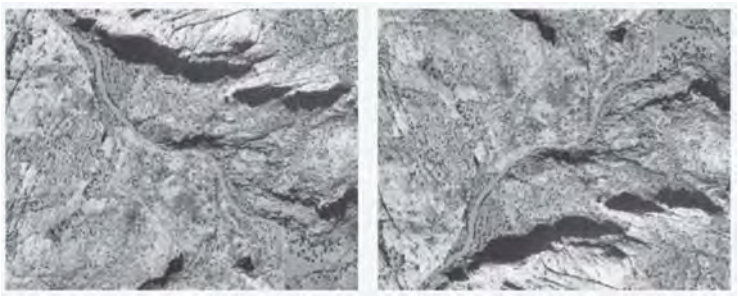


FIGURE 6.11 Observer perception of shadowing on the landscape. Photographs of landscapes with pronounced shadowing are usually perceived in correct relief when shadows fall toward the observer. Left: When shadows fall toward the observer, relief is correctly perceived. Right: When the image is rotated so that shadows fall in the opposite direction, away from the observer, topographic relief appears to be reversed. From USGS.

the lower right, most observers tend to perceive the relief as inverted (Figure 6.11, right). Experimentation with conditions that favor this effect confirms the belief that, like most illusions, relief inversion is perceived only when the context has confined the viewer's perspective to present an ambiguous visual situation.

Photointerpreters should remember that the human visual system has a powerful drive to impose its own interpretation on the neurological signals it receives from the eye and can easily create plausible interpretations of images when the evidence is uncertain, confused, or absent. Image analysts must strive always to establish *several independent lines of evidence and reasoning* to set the context that establishes the meaning of an image. When several lines of evidence and reasoning converge, then an interpretation can carry authority and credibility. When multiple lines of evidence and reasoning do not converge or are absent, then the interpretation must be regarded with caution and suspicion.

Image interpretation's successes illustrate the significance of establishing the proper context to understand the meaning of an image. The use of photointerpretation to identify the development and monitor the deployment of the German V-1 and V-2 missiles in World War II (Babington-Smith, 1957) and to identify at an early stage the deployment of Soviet missiles in Cuba during the 1962 Cuban Missile Crisis (Brugioni, 1991) was successful because it provided information that could be examined and evaluated in a broader context. Image interpretation proved to be less successful in February 2003 when U.S. Secretary of State Colin Powell presented images to the United Nations to document the case for an active threat from weapons of mass destruction in Iraq. Later, it became quite clear that there was insufficient information at hand to establish the proper meaning of the images.

6.10 STEREOVISION

Stereoscopy is the ability to derive distance information (or in the case of aerial photography, height information) from two images of the same scene. (Section 4.8 introduced the manner in which aerial cameras can collect duplicate coverage of a single region using overlapping images.) Stereovision contributes a valuable dimension to information derived from aerial photography. Full development of its concepts and techniques is encompassed in the field of photogrammetry (Wolf, 1974); here we can introduce some of its applications by describing some simple instruments.

Stereoscopes are devices that facilitate stereoscopic viewing of aerial photographs. The simplest and most common of these devices is the *pocket stereoscope* (Figure 6.12). This simple, inexpensive instrument provides an important image interpretation aid that can be employed in a wide variety of situations and introduce concepts that underlie more advanced instruments. Its compact size and inexpensive cost make it one of the most widely used remote sensing instruments, even in this era of digital imagery. The pocket stereoscope consists of a body holding two low-power lenses attached to a set of collapsible legs that can be folded so that the entire instrument can be stored in a space somewhat larger than a deck of playing cards. The body is usually formed from two separate metal pieces, each holding one of the two lenses, which can be adjusted to control the spacing between the two lenses to accommodate the individual user.

Other stereoscopes include the *mirror stereoscope* (Figure 6.13), which permits stereoscopic viewing of large areas, usually at low magnification, and the *binocular stereo-*

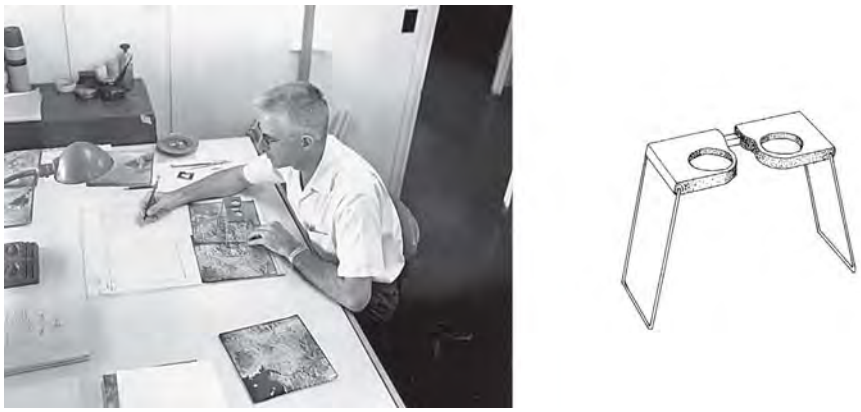


FIGURE 6.12 A USGS geologist using a pocket stereoscope to examine vertical aerial photography, 1957. (The pocket stereoscope is positioned on the desk, over the two photographs, and illustrated in the sketch at the right of the photograph.) From Photographic Library, USGS. Photograph by E. F. Patterson, no. 22.

scope (Figure 6.14), designed primarily for viewing film transparencies on light tables. Often the binocular stereoscope has adjustable magnification that enables enlargement of a portion of the image up to 20 or 40 times the size of the original.

Although at first glance the stereoscope appears designed to magnify images, magnification is really an incidental feature of the instrument. In fact, the purpose of the stereoscope is to assist the analyst in *maintaining parallel lines of sight*. Stereoscopic vision is based on the ability of our visual system to detect *stereoscopic parallax,* that is, differences in the appearance of objects from one eye to the next, which is caused by difference in perspectives. So, when we view a scene using only the right eye, we see a slightly different view than we do using only the left eye. This difference is known as stereoscopic parallax. Because stereoscopic parallax is greater for nearby objects than it is for more distant objects, our visual system can use this information to make accurate judgments about distance (see Figure 6.15).

FIGURE 6.13 Image interpretation equipment, Korean conflict, March 1952. A U.S. Air Force image interpreter uses a tube magnifier to examine detail of an aerial photograph. A mirror stereoscope, used for stereoscopic viewing of large photographs, is visible in the foreground. From U.S. Air Force, U.S. National Archives and Records Administration, ARC 542277.

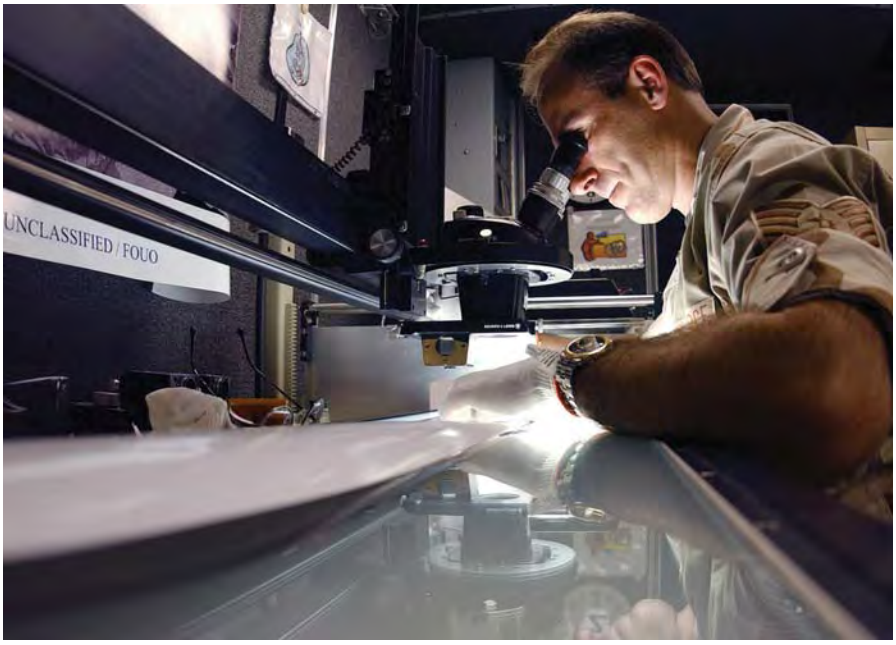


FIGURE 6.14 U.S. Air Force image analyst examining reconnaissance imagery using a binocular microscope. From USAF. Photograph by SSgt. Reynaldo Ramon, USAF.

FIGURE 6.15 Interpretation of aerial reconnaissance imagery. From U.S. Navy. Photograph by Mark J. Rebilas.

Stereoscopic aerial photographs are acquired in sequences designed to provide overlapping views of the same terrain; that is, they provide two separate perspectives of the same landscape, just as our eyes provide two separate images of a scene. We can use a stereo pair of aerial photographs to simulate a stereoscopic view of the terrain, provided we can maintain parallel lines of sight, just as we would normally do in viewing a distant object (Figure 6.16a). Parallel lines of sight ensure that the right and left eyes each see independent views of the same scene, to provide the parallax needed for the stereoscopic illusion. However, when we view objects that are nearby, our visual system instinctively recognizes that the objects are close, so our lines of sight converge (Figure 6.16b), depriving our visual system of the two independent views needed for stereoscopic vision. Therefore, the purpose of the stereoscope is to assist us in maintaining the parallel lines of sight that enable the stereoscopic effect (Figure 6.16c).

Although many students will require the assistance of the instructor as they learn to use the stereoscope, the following paragraphs may provide some assistance for beginners. First, stereophotographs must be aligned so that the flight line of the photographs passes left to right (as shown in Figure 6.16). Check the photo numbers to be sure that the photographs have been selected from adjacent positions on the flight line. Usually (but not always), the numbers and annotations on photos are placed on the *leading edge* of the image: the edge of the image nearest the front of the aircraft at the time the image was taken. Therefore, these numbers should usually be oriented in sequence from left to right, as shown in Figures 6.12 and 6.15. If the overlap between adjacent photos does not correspond to the natural positions of objects on the ground, then the photographs are incorrectly oriented.

Next, the interpreter should identify a distinctive feature on the image within the zone of stereoscopic overlap. The photos should then be positioned so that the duplicate images of this feature (one on each image) are approximately 64 mm (2.5 in.) apart. This distance represents the distance between the two pupils of a person of average size (referred to as the *interpupillary distance*). For many, however, it may be a bit too large or too small, so the spacing of photographs may require adjustment as the interpreter follows the procedure outlined here. The pocket stereoscope should be opened so that its

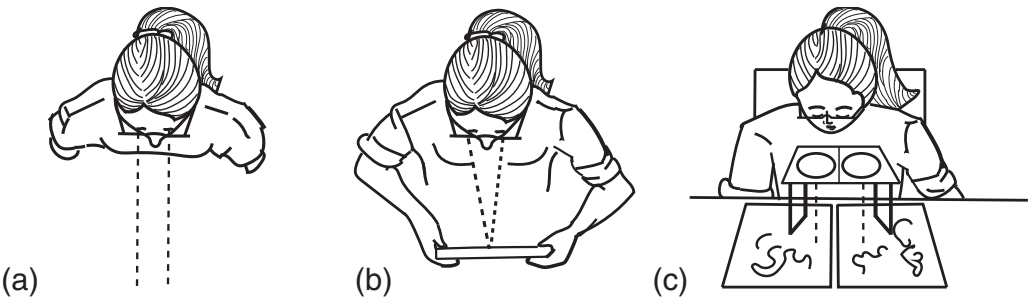


FIGURE 6.16 The role of the stereoscope in stereoscopic vision. (a) To acquire the two independent views of the same scene required for stereoscopic vision, we must maintain parallel lines of sight. (b) Normally, when we view nearby objects, our lines of sight converge, preventing us from acquiring the stereo effect. (c) The stereoscope is an aid to assist in maintaining parallel lines of sight even when the photographs are only a few inches away from the viewer. Image by Susmita Sen.

legs are locked in place to position the lens at their correct height above the photographs. The two segments of the body of the stereoscope should be adjusted so that centers of the eyepieces are about 64 mm (2.5 in) apart (or a slightly larger or smaller distance, as mentioned above).

Then the stereoscope should be positioned so that centers of the lenses are above the duplicate images of the distinctive feature selected previously. Looking through the two lenses, the analyst sees two images of this feature; if the images are properly positioned, the two images will appear to "float" or "drift." The analyst can, with some effort, control the apparent positions of the two images so that they fuse into a single image. As this occurs, the two images should merge into a single image that is then visible in three dimensions. Usually aerial photos show exaggerated heights, due to the large separation (relative to distance to the ground) between successive photographs as they were taken along the flight line. Although exaggerated heights can prevent convenient stereo viewing in regions of high relief, it can be useful in interpretations of subtle terrain features that might not otherwise be noticeable. The student who has successfully used the stereoscope to examine a section of the photo should then practice moving the stereoscope over the image to view the entire region within the zone of overlap. (As long as the axis of the stereoscope is oriented parallel to the flight line, it is possible to retain stereo vision while moving the stereoscope.) If the stereoscope is not aligned with respect to the flight line, the interpreter loses stereo vision. By lifting the edge of one of the photographs, it is possible to view in stereo image regions near edges of the photos. Although the pocket stereoscope is valuable for examining terrain, drainage, and vegetation patterns, it does not provide the detailed measurements within the realm of photogrammetry and more sophisticated instruments.

The stereoscope is only one of several devices designed to present separate images intended to create a stereo effect using the *optical separation* technique. Left and right images are presented side by side, with an optical device designed to separate the analyst's view of the left and right images. The *red/blue anaglyph* presents images intended for each eye in separate colors, blues for the left eye, reds for the right eye, and shades of magenta for those portions of the image common to both eyes. The analyst views the image using glasses with a red lens for the left eye and blue for the right eye. The colored lenses cause the image intended for the other eye to blend into the background; the image intended for its own eye will appear as black.

The anaglyph has been widely used for novelties, less often as an analytical device (see **Figure 6.17** and *https://serc.carleton.edu/download/images/8512/jge-jan07.jpg*).

The use of *polarized lenses* for stereovision is based on the projection of images for each eye through separate polarizing filters (e.g., horizontal for the left eye, vertical for the right eye). The combined image must be viewed through special glasses that use orthogonal polarizations for the left and right lenses. This technique is one of the most effective means of stereoviewing for instructional and analytical applications. A proprietary variation of this technique (CrystalEyes) displays left- and right-eye views of a digital image in sequential refresh scans on a monitor, then uses synchronized polarized shutter glasses to channel the correct image to the correct eye. This technique forms the basis for stereographic images in many virtual reality display environments. Many other techniques are effective to varying degrees for stereovision (including random dot stereograms, Magic Eye images, and others). Most of them are less effective for scientific and analytical applications than stereoscopes and polarized lenses.

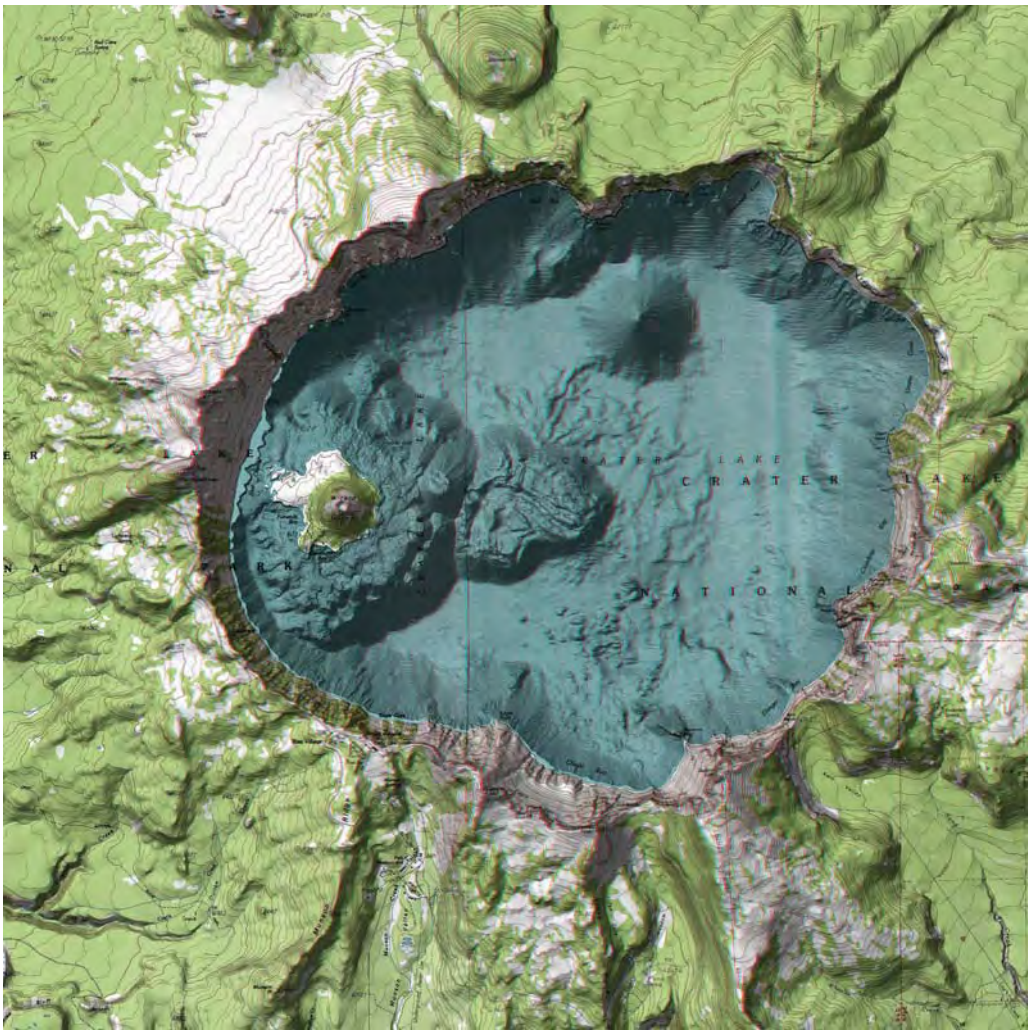


FIGURE 6.17 Anaglyph stereo maps of topographic surfaces. Crater Lake bathymetry produced by James V. Gardner, Larry A. Mayer, and Mark Butenica as merged into USGS topographic data. (Red/blue anaglyph glasses required to view these scenes in three dimensions.) Used by permission.

6.11 DIGITAL PHOTOINTERPRETATION

Increasing use of digital photography and softcopy photogrammetry (Section 4.9) has blurred a previously distinct separation between manual and digital photointerpretation. Analyses that previously were conducted by visual examination of photographic prints or transparencies can now be completed by examination of digital images viewed on computer screens. Analysts record the results of their interpretations as onscreen annotations, using the mouse and cursor to outline and label images. Figure 6.18 illustrates a digital record of interpreted boundaries recorded by onscreen digitization (left) and the outlines shown without the image backdrop (center). The right-hand image shows an enlargement

FIGURE 6.18 A digital record of image interpretation, depicting outlines traced by the analyst using onscreen digitization (left), outlined without the image backdrop (center), and detail of the raster structure of the digital outline (right).

of a portion of the labeled region, illustrating the raster structure of the image and land-use boundaries.

Some systems employ photogrammetric software to project image detail in its correct planimetric location, without the positional or scale errors that might be present in the original imagery. Furthermore, the digital format enables the analyst to easily manipulate image contrast to improve interpretability of image detail. Digital photogrammetric workstations (Figure 6.19), often based on the usual PC or UNIX operating systems, can accept scanned film imagery, airborne digital imagery, or digital satellite data. The full

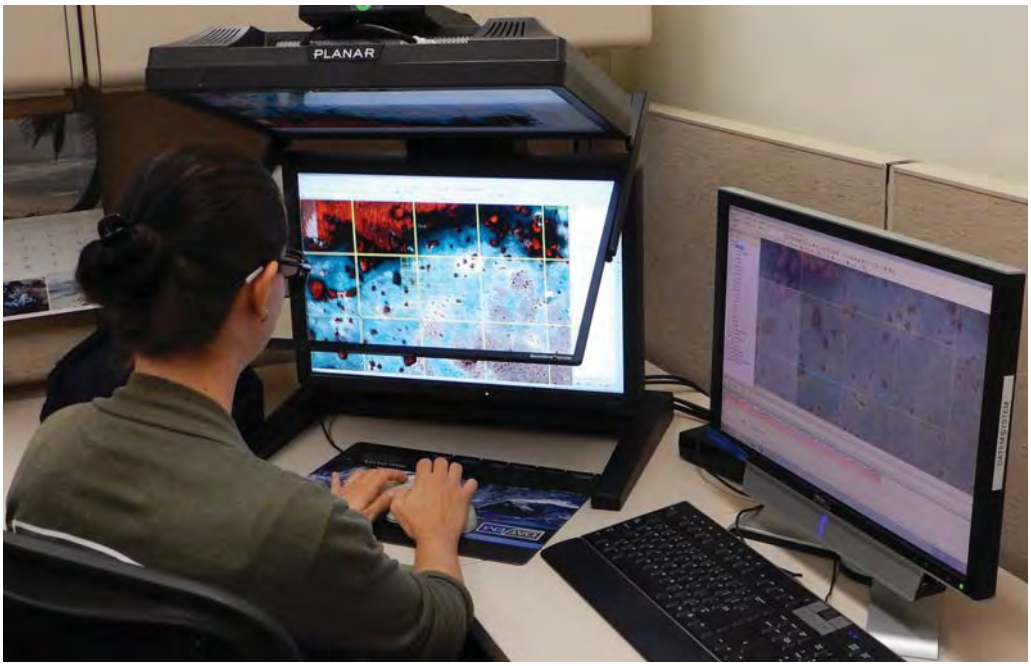


FIGURE 6.19 USGS photointerpreter using a 3D photogrammetric workstation to prepare multi-dimensional representation of terrain. From USGS Earth Resources Observation and Science Center.

range of photogrammetric processes can be implemented digitally, including triangulation, compilation of digital terrain models (DTMs), feature digitization, construction of orthophotos, mosaics, and flythroughs. Analysts can digitize features onscreen ("heads-up" digitization), using the computer mouse, to record and label features in digital format.

6.12 IMAGE SCALE CALCULATIONS

Scale is a property of all images. Knowledge of image scale is essential for making measurements from images and for understanding the geometric errors present in all remotely sensed images. *Scale* is an expression of the relationship of the *image distance* between two points and the *actual distance* between the two corresponding points on the ground. This relationship can be expressed in several ways.

The *word statement* sets a unit distance on the map or photograph equal to the correct corresponding distance on the ground—for example, "One inch equals 1 mi" or, just as correctly, "One centimeter equals 5 km." The first unit in the expression specifies the map distance, and the second is the corresponding ground distance. Another method of specifying scale is the *bar scale,* which simply labels a line with subdivisions that show ground distances. Yet another method, the *representative fraction* (RF), is more widely used and often forms the preferred method of reporting image scale. The RF is the ratio between image distance and ground distance. It usually takes the form "1:50,000" or "1/50,000," with the numerator set equal to 1 and the denominator equal to the corresponding ground distance.

The RF has meaning in any unit of length as long as both the numerator and the denominator are expressed in the same units. Thus, "1:50,000" can mean "1 in. on the image equals 50,000 in. on the ground" or "1 cm on the image equals 50,000 cm on the ground." A frequent source of confusion is converting the denominator into the larger units that we find more convenient to use for measuring large ground distance. With metric units, the conversion is usually simple. In the example given above, it is easy to see that 50,000 cm is equal to 0.50 km and that 1 cm on the map represents 0.5 km on the ground. With English units, the same process is not quite so easy. It is necessary to convert inches to miles to derive "1 in. equals 0.79 mi" from 1:50,000. For this reason, it is useful to know that 1 mi equals 63,360 in. Thus, 50,000 in. is equal to 50,000/63,360 = 0.79 mi.

A typical scale problem requires estimation of the scale of an individual photograph. One method is to use the focal length and altitude method (Figure 6.19):

$$RF = \frac{\text{Focal length}}{\text{Altitude}} \qquad \text{(EQ. 6.1)}$$

Both values must be expressed in the same units. Thus, if a camera with a 6-in. focal length is flown at 10,000 ft, the scale is 0.5/10,000 = 1:20,000. (Altitude always specifies the flying height above the terrain, *not* above sea level.) Because a given flying altitude is seldom the exact altitude at the time the photography was done, and because of the several sources that contribute to scale variations within a given photograph (Chapter 4), we must always regard the results of such calculations as an approximation of the scale of any specific portion of the image. Often such values are referred to as the "nominal" scale of an image, meaning that it is recognized that the stated scale is an approximation and that image scale will vary within any given photograph.

Another way to calculate the RF is to use a *known ground distance*. In instances in which accurate maps of the area represented on the photograph may not be available, the interpreter may not know focal length and altitude. Then, if possible, an approximation of image scale can be made to identify an object or feature of known dimensions. Such features might include a football field or baseball diamond; measurement of a distance from these features as they are shown on the image provides the "image distance" value needed to use the relationship given above. The "ground distance" is derived from our knowledge of the length of a football field or the distance between bases on a baseball diamond. Some photointerpretation manuals provide tables of standard dimensions of features commonly observed on aerial images, including sizes of athletic fields (soccer, field hockey, etc.), lengths of railroad boxcars, distances between telephone poles, and so on, as a means of using the known ground distance method. Similarly, we can identify two points at the same elevation on the aerial photograph that are also represented on a map. For example, in Figure 6.20, the image distance between points *A* and *B* is measured to be approximately 2.2 in. (5.6 cm). From the map, the same distance is determined to correspond to a ground distance of 115,000 in. (about 1.82 mi). Thus, the scale is found to be:

$$RF = \frac{\text{Focal length}}{\text{Ground distance}} = \frac{2.2 \text{ in.}}{1.82 \text{ mi.}} = \frac{2.2 \text{ in.}}{115{,}000 \text{ in.}} = \frac{1}{52{,}273} \qquad \text{(EQ. 6.2)}$$

A second kind of scale problem is the use of a known scale to measure a distance on the photograph. Such a distance might separate two objects on the photograph but not

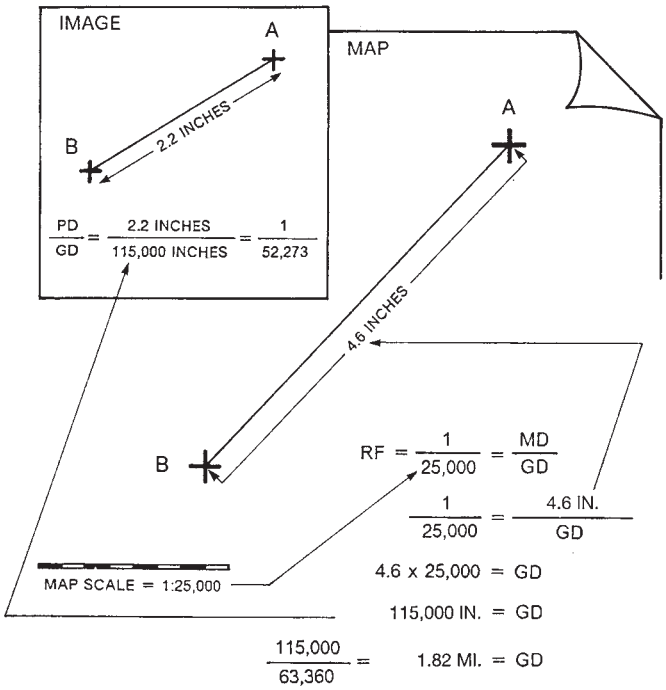


FIGURE 6.20 Measurement of image scale using a map to derive ground distance.

be represented on the map, or the size of a feature may have changed since the map was compiled. For example, we know that image scale is 1:15,000. A pond not shown on the map is measured on the image as 0.12 in. in width. Therefore, we can estimate the actual width of the pond to be:

$$\begin{aligned} \frac{1}{15{,}000} &= \frac{\text{Image distance}}{\text{Ground distance}} \\ \frac{0.12 \text{ in.}}{\text{Unknown } GD} &= \frac{\text{Image distance}}{\text{Ground distance}} \\ GD &= 0.12 \times 15{,}000 \text{ in.} \\ GD &= 1{,}800 \text{ in., or } 150 \text{ ft} \end{aligned} \qquad \text{(EQ. 6.3)}$$

This example can illustrate two other points. First, because image scale varies throughout the image, we cannot be absolutely confident that our distance for the width of the pond is accurate; it is simply an estimate, unless we have high confidence in our measurements and in the image scale at this portion of the photo. Second, measurements of short image distances are likely to have errors simply due to our inability to make accurate measurements of very short distances (e.g., the 0.12 in. distance measured above). As distances become shorter, our errors constitute a greater proportion of the estimated length. Thus, an error of 0.005 in. is 0.08% of a distance of 6 in. but 4% of the distance of 0.12 in. mentioned above. Accordingly, the interpreter should exercise a healthy skepticism regarding measurements made from images unless he or she has taken great care to ensure maximum accuracy and consistency.

6.13 SUMMARY

Image interpretation was once practiced entirely within the realm of photographic prints and transparencies, using the equipment and techniques outlined in the preceding sections. As digital analyses have increased in significance, so has the interpretation of imagery presented on computer displays. Although such interpretations are based on the same principles outlined here for traditional imagery, digital data have their own characteristics that require special treatment in the context of visual interpretation.

Despite the increasing significance of digital analysis in all aspects of remote sensing, image interpretation still forms a key component in the way that humans understand images. Analysts must evaluate imagery, either as paper prints or as displays on a computer monitor, using the skills outlined in this chapter. The fundamentals of manual image interpretation were developed for application to aerial photographs at an early date in the history of aerial survey, although it was not until the 1940s and 1950s that they were formalized in their present form. Since then, these techniques have been applied, without substantial modification, to other kinds of remote sensing imagery. As a result, we have a long record of experience in their application and comprehensive knowledge of their advantages and limitations.

Interesting questions remain. In what ways might image interpretation skills be modified in the context of interpretation using computer monitors? What new skills might be

necessary? How have analysts already adjusted to new conditions? How might equipment and software be improved to facilitate interpretation in this new context?

SOME TEACHING AND LEARNING RESOURCES

- Introduction to Photo Interpretation
  www.youtube.com/watch?v=LlBDGBopt_g
  This 1955 film has separate sections addressing photointerpretation for hydrology, soils, geology, and forestry. Although the production methods and presentation are clearly dated, many of the basic principles are effectively presented. Many viewers can safely skip the introduction, which concludes at about 4:30.
- Map and Compass Basics: Understanding Map Scale
  www.youtube.com/watch?v=jC1w2jb13GQ
- Map Reading: Understanding Scale
  www.youtube.com/watch?v=93xYDoEA7CQ&feature=related

REVIEW QUESTIONS

1. A vertical aerial photograph was acquired using a camera with a 9-in. focal length at an altitude of 15,000 ft. Calculate the nominal scale of the photograph.
2. A vertical aerial photograph shows two objects to be separated by 6¾ in. The corresponding ground distance is 9½ mi. Calculate the nominal scale of the photograph.
3. A vertical aerial photograph shows two features to be separated by 4.5 in. A map at 1:24,000 shows the same two features separated by 9.3 in. Calculate the scale of the photograph.
4. Calculate the area represented by a 9 in. × 9 in. vertical aerial photograph taken at an altitude of 10,000 ft using a camera with a 6-in. focal length.
5. You plan to acquire coverage of a county using a camera with 6-in. focal length and a 9 in. × 9 in. format. You require an image scale of 4 in. equal to 1 mi, 60% forward overlap, and sidelap of 10%. Your county is square in shape, measuring 15.5 mi on a side. How many photographs are required? At what altitude must the aircraft fly to acquire these photos?
6. You have a flight line of 9 in. × 9 in. vertical aerial photographs taken by camera with a 9-in. focal length at an altitude of 12,000 ft above the terrain. Forward overlap is 60%. Calculate the distance (in miles) between ground nadirs of successive photographs.
7. You require complete stereographic coverage of your study area, which is a rectangle measuring 1.5 mi × 8 mi. How many 9 in. × 9 in. vertical aerial photographs at 1:10,000 are required?
8. You need to calculate the scale of a vertical aerial photograph. Your estimate of the ground distance is 2.8 km. Your measurement of the corresponding image distance is 10.4 cm. What is your estimate of the image scale?
9. You have little information available to estimate the scale of a vertical aerial photograph, but you are able to recognize a baseball diamond among features in an athletic

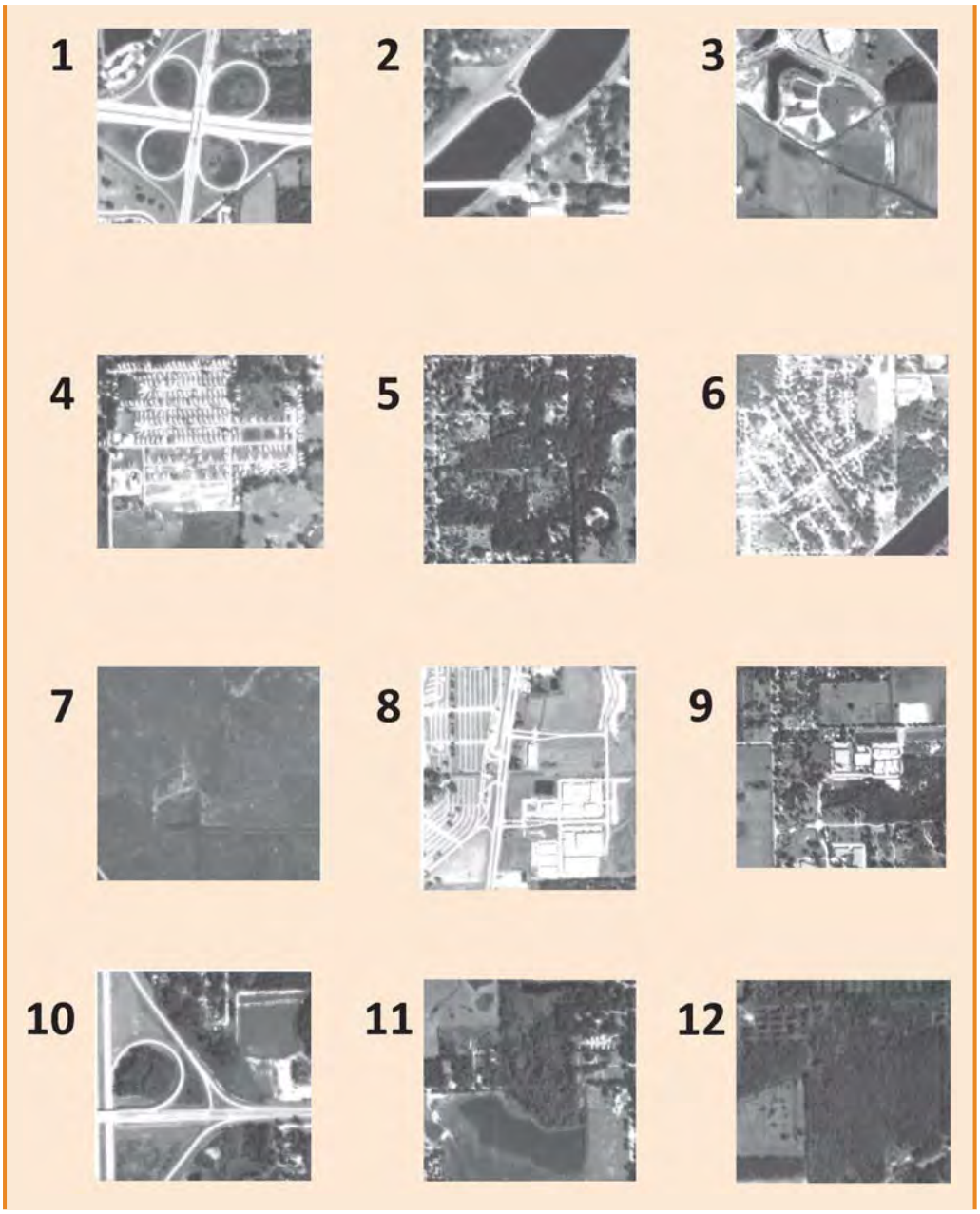
complex. You use a tube magnifier to measure the distance between first and second base, which is 0.006 ft. What is your estimate of the scale of the photo?

10. Assume you can easily make an error of 0.001 in your measurement for Question 9. Recalculate the image scale to estimate the range of results produced by this level of error. Now return to Question 3 and assume that the same measurement error applies. (Do not forget to consider the different measurement units in the two questions.) Calculate the effect on your estimates of the image scale. The results should illustrate why it is always better whenever possible to use long distances to estimate image scale.

11. **Visual search exercise.** This exercise develops skills in matching patterns and developing a sense of spatial context for image interpretation. Examine each of the smaller images, and match it to the correct location on the source image shown below, using the coordinates marked at the edge of the source image. As an example, image 1 has been identified.

Answers:

1. B-3	3. ___	5. ___	7. ___	9. ___	11. ___
2. ___	4. ___	6. ___	8. ___	10. ___	12. ___


1
2
3
4
5
6
7
8
9
10
11
12

12. **Identification exercise.** Identify the principal features depicted in each image, using the elements of image interpretation listed in the text. Be prepared to identify the key elements important for each image.

1 

2 

3 

4 

5 

6 

7 

8 

9 

10 

11 

12 

13 

14 

15 

16 

17 

18 

REFERENCES

Avery, T. E., and G. L. Berlin. 2003. *Fundamentals of Remote Sensing and Airphoto Interpretation* (6th ed.). New York: Macmillan, 540 pp.

Babington-Smith, C. 1957. Air Spy: The Story of Photo Interpretation in World War II. New York: Ballantine, 190 pp. (Reprinted 1985. Bethesda, MD: American Society for Photogrammetry and Remote Sensing, 266 pp.)

Brugioni, D. 1991. *Eyeball to Eyeball: The Inside History of the Cuban Missile Crisis*. New York: Random House, 622 pp.

Brugioni, D. 1996. The Art and Science of Photoreconnaissance. *Scientific American,* Vol. 274, pp. 78–85.

Brugioni, D. A., and F. J. Doyle. 1997. Arthur C. Lundahl: Founder of the Image Exploitation Discipline. In R. McDonald (Ed.), *Corona between the Sun and the Earth: The First NRO Reconnaissance Eye in Space* (pp. 159–166). Bethesda, MD: American Society for Photogrammetry and Remote Sensing.

Campbell, J. B. 2005. Visual Interpretation of Aerial Imagery. Chapter 10 in *Remote Sensing for GIS Managers* (S. Aronoff, ed.). Redlands, CA: ESRI Press, pp. 259–283.

Campbell, J. B. 2008. Origins of Aerial Photographic Interpretation, U.S. Army, 1916 to 1918. *Photogrammetric Engineering and Remote Sensing,* Vol. 74, pp. 77–93.

Campbell, J. B. 2010. Information Extraction from Remotely Sensed Data. Chapter 19 in *Manual of Geospatial Science and Technology* (J. D. Bossler, ed.). Boca Raton, FL: CRC Press, pp. 363–389.

Cihlar, J., and R. Protz. 1972. Perception of Tone Differences from Film Transparencies. *Photogrammetria,* Vol. 8, pp. 131–140.

Coburn, C., A. Roberts, and K. Bach. 2001. Spectral and Spatial Artifacts from the Use of Desktop Scanners for Remote Sensing. *International Journal of Remote Sensing,* Vol. 22, pp. 3863–3870.

Coiner, J. C., and S. A. Morain. 1972. SLAR Image Interpretation Keys for Geographic Analysis (Technical Report 177–19). Lawrence, KS: Center for Research, 110 pp.

Colwell, R. N. (ed.). 1960. *Manual of Photographic Interpretation*. Falls Church, VA: American Society of Photogrammetry, 868 pp.

Colwell, R. N. 1977. *Manual of Photographic Interpretation*. Washington, DC: American Society for Photogrammetry and Remote Sensing.

Erdman, C., K. Riehl, L. Mayer, J. Leachtenauer, E. Mohr, J. Odenweller, . . . and D. Hothem. 1994. Quantifying Multispectral Imagery Interpretability. In *International Symposium on Spectra Sensing Research,* Vol. 1, pp. 468–476. Alexandria, VA: U.S. Corps of Engineers.

Goddard, G. C. (with D. S. Copp). 1969. *Overview: A Lifelong Adventure in Aerial Photography*. Garden City, NY: Doubleday, 415 pp.

Gorrell, E. S. 1917–1919. *Gorrell's History of the U.S. Army Air Service*. Microfilmed typescript, T-619 (58 rolls). College Park, MD: U.S. National Archives and Records Administration.

Haack, B., and S. Jampoler. 1995. Colour Composite Comparisons for Agricultural Assessments. *International Journal of Remote Sensing,* Vol. 16, pp. 1589–1598.

Hamburger, K. E. 1997. *Learning Lessons in the American Expeditionary Forces*. U.S. Army Center of Military History, CMH Publication 24–1, 28 pp.

Jensen, J. R. 2000. *Remote Sensing of the Environment*. New York: Prentice Hall, 592 pp.

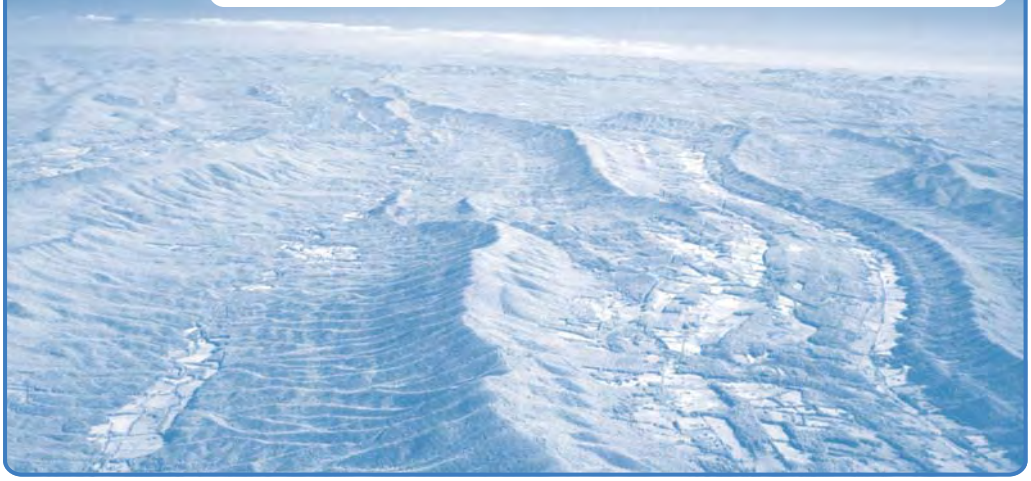
Lee, W. T. 1922. *The Face of the Earth as Seen from the Air*. Special Publication No. 4. New York: American Geographical Society.

Lillesand, T., R. Keifer, and J. W. Chipman. 2008. *Remote Sensing and Image Interpretation* (6th ed.). New York: John Wiley, 756 pp.

Lueder, D. R. 1959. *Aerial Photographic Interpretation: Principles and Applications*. New York: McGraw-Hill, 462 pp.

Melton, F. A. 1939. Aerial Photographs and the First Course in Geology. *Photogrammetric Engineering*, Vol. 5, pp. 74–77.
O'Conner, J. 2015. *NPIC: Seeing the Secrets and Growing the Leaders*. Alexandria, VA: Acumensa Solutions.
Olson, C. E. 1960. Elements of Photographic Interpretation Common to Several Sensors. *Photogrammetric Engineering*, Vol. 26, No. 4, pp. 651–656.
Paine, D. P., and J. D. Kiser. 2003. *Aerial Photography and Image Interpretation*. New York: John Wiley, 648 pp.
Philipson, W. R. (ed.). 1996. *Manual of Photographic Interpretation* (2nd ed.). Bethesda, MD: American Society for Photogrammetry and Remote Sensing, 689 pp.
Rapp, D. N., S. A. Culpeper, K. Kirkby, and P. Morin. 2007. *Journal of Geoscience Education*. Vol. 55, No. 1, pp. 5–16.
Smith, H. T. U. 1942. Aerial Photographs in Geomorphic Studies. *Photogrammetric Engineering*. Vol. 8, pp. 129–155.
Smith, H. T. U. 1943. *Aerial Photographs and Their Applications*. New York: Century Crofts, 372 pp.
Stone, K. H. 1964, September. A Guide to the Interpretation and Analysis of Aerial Photos. *Annals of the Association of American Geographers*, Vol. 54, No. 3, pp. 318–328.
Wolf, P. R. 1974. *Elements of Photogrammetry*. New York: McGraw-Hill, 562 pp.

7 Land Observation Satellites



MAJOR TOPICS TO UNDERSTAND

- Current Satellite Systems
- Landsat Origins and Evolution
- Landsat-Like Systems
- Broad-Scale/Coarse Resolution Satellites
- Fine-Resolution Satellite Systems
- SmallSats
- Land Observation Satellite Orbits
- Data Archives and Indexing Systems for Land Observation Satellites
- International Charter

7.1 INTRODUCTION

Satellite remote sensing was first introduced in Chapter 3. Today, many nations, and a broad range of corporations, operate satellite remote sensing systems specifically

designed for observation of the Earth's surface to collect information concerning crops, forests, water bodies, land use, cities, and mineral resources. As first described in Section 3.5, satellite platforms offer numerous advantages, including their *synoptic perspective* (observation of large areas in a single image), fine detail, and systematic, repetitive coverage. Section 3.5 describes the basic components of satellite systems and their orbits. Here we discuss the evolution of some of the key satellite systems that have contributed to our understanding of the Earth's systems and human impacts on the environment. Due to the vast number of satellite observation systems in use and rapid changes in their design, it is not feasible to discuss them all. Nonetheless, an examination of the past, present, and future of key systems can provide readers with a basic framework to assist in understanding key aspects of Earth observation satellites (EOS). Thus, this chapter outlines essential characteristics of some of the most important systems as a basis for understanding other systems not specifically discussed here. We focus on satellite systems that support optical sensors; subsequent chapters introduce other satellite systems that collect microwave and thermal imagery.

7.2 CURRENT SATELLITE SYSTEMS

Since the 1980s, the number and varied characteristics of such systems has increased so rapidly that it is impractical to list them, let alone describe their characteristics. Understanding Earth observation satellite systems is facilitated by considering them as members of four families of satellites.

1. The first group consists of *Landsat* and *Landsat-like systems,* designed for acquisition of rather broad geographic coverage at moderate levels of detail. Data from these systems have been used for an amazingly broad range of applications, which can be generally described as focusing on surveying and monitoring land and water resources. The 30-m pixel size of Landsat has proven useful for recording imagery that permits examination of patterns that match to scales that reflect how humans occupy the Earth's surface (agricultural patterns, water resources, urban systems, coastal landforms, forest resources, etc.).
2. A second group is formed by those satellite observation systems designed to acquire *very broad-scale images at coarse resolutions,* intended in part to acquire images that can be aggregated to provide continental or global coverage. Such images enable scientists to monitor broad-scale environmental dynamics. We can think of these systems as providing imagery with pixels several kilometers in size, including some that extend to many tens of kilometers in size.
3. A third family of satellite systems provides *very fine detail of small regions* to acquire imagery that might assist in urban planning, design of highway or pipeline routes, or precision agriculture, for example. These satellites typically provide fine spatial detail (small-footprint images) at about 10-m detail or finer, reaching submeter detail in some instances.
4. A fourth category of satellites, *SmallSats,* consists of satellite systems defined by their very small sizes and by characteristics that derive from their low costs, small dimensions, light weights, and economical designs and operational costs.

Although this categorization is imperfect, it does provide a framework that helps explain the capabilities of a very large number of satellite systems now available. For systems not discussed here, Stoney (2008) provides a comprehensive catalog of earlier land observation satellite systems.

7.3 LANDSAT ORIGINS AND EVOLUTION

Today's land observation satellites have developed from earlier systems. Landsat ("land satellite"), designed in the 1960s and launched in 1972, was the first satellite tailored specifically for broad-scale observation of the Earth's land areas, accomplishing for land resource studies what meteorological satellites had accomplished for meteorology and climatology. Today, the Landsat system is important in its own right—both as an innovative remote sensing system that has contributed greatly to Earth resource studies and as a template for similar land observation satellites operated by national governments and private enterprises throughout the world.

Landsat was proposed by scientists and administrators in the U.S. government who envisioned application of basic principles of remote sensing to broad-scale, repetitive survey of the Earth's land areas. This system has since supported many years of satellite observation of the Earth's land areas, surveying both natural and human facets of the Earth's surface as it changes. Here, we offer a concise introduction to Landsat's history and some of its applications. We encourage readers to consult *Landsat's Enduring Legacy* (Landsat Legacy Project Team, 2017), a thoughtful, comprehensive review of Landsat and its applications.

The first Landsat sensors recorded energy in the visible and near-infrared spectra. Although these regions of the spectrum had long been used for aircraft photography, it was by no means certain that they would also prove practical for observation of Earth resources from satellite altitudes. Scientists and engineers were not completely confident that the sensors would work as planned, that they would prove to be reliable, that detail would be satisfactory, or that a sufficient proportion of scenes would be free of cloud cover. Although many of these problems were encountered, the feasibility of the basic concept was soon demonstrated, and Landsat has formed the model for similar systems now operated by other organizations throughout the world.

The Landsat system consists of spacecraft-borne sensors that observe the Earth and then transmit information by microwave signals to ground stations that receive and process data for dissemination to a community of data users. The first Landsat vehicles carried two sensor systems: (1) the *Return Beam Vidicon* (*RBV*) and (2) the *Multispectral Scanner System* (*MSS*) (Table 7.1). The RBV camera, designed as the primary sensor for Landsat 1 in 1972, generated high-resolution, television-like, images of the Earth's surface. Its significance for our discussion is that it was designed to apply then-current remote sensing technologies for use from orbital, rather than aircraft, altitudes. It provided three spectral channels, in the green, red, and near infrared, to replicate information conveyed by color infrared films, valued then as the most effective aerial film. The RBV was designed to provide a camera-like perspective, using a shutter to provide an electronic image that could be analyzed using photogrammetic principles, in much the same way that photogrammetry had been used to analyze aerial photographs acquired at aircraft altitudes. However, once the RBV instrument attained orbit, analysts discovered

TABLE 7.1 Landsat Missions

Satellite	Launched	End of service[a]	Principal sensors[b]
Landsat 1	July 23, 1972	January 6, 1978	MSS, RBV
Landsat 2	January 22, 1975	January 25, 1982	MSS, RBV
Landsat 3	March 5, 1978	March 3, 1983	MSS, RBV
Landsat 4	July 19, 1982	[c]	TM, MSS
Landsat 5	March 1, 1984	June 5, 2013	TM, MSS
Landsat 6	October 5, 1993	Lost at launch	ETM
Landsat 7	April 15, 1999	[d]	ETM+
Landsat 8	February 11, 2013		OLI, TIRS

Note: See *www.usgs.gov/land-resources/nli/landsat/landsat-satellite-missions* for a complete chronology.
[a]Satellite systems typically operate on an intermittent or standby basis for considerable periods prior to formal retirement from space.
[b]Sensors are discussed in Section 7.3. MSS = Multispectral Scanner System; RBV = Return Beam Vidicon; TM = Thematic Mapper; ETM = Enhanced Thematic Mapper; ETM+ = Enhanced Thematic Mapper Plus; OLI = Operational Land Imager; TIRS = thermal infrared sensor.
[c]Transmission of TM data failed in August 1993.
[d]Malfunction of TM scan line corrector has caused data gaps in the imagery since May 2003.

that the RBV did not provide reliable imagery, leaving success of the mission with its remaining sensor—the *multispectral scanner*.

As a result of the unexpected unreliable performance of the RBV, the MSS became the primary Landsat sensor. Its design originated as an unsolicited proposal to NASA from Hughes Aerospace Corporation, prompted by the work of Virginia Norwood, a Hughes Aerospace Engineer (**Figure 1.15**), who proposed using a nascent digital scanning technology for the first Landsat mission (**Figure 7.1**), even though its design was as yet unproven in the context of its proposed use as an orbital sensor. In contrast to the RBV, the MSS was designed to provide finer detail concerning spectral characteristics of the Earth but less accurate positional detail. The IFOV was 68 × 83 m, which gives an area somewhat less than that of a U.S. football field. There are four spectral channels, located in the green, red, and near-infrared portions of the electromagnetic spectrum (Band 1: 0.5–0.6 μm; Band 2: 0.6–0.7 μm; Band 3: 0.7–0.8 μm; Band 4: 0.8–1.1 μm).

The MSS then formed the operational core of the Landsat system and introduced the remote sensing community to a new imaging system, which required rapid innovation to develop and evaluate new analytical strategies for image analysis. In subsequent years, routine availability of MSS digital data formed a foundation for a dramatic increase in the number and sophistication of digital image processing capabilities available to the remote sensing community. Although the MSS has since been replaced by more advanced systems, it remains a foundation of remote sensing because it introduced, and established the validity of, concepts used later for more complex imaging systems. Furthermore, it has an important historic significance, as the techniques developed by scientists to analyze MSS imagery form the origins of the practice of digital image processing, which is now the foundation for current strategies for examining satellite imagery.

Subsequent missions improved on the basic design of the MSS with a new instrument known as the Thematic Mapper (TM). Landsats 4 and 5 carried both the TM and an

MSS on an improved platform that maintained a high degree of stability as a means of improving image geometry, with additional sensor systems using improved technologies. The TM continued as the sensor system for Landsats 4 and 5 (1982–2013). The TM provided an impressive record of reliability, high-quality imagery, and continuity.

The TM was based on some of the same principles as the MSS but employed improved technologies to provide finer spatial resolution, improved geometric fidelity, greater radiometric detail, and more detailed spectral information in more precisely defined spectral regions. Despite the historical relationship between the MSS and the TM, the two sensors are distinct. Whereas the MSS used four broadly defined spectral regions, the TM recorded seven spectral bands (Table 7.2). Unlike the MSS, TM spectral channels were specifically tailored to record radiation of interest to specific scientific investigations rather than the arbitrary definitions used for the MSS. TM sensors used a spatial resolution of 30 m (about 0.09 ha, or 0.22 acre), which provided much finer spatial detail than the MSS. (TM band 6 used a coarser spatial resolution of about 120 m.) Digital values were quantized at 8 bits (256 brightness levels), which provided a much larger range of brightness values relative to the MSS. These modifications greatly improved the practical value of Landsat imagery, increasing the variety of applications and leading to a growing reliance on Landsat imagery for broad-scale monitoring and assessing landscapes.

The Landsat 7 Enhanced Thematic Mapper Plus (ETM+) was designed to extend the capabilities of previous TMs by adding modest improvements to the TM design (Figure 7.2). In the visible, near-infrared, and mid-infrared channels, its spectral channels duplicate those of TM (Table 7.2). The thermal channel has 60-m resolution, improved from the 120-m resolution of earlier TMs. The ETM+ also has a 15-m panchromatic channel. The system is characterized by improvements in accuracy of calibration, data transmis-

FIGURE 7.1 Multispectral Scanner System (MSS). The insight and vision of Virginia Norwood led to the design of the MSS, which, although intended as an experimental instrument, became the primary sensor for Landsat 1 when the primary instrument failed to perform as intended. The MSS's reliability and high-quality imagery formed the model for later Landsat instruments. From NASA.

TABLE 7.2 Summary of Landsat TM Sensor Characteristics

Band	Resolution	Spectral definition	Some applications
1	30 m	Blue-green, 0.45–0.52 μm	Penetration of clear water, bathymetry, mapping of coastal waters, chlorophyll absorption, distinction between coniferous and deciduous vegetation
2	30 m	Green, 0.52–0.60 μm	Reflection from green vegetation, plant vigor, turbid water
3	30 m	Red, 0.63–0.69 μm	Chlorophyll absorption, important for plant-type delineation
4	30 m	Near infrared, 0.76–0.90 μm	Indicator of plant cell structure, biomass, absorption by water facilitates shoreline delineation
5	30 m	Mid infrared, 1.55–1.75 μm	Vegetation moisture content, soil moisture mapping, differentiating snow from clouds, penetration of thin clouds
6	120 m	Far infrared, 10.4–12.5 μm	Vegetation heat stress analysis, evapotranspiration, thermal mapping
7	30 m	Mid infrared, 2.08–2.35 μm	Discrimination of rock types, alteration zones for hydrothermal mapping, hydroxylation absorption

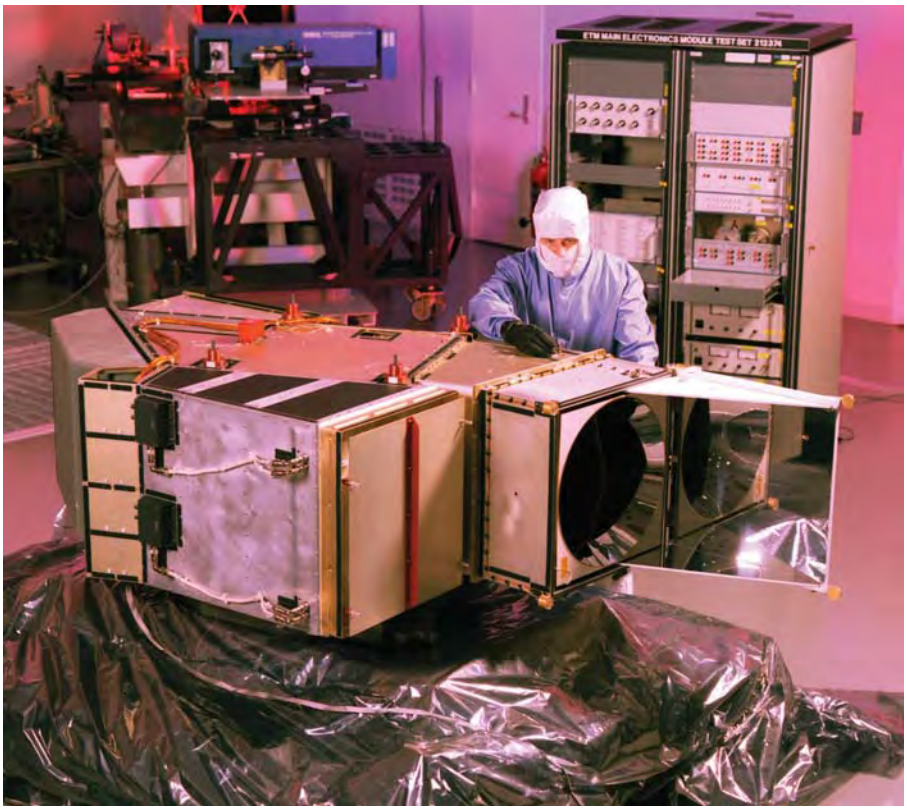


FIGURE 7.2 Pre-launch inspection, Enhanced Thematic Mapper Plus (ETM+) sensor for the Landsat 7 satellite. The ETM+ has eight bands in visible and infrared wavelengths. From NASA.

sion, and other characteristics. ETM+ extends the continuity of earlier Landsat data by maintaining consistent spectral definitions, resolutions, and scene characteristics.

In 2013, the launch of Landsat 8 introduced the *Operational Land Imager (OLI)* to replace the TM and ETM instruments. This action replaced the electro-optical technologies used for earlier Landsat TM and ETM systems with charge-coupled device (CCD) sensors (see Chapter 4) as the basis for the OLI. In comparison to earlier systems, Landsat 8 OLI (and the upcoming Landsat 9) have nine visible, near-infrared, and mid-infrared channels. These include an additional blue band for coastal aerosols (0.43–0.45 μm) and a cirrus band (1.36–1.38 μm). The wavelength ranges for each band are narrower than those in earlier systems and have higher radiometric resolution (12-bit) and a better signal-to-noise ratio. Landsat 8 also has two thermal infrared bands on the thermal infrared sensor (TIRS), which are described in Chapter 10. In addition, a quality assurance band provides information that supports, on a pixel-by-pixel basis, assessment of water, snow cover, or cloud conditions that may influence the accuracy and usability of pixels in other channels. As with Landsat 7, the OLI has a panchromatic channel that provides 15-m detail for channels within the visible spectrum.

Throughout the decades-long life of the Landsat, NASA engineers have devoted careful attention to engineering detail that has enabled several generations of Landsat imagery to provide imagery of consistent quality. Such instruments have permitted the collection of imagery from one system to another. In this respect, Landsat systems, unlike most other satellite systems, have collected a large, comprehensive image archive of consistent quality.

7.4 LANDSAT-LIKE SYSTEMS

The Landsat model has formed a template for the land remote sensing systems proposed and in use by many other nations. Lauer, Morain, and Salomonson (1997) listed 33 separate Landsat or Landsat-like systems launched or planned for launch between 1972 and 2007. This number attests to the value of observational systems designed to acquire regional overviews at moderate to coarse levels of spatial detail. Landsat has formed the model for these systems with respect to essential imaging capabilities, technological design, data management, and overall missions. For example, the first Landsat systems established the value of digital data for the general community of remote sensing practitioners, set expectations that imagery would be easily accessible, and established a model for organizing, archiving, and cataloging imagery. Despite differences, related systems have followed Landsat's lead in many respects.

SPOT

The French SPOT satellite system is an advanced Earth observation satellite system based on the Landsat model but with distinct technological advantages. SPOT, an acronym for *Satellite Pour l'Observation de la Terre* (Satellite for Observation of the Earth), was first lauched in 1986. The SPOT system was designed to provide data for land-use studies, assessment of renewable resources, exploration of geologic resources, and cartographic work at scales of 1:50,000 to 1:100,000. Design requirements included provision for complete world coverage, rapid dissemination of data, stereo capability, and sensitivity in spectral regions responsive to reflectance from vegetation.

SPOT satellites (Table 7.3) are placed into circular sun-synchronous orbits (SSOs), completing about 14 revolutions per day. Initially, for SPOT 1–3, the primary SPOT sensors were the two High Resolution Visible (HRV) sensors that could be operated independently, or in coordination, to acquire either panchromatic imagery (viewing a single wide band across the visible region) or multispectral imagery (viewing at coarser detail, but in several spectral channels, in the red, green, and near-infrared regions). Because the HRV sensors could be aimed laterally, they could acquire oblique views, in effect improving the system's revisit capability. As the SPOT sensors evolved, they significantly increased their spatial resolution. SPOT 4 improved upon the HRV design with the high-resolution visible and infrared (HRVIR) instrument, which included the addition of a mid-infrared band (1.58–1.75 μm) at 20-m resolution and red wavelengths at 10-m resolution. SPOT 5 enables 2.5-m or 5-m panchromatic imagery, with stereoimagery at 5 m, intended for large-scale topographic maps and data (Figure 7.3). SPOT 6 and 7 further increased the spatial resolution to provide 1.5-m panchromatic and 6-m multispectral imagery. These instruments provide high-resolution imagery with broad spatial coverage and therefore play an important role in high-resolution environmental applications.

Sentinel-2

The European Space Agency (ESA) has developed a comprehensive satellite observation program consisting of several remote sensing systems. (Sentinel-1 provides two satellites with active microwave; this type of system and data are discussed in Chapter 8.) Here we discuss Sentinel-2, modeled to provide a measure of compatibility with the Landsat and SPOT systems.

Sentinel-2A and -2B (launched in June 2015 and March 2017, respectively) are designed in part to provide systematic multispectral imagery to support land management, agriculture, forestry, disaster response, and security programs. Sentinel-2 satellites each provide 12-bit, high-resolution multispectral imagery, including 13 spectral bands covering a 290-km swath at 10-m, 20-m, and 60-m resolution (Table 7.4), acquired at

TABLE 7.3 Summary of SPOT Systems

Satellite	Launched	End of service[a]	Sensors
SPOT 1	February 22, 1986	December 31, 1990	HRV for SPOTs 1, 2, 3[a]
SPOT 2	January 22, 1990	July 2009	
SPOT 3	September 26, 1993	November 14, 1997	
SPOT 4	March 24, 1998	July 2013	HRVIR[b]
SPOT 5	May 4, 2002	March 31, 2015	HRG[c]
SPOT 6	September 9, 2012		NAOMI[d]
SPOT 7	June 30, 2014		

[a]High-resolution visible, with 10-m panchromatic and 20-m multispectral visible bands in the green, red, and NIR.
[b]High-resolution visible and mid infrared (20 m). A 10-m band covers the red wavelengths.
[c]High-resolution geometrical. Panchromatic imagery at 2.5 or 5 m. Stereo imagery at 5-m resolution. Spectral coverage of SPOT-4.
[d]New Astrosat Optical Modular Instrument, 1.5-m panchromatic and 6-m multispectral imagery.

FIGURE 7.3 SPOT 5 satellite image, Agger Tange, Denmark, 2002. From Airbus Defense and Space. Used by permission.

TABLE 7.4 Common Wavelength Range for Sentinel-2 Multispectral Instrument (MSI) Bands

Band	Wavelength range (μm)	Spatial resolution	Notes
1	0.4322–0.4527	60 m	Aerosol retrieval
2	0.4594–0.5251	10 m	
3	0.5418–0.5770	10 m	
4	0.6494–0.6801	10 m	
5	0.6966–0.7116	20 m	
6	0.7330–0.7466	20 m	
7	0.7728–0.7897	20 m	
8	0.7799–0.8858	10 m	
8a	0.8542–0.8750	20 m	
9	0.9351–0.9537	60 m	Water vapor correction
10	1.3619–1.389	60 m	Cirrus detection
11	1.5682–1.6574	20 m	
12	2.1149–2.2782	20 m	

Note: There are slight differences in the central wavelengths and bandwidths for Sentinel-2A and -2B. Most are very small (<1 nm), with the largest difference for Band 12 (17 nm central wavelength and 10 nm difference in bandwidth). Only the wavelength range in common is shown here. Please see *https://sentinel.esa.int/web/sentinel/user-guides/sentinel-2-msi/resolutions/radiometric* for exact band specifications.

10:30 A.M. local solar time. The two satellites have identical instrumentation, positioned in identical orbits but separated by 180° to provide frequent repeat coverage. At the equator, the Sentinel system provides a five-day coverage cycle when imagery from both satellites is used (shorter at higher latitudes). The four 10-m resolution bands allow for synergies with SPOT 4 and 5 data. The 20-m bands allow for analysis similar to what has been described for Landsat data but at higher spatial resolution. These synergies have enabled time-series applications in which the data from Landsat and Sentinel-2 are combined to improve the temporal coverage, including the development of a Harmonized Landsat and Sentinel-2 reflectance product (Claverie et al., 2018). The synergies of Sentinel-2 with SPOT and Landsat data reflect the scientific community's understanding of the importance of data continuity for the analysis of longer-term environmental processes and change.

Indian Remote Sensing

After operating two coarse-resolution remote sensing satellites in the 1970s and 1980s, India began to develop multispectral remote sensing programs in the style of the Landsat system. During the early 1990s, two India Remote Sensing (IRS) satellites were in service. IRS-1A, launched in 1988, and IRS-1B, launched in 1991, carried the LISS-I and LISS-II pushbroom sensors. These instruments collect data in four bands: blue (0.45–0.52 μm), green (0.52–0.59 μm), red (0.62–0.68 μm), and near infrared (0.77–0.86 μm), creating images of 2,400 lines in each band. LISS-I provides resolution of 72.5 m in a 148-km swath, and LISSII has 36.25-m resolution. Two LISS-II cameras acquire data from 74-km-wide swaths positioned within the field of view of LISS-I, so that four LISS-II images cover the area imaged by LISS-I, with an overlap of 1.5 km in the cross-track direction and about 12.76 km in the along-track direction. Repeat coverage is 22 days at the equator, with more frequent revisit capabilities at higher latitudes. In the United States, IRS imagery can be purchased from a variety of image services that can be located using the internet.

The LISS-III instrument is designed for the IRS-1C and IRS-1D missions, launched in 1995 and 1997, respectively. They acquire data in four bands: green (0.52–0.59 μm), red (0.62–0.68 μm), near infrared (0.77–0.86 μm), and shortwave infrared (1.55–1.70 μm). LISS-III provides 23-m resolution for all bands except for the shortwave infrared, which has a 70-m resolution. Swath width is 142 km for bands 2, 3, and 4, and 148 km in band 5. The satellite provides a capability for 24-day repeat coverage at the equator.

In 2003, India Remote Sensing launched *Resourcesat 1,* which includes three instruments of interest. The first is LISS-3, as described above. The second is LISS-4, which acquires a 23.9-km image swath in multispectral mode and a 70.3-km swath in panchromatic mode. LISS-4 has a spatial resolution of 5.5 m, and three spectral channels matching to channels 1, 2, and 3 of the LISS-3 instrument. The field of view can be aimed along the across-track dimension to acquire stereoscopic coverage and provide a five-day revisit capability. The third instrument is the Advanced Wide Field Sensor (AWiFS) instrument, which provides imagery of a 760-km swath at 56-m resolution, using the same four spectral channels as LISS-3.

IRS's *Cartosat-1* (May 2005) uses two panchromatic cameras, aimed fore and aft, to acquire stereoscopic imagery within a 30-km swath in the along-track dimension in a single panchromatic band (0.50–0.85 μm). Its objective is to acquire imagery to support detailed mapping and other cartographic applications at the cadastral level, urban and

rural infrastructure development and management, as well as applications in land information systems (LIS) and geographical information systems (GIS).

China-Brazil Earth Resources Satellite Program

Development of the China-Brazil Earth Resources Satellite Program (CBERS) dates from 1988, when China and Brazil agreed to pursue a joint collaboration in science and technology. Initially, the satellite program was administered by the Chinese Academy of Space Technology and Brazil's Instituto de Pesquisas Espaciais and focused on development and deployment of two satellites, CBERS-1 and CBERS-2. Later agreements included three additional satellites, CBERS-3, -4, and -4B. Combined, this suite of sensors captures panchromatic, visible, infrared, and thermal infrared at increasing spatial resolutions as the sensors evolved, similar to Landsat. However, these data have not been widely used outside of Brazil and China. This may change moving forward, as researchers are now increasing the use of multiple data sources to increase temporal coverage and improve data continuity to monitor changing Earth environments.

Advanced Spaceborne Thermal Emission and Reflection Radiometer

The Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER) is an imaging system carried on the TERRA satellite (used also for the Moderate Resolution Imaging Spectroradiometer [MODIS] system, as described in Section 7.5), launched in December 1999 in an SSO. The ASTER system was designed and constructed as a cooperative effort among NASA, Japan's Ministry of Economy, Trade and Industry (METI), and Japan Space Systems. ASTER does not fit well into the typology for classifying remote sensing satellite systems, as it covers a small area relative to other systems in the group. It is listed here with the Landsat-class satellites not because it offers broad coverage, but because it offers similar imaging capabilities.

ASTER's three sensor systems provide unique capabilities, based on 14 different spectral bands (see **Table 7.5** later in the chapter): (1) the 8-bit VNIR subsystem (two visible channels and a repeating NIR, one of which is nadir [3N] and the other backwards [3B], to capture a stereo pair in successive images); (2) the 8-bit SWIR subsystem (six channels at 30-m detail), and (3) a five-channel 12-bit thermal infrared (TIR) subsystem to capture a thermal perspective at 90-m detail. The ASTER SWIR channels are no longer active, although archived data are available. The three ASTER instruments (VNIR, SWIR, and TIR) can view the Earth in the cross-track orientation. Given its high resolution and its ability to change viewing angles, ASTER can capture stereoscopic images and generate terrain models.

Unlike the other satellite systems discussed here, ASTER will not collect imagery continuously; rather, it collects data as specifically requested by customers and scientists, for distribution shortly after the overpass. Current and archived ASTER data are available to the public without charge. Further information about the ASTER system can be found at:

- *https://asterweb.jpl.nasa.gov/characteristics.asp*
- *https://lpdaac.usgs.gov/documents/262/ASTER_User_Handbook_v2.pdf*

The data can be ordered from GloVIS (discussed in Section 7.9) or other data distribution sites.

7.5 BROAD-SCALE/COARSE RESOLUTION SATELLITES

Broad-scale/coarse resolution satellites have often been defined as systems that collect imagery at spatial resolutions between 250 m and 5 km, permitting broad-scale observation of the Earth at frequent revisit intervals. Such images, collected over a period of several weeks, can be used to generate cloud-free composites that represent large areas of the Earth without the cloud cover that would be present in any single scene. These images have opened a new perspective for remote sensing by allowing scientists to examine topics that require examination at continental or global scale that previously were outside the scope of direct observation.

Advanced Very High Resolution Radiometer

The Advanced Very High Resolution Radiometer (AVHRR) is a scanning system that has been carried on NOAA's polar orbiting environmental satellites since 1978 and on the European Organisation for the Exploitation of Meteorological Satellites Polar System since 2006. AVHRR is one of the oldest and most widely used of the broad-scale EOS systems. It provides 14 passes each day, viewing a 2,399-km swath. AVHRR collects data in four to six channels (depending on the satellite) covering the red, NIR, mid-IR, and thermal IR portions of the electromagnetic spectrum. The data support meteorology and climatology by providing synoptic information on snow/ice cover, water bodies, vegetation cover, and phenology.

At nadir, the AVHRR system provides an IFOV of about 1.1 km. Data presented at that level of detail is referred to as *local area coverage* (LAC). In addition, a *global areal coverage* (GAC) data set is generated by onboard averaging of the full-resolution data. GAC data are formed by the selection of every third line of data in the full-resolution data set. For each of these lines, four out of every five pixels are used to compute an average value that forms a single pixel in the GAC data set. This generalized GAC coverage provides pixels of about 4-km × 4-km resolution at nadir. A third AVHRR image product, high-resolution picture transmission (HRPT) data, is created by direct transmission of full-resolution data to a ground receiving station as the scanner collects the data. As for LAC data, resolution is 1.1 km. (LAC data are stored onboard for later transmission when the satellite is within line of sight of a ground receiving station; HRPT data are directly transmitted to ground receiving stations.)

Data from several passes are collected to create georeferenced composites that show large regions of the Earth without cloud cover. Scientists interested in broad-scale environmental issues can use such data to examine patterns of vegetation, climate, and temperature.

Further information describing AVHRR is available at:

- *www.usgs.gov/centers/eros/science/usgs-eros-archive-advanced-very-high-resolution-radiometer-avhrr*
- *www.esa.int/Applications/Observing_the_Earth/Meteorological_missions/MetOp/About_AVHRR_3*

Moderate Resolution Imaging Spectroradiometer

The Moderate Resolution Imaging Spectroradiometer (MODIS) is an imaging system carried by two satellite systems (Terra and Aqua) within NASA's Earth Observation System (EOS). MODIS is designed for long-term observation of the Earth's surface, atmosphere, and oceans. Terra ("Earth") follows an orbit timed for a morning equatorial (descending) pass, while Aqua ("Water") passes south to north (an ascending pass), crossing the equator in the afternoon. Terra was launched in December 1999, Aqua in May 2002. Both systems were still in operation as of spring 2022, far beyond their design life of 6 years, providing a valuable long-term data record.

MODIS established a new imaging capability, characterized by daily revisits (A.M. and P.M.) with a calibrated, science-quality, coarse-resolution satellite observation. MODIS records 36 image bands in the visible and infrared regions designed to observe, over time, key ecological, atmospheric, land cover, and atmospheric variables. These bands provide imagery at 250-m detail at nadir for two bands, 500-m resolution for five bands, and the remaining 29 bands at 1-km detail, covering the visible, NIR, mid-IR, and thermal wavelengths. Both satellites follow SSOs at 705 km, with a swath width of 2330 km, observing the Earth every 1 to 2 days.

MODIS imagery has enabled the compilation of global data sets at frequent intervals. Depending on the product of interest, NASA provides MODIS data as daily scenes or as derived products in multi-day composites (4-, 8-, and 16-day) at different spatial resolutions (Figure 7.4). Composites are designed to eliminate most cloud cover and many other errors that characterize individual data. Image composites are formed by selecting the highest-quality pixels within the interval, then preparing an image (the "composite"). Over time, many products developed from MODIS data have been widely used by scientists to study global change. They include products related to the atmosphere (aerosols, clouds, etc.), the land (land cover, surface temperature and emissivity, vegetation indices, leaf area index [LAI], evapotranspiration, etc.), the cryosphere (snow cover and sea ice), and the ocean (sea surface temperature, particulate carbon, and information related

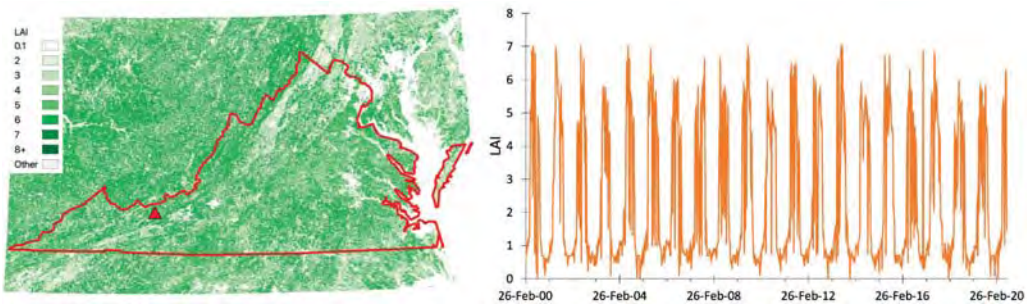


FIGURE 7.4 Left: Moderate Resolution Imaging Spectroradiometer (MODIS) Terra Leaf Area Index (LAI) 8-day composite around August 4, 2020, for an area on the east coast of the United States that includes Virginia (outlined in red). The red triangle in the western portion of Virginia indicates the location of a forested area in Bland County, for which the entire 8-day LAI time series was extracted. Right: MODIS 8-day LAI time series for the same forested area in Bland County. Note the seasonal trend of LAI, with peak growth in summer months and minimum LAI in winter, which is typical of the northern hemisphere.

to photosynthesis). More information about MODIS products can be found at *https://modis.gsfc.nasa.gov/data/dataprod*.

Visible Infrared Imaging Radiometer Suite

Building on the success of the MODIS systems, the VIIRS (*Visible Infrared Imaging Radiometer Suite*) system was launched in 2011 and complements MODIS's long-term data record. We now have 10 years of overlap between these two systems. VIIRS is a 22-band NASA-sponsored scanning radiometer operating in the visible and infrared regions (from 0.4 to 12.5 μm), tailored to collect calibrated, broad-scale data supporting oceanic, ecological, atmospheric, and oceanographic studies. VIIRS is one of five instruments onboard the Suomi National Polar-Orbiting Partnership (Suomi NPP) weather satellite (**Figure 7.5**). This satellite is named for the Finnish-American scientist Verner Edward Suomi, an early leader of satellite meteorology. Like MODIS, the VIIRS instrument collects surface observations that are comparable to conventional Earth observation data.

The principal data characteristics of VIIRS include 16 moderate-resolution (750 m at nadir) bands of which eight are in the visible-NIR wavelengths, three are SWIR channels, two are mid-IR channels, and three are thermal IR channels. There are also five imaging channels (325-m resolution at nadir), with two visible NIR wavelengths, and one each of SWIR, mid-infrared, and thermal wavelengths. In addition, there is a panchromatic band with a spatial resolution of 750 m throughout the entire scan. The swath width of the system is 3060 km.

As with MODIS data, numerous VIIRS data products can be freely downloaded at different spatial and temporal scales. These include the reflectance and emittance data, clouds, sea surface temperature, polar wind, ocean color, aerosols, fire, snow and ice, vegetation indices, thermal anomalies, land surface temperature, albedo, LAI/fPAR land surface phenology (fPAR stands for the fraction of photosynthetically active radiation absorbed by green vegetation), and nighttime lights. This has allowed for data continuity between MODIS and VIIRS and time-series products that combine the two systems.

7.6 FINE-RESOLUTION SATELLITE SYSTEMS

The third class of land observation satellites consists of systems designed to provide detailed coverage for small-footprint images. In the 1970s and 1980s, Landsat, SPOT, and other systems established the technical and commercial value of the land observation

FIGURE 7.5 The Visible Infrared Imaging Radiometer Suite (VIIRS), used for ocean, land, and atmosphere research. VIIRS will replace the Advanced Very High Resolution Radiometer (AVHRR) and MODIS, now approaching the ends of their service lives. From NASA/GSFC.

satellite concept, but they also revealed the limitations of systems intended to provide general-purpose data for a broad community of users who may require more specialized image resources.

The success of these early satellite systems revealed the demand for fine spatial resolution imagery required for tasks not satisfied by coarser resolutions, especially urban infrastructure analysis, transportation planning, precision agriculture, strategic reconnaissance, and construction support. By the late 1980s and early 1990s, technological, analytical, and commercial infrastructure was in place to support the development of smaller, special-purpose EOS focused on the requirements of these more narrowly defined markets. By 1994, the U.S. government relaxed its restrictions on commercial applications of high-resolution satellite imagery. These changes in policy, linked with advances in technology, opened opportunities for deployment of satellites designed to provide high-resolution image data with characteristics that are distinct from the design of the other, coarser-resolution systems described here.

Relative to the moderate-resolution class of satellite systems, fine-resolution systems offer highly specialized image products tailored for a specific set of applications. They provide imagery characterized by high spatial detail, high radiometric resolution with narrow fields of view, and small footprints targeted on specified areas. They acquire data in the optical and near-infrared regions of the spectrum, often with the capability to provide panchromatic imagery at even finer spatial detail. Some systems offer stereo capability and may offer orthocorrected imagery to fully account for the positional effects of terrain relief within the image area. Fine-resolution systems are organized as commercial enterprises, although it is quite common for many to have strategic relationships with government organizations. Image providers often collect image data at 10- or 11-bit resolution, although some products are distributed at 8-bit resolution for customers who prefer smaller file sizes.

Sometimes fine-resolution satellite data are prepared as *pan-sharpened* imagery—that is, fused or merged multispectral channels that combine multispectral data with the finer detail of the corresponding panchromatic channel. For visual interpretation, such products can provide the benefit of multispectral imagery at the finer spatial detail of the panchromatic channel. However, unlike in a true multispectral image, the components of the pan-sharpened image are, so to speak, baked together such that they cannot be independently manipulated to change the image presentation. Therefore, they have limited utility in the analytical context, although they provide value for their effectiveness in visual interpretations.

Stated broadly, image archives for fine-resolution imaging systems began in the late 1990s, so they will not have the comprehensive temporal or geographic scope of those moderate-resolution systems due to smaller footprints and acquisition strategies designed to respond to customer requests rather than to systematically build geographic coverage. The archives of the numerous commercial enterprises are not centralized, and users may find that some organizations have changed as they were reorganized or acquired by other enterprises.

Fine-resolution satellite imagery has created viable alternatives to aerial photography, particularly in remote regions where the logistics of aerial photography may generate high costs. Fine-resolution satellite imagery has found important application niches for mapping of urban utility infrastructure, floodplain mapping, engineering and construction analysis, topographic site mapping, change detection, transportation planning, and precision agriculture.

In the following capsule descriptions of fine-resolution systems, readers should refer to websites of companies that operate these systems for specifics concerning stereo capabilities, positional accuracy revisit times, and other characteristics that are too intricate for the concise descriptions presented here. There are too many high-resolution systems in orbit to comprehensively review in this book. Here we highlight selected systems operated by Maxar and ImageSat International, which include well-known satellite systems, several of which originated from stand-alone companies that have merged with the parent company over the years.

GeoEye-1 (Maxar)

GeoEye-1, launched in September 2008 and operated by Maxar, acquires panchromatic imagery at 0.41-m resolution and multispectral imagery at 1.65 m. Its footprint is 15.2 km at nadir.

- *Band 1:* 0.450–0.800 μm (panchromatic)
- *Band 2:* 0.450–0.510 μm (blue)
- *Band 3:* 0.510–0.580 μm (green)
- *Band 4:* 0.655–0.690 μm (red)
- *Band 5:* 0.780–0.920 μm (near infrared)

Details are available at *https://satimagingcorp.com/satellite-sensors/geoeye-1*.

IKONOS (DigitalGlobe)

The IKONOS satellite system (named from the Greek word for "image"), now decommissioned, was launched in September 1999 and operated by DigitalGlobe (now Maxar) as the first commercial satellite able to collect submeter imagery (0.80-m resolution in panchromatic mode [0.45–0.90 μm] and 3.2-m resolution in multispectral mode):

- *Band 1:* 0.45–0.52 μm (blue)
- *Band 2:* 0.52–0.60 μm (green)
- *Band 3:* 0.63–0.69 μm (red)
- *Band 4:* 0.76–0.90 μm (near infrared)

IKONOS imagery had an image swath of 11.3 km at nadir; imagery was acquired from an SSO, with a 10:30 A.M. equatorial crossing. The revisit interval varied with latitude; at 40°, repeat coverage could be acquired at about 3 days in the multispectral mode and at about 11 to 12 days in the panchromatic mode.

For details, see *https://satimagingcorp.com/satellite-sensors/ikonos*.

QuickBird (DigitalGlobe)

QuickBird, now decomissioned, launched in October 2001 by DigitalGlobe (now Maxar), was able to acquire panchromatic imagery at 0.60-m resolution and multispectral imagery at 2.4-m resolution, with a revisit capability of 3.5 days, depending on latitude. Its

footprint at nadir was 16.5 km. It collected both panchromatic and multispectral imagery as follows:

- *Band 1:* 0.45–0.52 μm (blue)
- *Band 2:* 0.52–0.60 μm (green)
- *Band 3:* 0.63–0.69 μm (red)
- *Band 4:* 0.76–0.89 μm (near infrared)
- *Band 5:* 0.45–0.90 μm (panchromatic)

For more information, see *https://satimagingcorp.com/satellite-sensors/quickbird*.

WorldView-1, -2, -3, and -4 (Maxar)

Maxar's WorldView satellites have provided high-resolution commercial satellite imagery since the launch of WorldView-1 in September 2007. The systems are characterized by increasing spatial resolution and number of spectral bands from WorldView-1, -2, and -3, as shown in Table 7.5. WorldView-3 also includes moderate-resolution CAVIS bands (Clouds, Aerosols, Vapors, Ice, Snow; 30 m) to map clouds, ice, and snow and to correct for aerosols and water vapor. Until it stopped operating in 2019, WorldView-4 was similar to WorldView-3 in the panchromatic and some VNIR bands.

EROS (ImageSat International)

ImageSat International operates Earth Resources Observation Satellites (EROS), deployed in sun-synchronous, near-polar orbits. They are focused primarily on intelligence, homeland security, and national development missions, but they are also employed in a wide range of civilian applications, including mapping, border control, infrastructure planning, agricultural monitoring, environmental monitoring, and disaster response. EROS A (launched in December 2000 and ended in 2016) collected panchromatic imagery at 1.9-m to 1.2-m resolution. EROS B (launched in April 2006 and still active) acquires

TABLE 7.5 Summary of WorldView Satellites

Sensor	Operation	Description
WorldView-1	September 18, 2007–present	0.5 m panchromatic (PAN) (400–900 nm)
WorldView-2	October 8, 2009–present	0.46 m PAN (450–800 nm), 1.85 m multispectral 8 bands in visible, NIR, and red-edge
WorldView-3	August 13, 2014–present	0.31 m PAN (450–800nm), 1.24 m VNIR, 3.7 m SWIR, 30 m CAVIS 8 VNIR bands (397–1039 nm) 8 SWIR bands (1184–2373 nm) 12 CAVIS bands (405–2245 nm)
WorldView-4	November 11, 2016–January 7, 2019	0.31 m PAN (450–800nm), 1.24 m VNIR 4 VNIR bands

Note: Additional details are available at *https://satimagingcorp.com/satellite-sensors*.

panchromatic imagery at 70-cm resolution. Their "next-generation" constellation (EROS NG Constellation) has three satellites at 30-cm resolution. More information about these systems is available at *www.imagesatintl.com/about*.

7.7 SMALLSATS

SmallSats are a category of Earth observation satellites defined by their small size, characterized typically by a mass less than 180 kg, a size sometimes compared to that of a large refrigerator. Variations are designated by the mass of each device:

- Minisatellite: 100–180 kg
- Microsatellite: 10–100 kg
- Nanosatellite: 1–10 kg
- Picosatellite: 0.01–1 kg
- Femtosatellite: 0.001–0.01 kg

Early versions of larger SmallSats were designed and launched as early as the 1970s, leading to well-known interplanetary satellites, such as Voyagers 1 and 2. For some missions, they are favored for their inexpensive launch and operational costs. Our interest here is in the use of such small systems for Earth observation and some of their early applications.

CubeSats, first introduced in Chapter 3, are a subclass of nanosatellites manufactured to specific sizes (Figure 7.6). This concept originated in the 1990s when universities developed a standardized template for educational and experimental satellite research. The standard CubeSat is based on a 1 unit (1 U), defined as a cube of 10 × 10 × 10 cm, expandable to form larger variations, such as 1.5, 2, 3, 6 U's or larger.

A *satellite constellation* is a system of artificial Earth-orbiting satellites operating to achieve a common mission, such as collection of imagery of a specified region of the

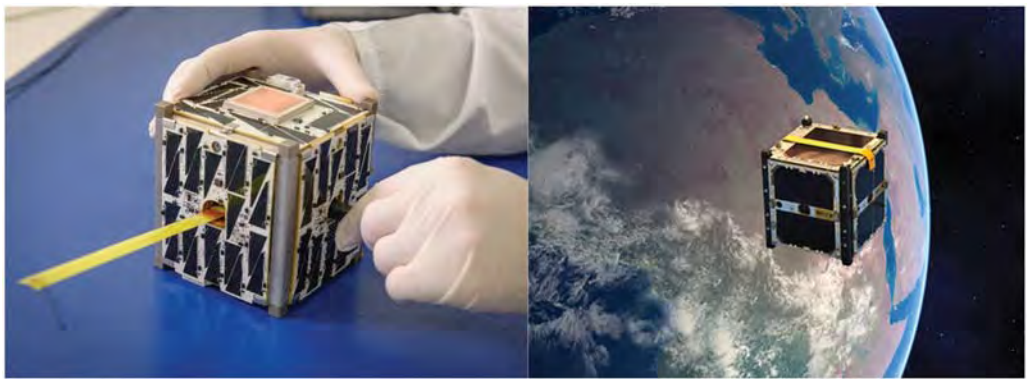


FIGURE 7.6 CubeSats. Left: PhoneSat 2.5, a CubeSat (roughly 4 × 4 × 4 in.) built at NASA's Ames Research Center with components from commercially available smartphones. From NASA Ames. Right: Artist's rendition of a CubeSat in orbit. From NASA, Jet Propulsion Laboratory.

Earth's surface, or to observe the same region from different perspectives. Currently, CubeSats are recognized for their role in developing satellite constellations.

Planet

Planet (or Planet Labs) operates a constellation of three systems that acquire global high-resolution data, with more than 150 satellites. Most of the satellites (over 130 CubeSats) are part of PlanetScope, which has been active since 2009 and provides data at 3.7-m (resampled to 3-m) resolution in the VNIR wavelengths. Another system, RapidEye, has five satellites that provide imagery at 5 m (resampled from 6.5 m) in the visible, red-edge, and NIR wavelengths. The third system is SkySat (also known as Terra Bella, or SkyBox), a SmallSat system that consists of 15 satellites collecting panchromatic and VNIR data. In this system, panchromatic and pansharpened VNIR data are available at 0.5 m (resampled from 0.8 m), with normal VNIR at 1-m spatial resolution. More information about these systems can be found at *www.planet.com/products/planet-imagery*.

7.8 LAND OBSERVATION SATELLITE ORBITS

Satellites are placed in orbits tailored to match the mission objectives of each satellite and the capabilities of the sensors they carry. Types of orbits were introduced in Chapter 3. The land observation satellite systems that form the principal focus for this chapter are positioned in SSOs, designed such that each orbital pass on the daylight side of the Earth is shifted westward each day to minimize variations in solar angle at the satellite's overpass. Figure 7.7 illustrates this concept by showing that, as observed from above the North Pole, the orbital plane of the satellite maintains a constant angle with respect to the solar beam. Ideally, all satellite images would be acquired under conditions of uniform illumination, so that brightnesses of features within each scene could reliably indicate conditions on the ground. In reality, brightnesses recorded by satellite images are not directly indicative of ground conditions because differences in latitude, time of day, and season lead to variations in the nature and intensity of light that illuminates each scene. See Chapters 5 and 6 for further discussion of this issue.

For simplicity, this section describes *normal* orbits, an idealization based on the assumption that the Earth's gravitational field is spherical, although satellites actually follow *perturbed* orbits, due in part to distortion of the Earth's gravitational field by the Earth's *oblate* shape (flattened at the poles and bulging at the equator) and in part to influences of lunar and solar gravity, atmospheric drag, and solar wind. A *normal orbit* forms an ellipse with the center of the Earth at one focus (Figure 7.8, left), characterized by an *apogee* (*A*; point farthest from the Earth), *perigee* (*P*; point closest to the Earth), *ascending node* (*AN*; point where the satellite crosses the equator moving south to north), and *descending node* (*DN*; point where the satellite crosses the equator passing north to south). For graphical simplicity, the *inclination* (*i*) is shown in Figure 7.8 (right) as the angle that a satellite track forms with respect to the equator at the descending node.

Thus, SSOs are designed to reduce variations in illumination by systematically moving the orbital track westward (*precessing*) 360° each year. Illumination observed under such conditions varies throughout the year but repeats on a yearly basis. Careful selection of orbital height, eccentricity, and inclination exploit the gravitational effects of the

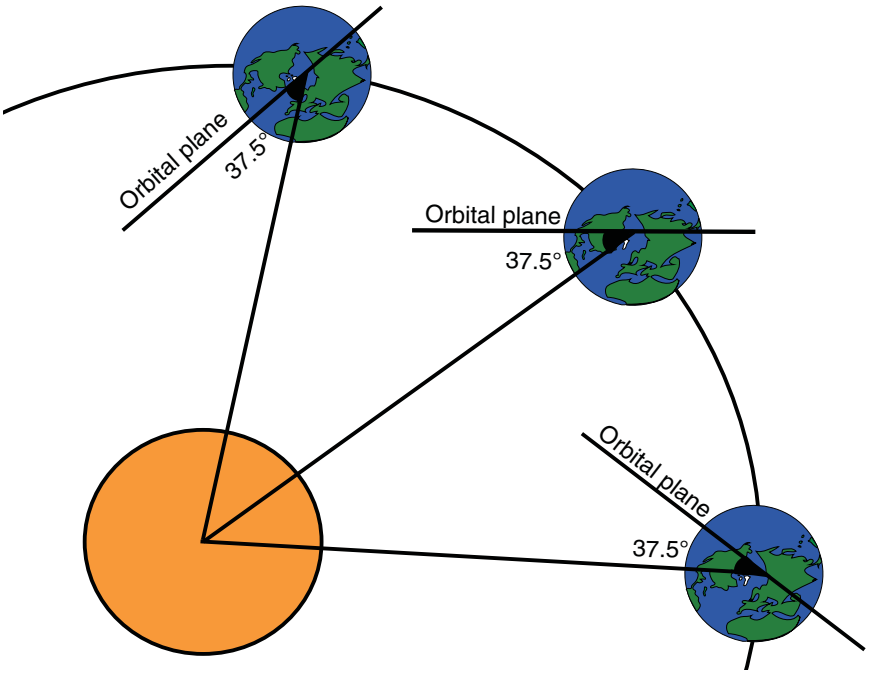


FIGURE 7.7 Schematic view of the Earth, observed as an orbiting satellite as viewed from above the North Pole. An SSO passes over a given point on Earth at the same mean solar time. This effect is achieved by selecting an orbital altitude and inclination such that the oblateness of the Earth causes a desired precession of the orbit. This orbit has many desirable features, including retention of the same angle with respect to the Sun. Adapted from U.S. Geological Survey (USGS). Image by Susmita Sen.

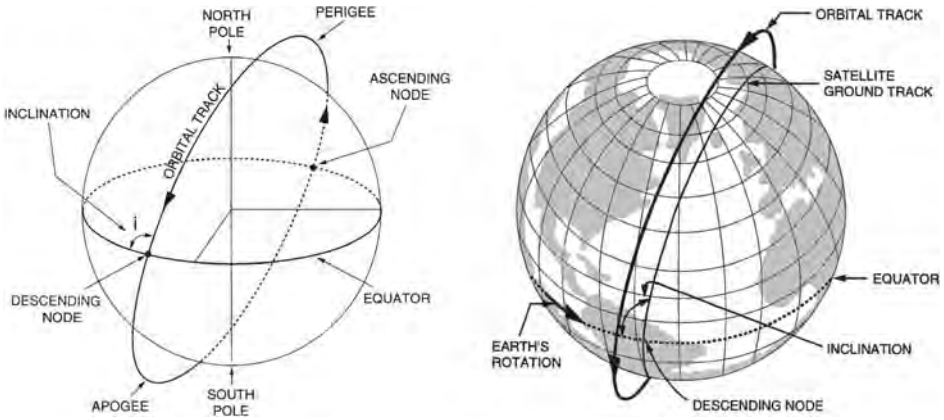


FIGURE 7.8 Satellite orbits. Left: Definitions. Right: Representation of an SSO, showing the orbital and satellite ground tracks.

Earth's equatorial bulge to cause the plane of the satellite's orbit to rotate with respect to the Earth, to match seasonal motion of the solar beam. That is, the nodes of the satellite's orbit will move eastward about 1° each day (Figure 7.9), so that over a year's time, the orbit will move through a complete 360° cycle. A satellite placed in an SSO will observe each part of the Earth within its view at the same local Sun time each day, thereby removing time of day as a source of variation in illumination. Although optimum local Sun time varies with the mission of the specific satellite, most Earth observation satellites are placed in orbits designed to acquire imagery between approximately 9:30 A.M. and 10:30 A.M. local Sun time—a time that provides a trade-off between preferred illumination and time of minimum cloud cover (as considered globally).

7.9 DATA ARCHIVES AND INDEXING SYSTEMS FOR LAND OBSERVATION SATELLITES

Because of the unprecedented amount of data generated by Earth observation satellite systems, images are organized and indexed using computerized databases. Records describing coverages of satellite data, known as *metadata,* are important to enable users to search these vast databases (Mather and Newman, 1995). In the context of remote sensing, metadata usually consist of text describing images: dates, spectral regions, quality ratings, cloud cover, geographic coverage, and so on, for each image or scene. Usually, users can examine such archives on the internet. One example archive is *Earth Explorer,* offered by the USGS as an index for USGS and Landsat imagery and cartographic data (*http://earthexplorer.usgs.gov*). In this system, users can specify geographic regions of

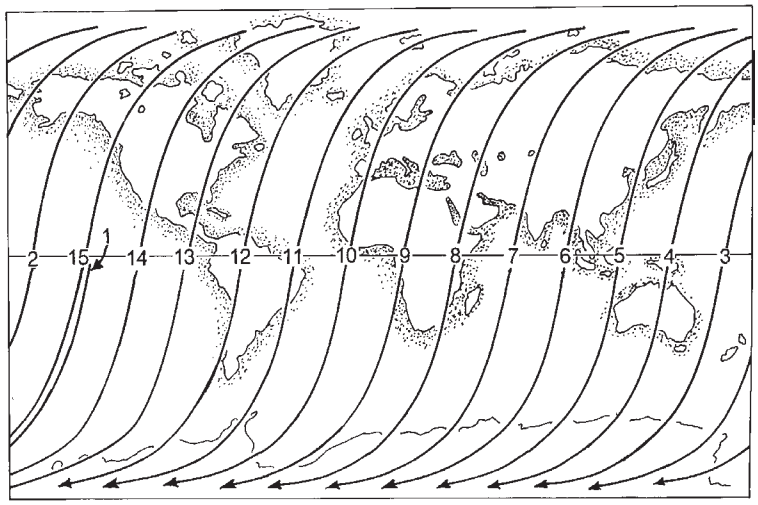


FIGURE 7.9 An example of a land observation satellite coverage cycle, in this case for Landsats 1, 2, and 3. Each numbered line designates a northeast-to-southeast pass of the satellite. In a single 24-hour interval, the satellite completes 14 orbits; the first pass on the next day (orbit 15) is immediately adjacent to pass 1 on the preceding day.

interest by using geographic coordinates, by outlining a region on a map, or by entering a place name. Subsequent screens allow users to select specific kinds of data, specify desirable dates of coverage, and indicate the minimum quality of coverage. The result is a computer listing that provides a tabulation of coverage meeting the constraints specified by the user.

Just as there are numerous and increasing land observation satellites, there are also different ways that each government agency or corporation indexes that data so that they can be archived, searched, and accessed by their locations. Landsat data are indexed according to how scenes are acquired along their orbital paths (described below), and there is some overlap at the edges of scenes. Other systems archive and release their data according to nonoverlapping tiling grids. For example, the European Space Agency Sentinel-2 system (described above) uses a tiling scheme of 100-km × 100-km tiles projected into UTM/WGS84. As the spatial resolution becomes coarser, the Earth may be tiled according to degrees, using different projections. A good example of this is MODIS, which has data products at 250 m, 500 m, 1 km, and 0.05 degrees. For MODIS data, there are three projections, depending on intended application (sinusoidal, Lambert azimuthal equal area, and geographic latitute/longitude for climate modeling). In general, there are too many satellites and different indexing systems to describe fully here, and it is more important that the user be aware of how the data are organized for their specific application and system of choice. For data archiving and access purposes, most systems have a tiling naming convention, such that users can access the data according to the tile name (or the path/row for Landsat) or according to a coordinate of interest. The particulars of the Landsat Worldwide Reference Systems are described below because there is now such a large archive of Landsat data stored using the system, and because the tiles are overlapping.

There are several well-known data archiving systems in which satellite data from different systems can be queried and ordered. One common data portal is the USGS Global Visualization Viewer (GloVis, *https://glovis.usgs.gov*), which allows users to constrain the date, cloud cover, scene quality, and so on (**Figure 7.10**). Two others are EarthExplorer (*https://earthexplorer.usgs.gov*) and the Copernicus Data and Information Access Service (DIAS) (*https://sentinel.esa.int/web/sentinel/sentinel-data-access*). There are also systems such as Google Earth Engine (*earthengine.google.com*), which archive data and allow image processing in the same interface, eliminating the need for massive data transfer and downloads for time series analysis or studies over large regions.

Worldwide Reference Systems

The worldwide reference systems (WRS and WRS-2) are a concise designation of nominal center points of Landsat scenes, used to index Landsat scenes by location. The reference system is based on a coordinate system in which there are 233 north–south paths corresponding to orbital tracks of the satellite and 119 rows representing latitudinal centerlines of Landsat scenes. A pair of path (E-W) and row (N-S) numbers uniquely identify a nominal scene center (**Figure 7.11**). Because of the drift of satellite orbits over time, actual scene centers may not match exactly to path-row locations, but the method provides a convenient and effective means of indexing locations of Landsat scenes. The initial WRS (WRS-1) was designed for MSS imagery; development of the TM led to a second version, WRS-2, tailored to index TM and OLI imagery (see also *https://landsat.gsfc.nasa.gov/about/the-worldwide-reference-system*).

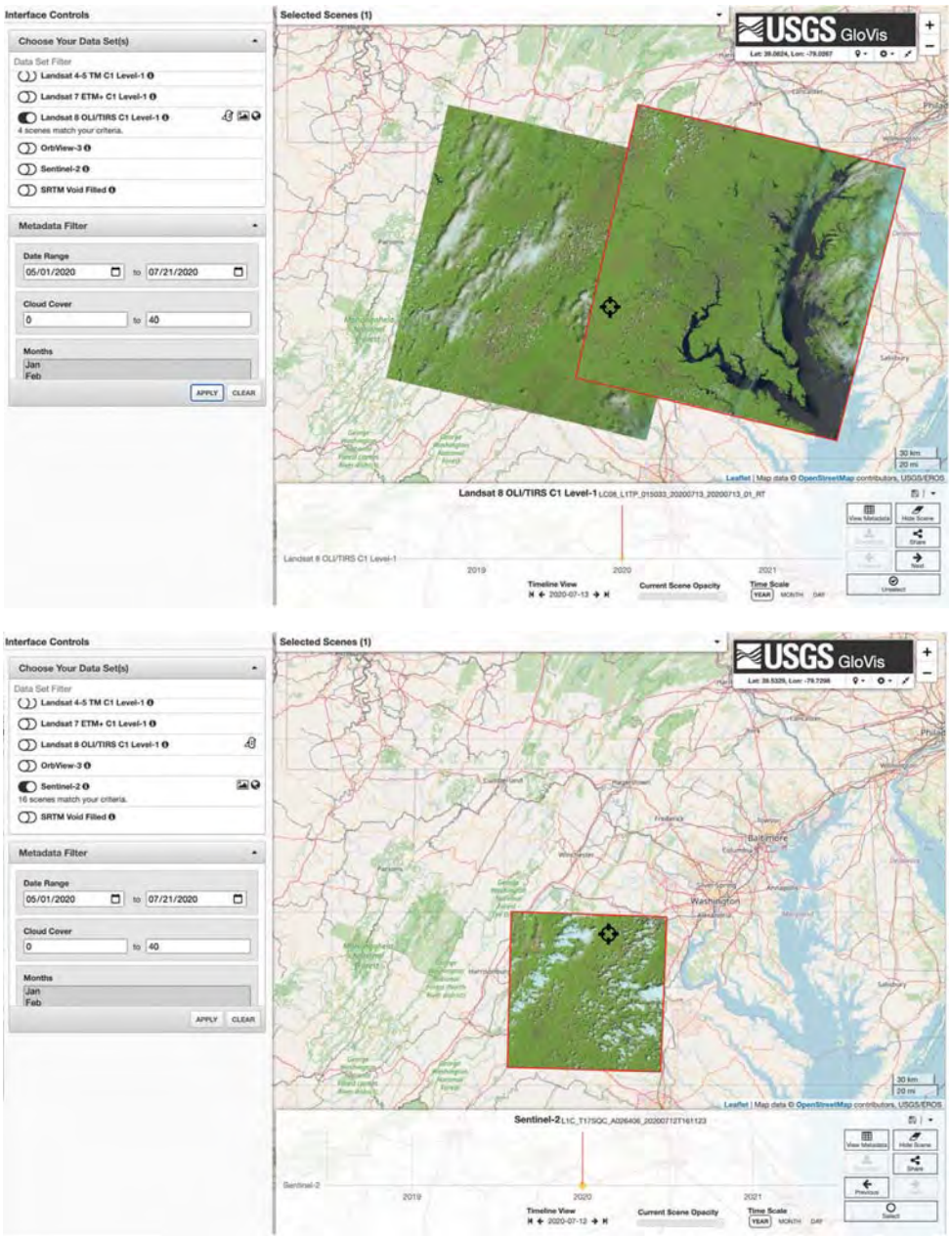


FIGURE 7.10 GloVis screenshot depicting Washington, D.C., and the surrounding area. Users can choose the data set (Landsat, Sentinel-2, Orbview-3, Aster, etc.) and set constraints on the date range, cloud cover, scene quality, and so on to select images. In this case, the date range is constrained from May 1, 2020, to July 21, 2020, with 40% maximum cloud cover. The upper image shows Landsat 8 scenes of the Washington, D.C., area (path 15, row 33) and the neighboring scene to the west (path 16, row 33). There are four total Landsat scenes in this area for these constraints (i.e., different dates). The lower image shows a Sentinel-2 tile for the same area. There are 16 possible Sentinel-2 tiles in this area for these constraints. From USGS.

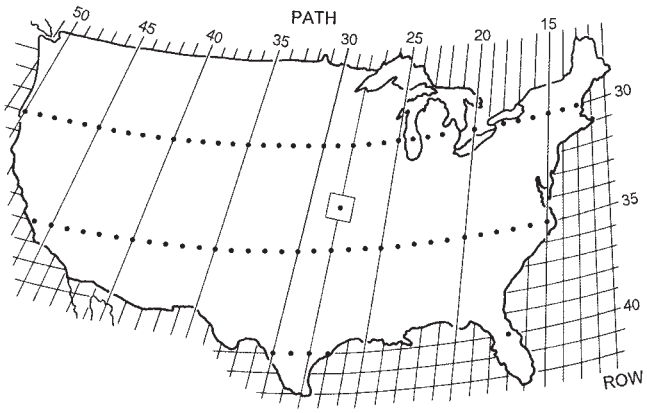


FIGURE 7.11 Path-row coordinates for the coterminous United States, which provide an indexing system for Landsat imagery. Here, the outline positioned at path 28, row 36 provides an example of the position and outline of a Landsat scene. Based on a NASA image.

In a single 24-hour interval, the satellite completes 14 orbits; the first pass on the next day (orbit 15) is immediately adjacent to pass 1 on the preceding day, which provides an indexing system for Landsat imagery (Figure 7.12).

7.10 INTERNATIONAL CHARTER

The *International Charter,* now in service for over 20 years, is a cooperative network of national governments and commercial satellite organizations that formally agree to share, without cost, image data with other nations to support disaster relief (such as floods, fire, and earthquakes) (Sivanpillai, Jones, and Lamb, 2017). This international network provides an ability to offer reporting covering a range of image dates and image scales. The International Charter allows for designation of authorized users qualified to use image data. Each agency member has committed resources to support the mission of the Charter and mitigate the effects of disasters on human life and property. The Interna-

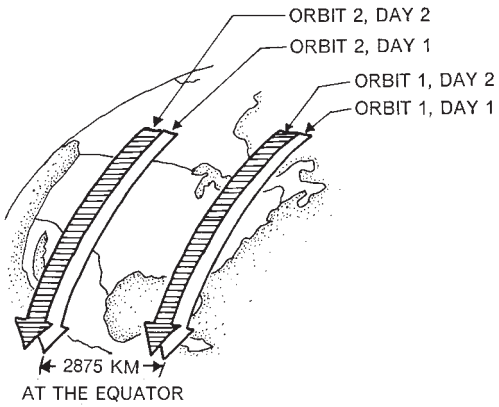


FIGURE 7.12 Progressive accumulation of Landsat coverage. On successive days, orbital tracks begin to fill in gaps left during westward displacement of orbits during the previous day. After 18 days, progressive accumulation of coverage fills all gaps left on Day 1. This system (WRS-2) was designed for Landsats 1–3, applied to Landsats 4, 5, and 7, and later to Landsat 8. Based on a NASA image.

tional Charter has provided data to support efforts to address the impacts of catastrophic events such as fires, earthquakes, floods, oil spills, and landslides.

In the United States, the USGS is the lead agency for participation in the International Charter, calling on governmental and commercial assets to support Charter needs. USGS staff have supported coordination, organization, and training of volunteer participants to be prepared when emergencies occur. The USGS is able to provide direct historical and current satellite data to support disaster management of agencies and international relief organizations.

As a participant in the International Charter, USGS provides the following forms of support:

- Submits requests for activation of the International Charter for international disasters on behalf of responsible emergency management authorities
- Provides access to imagery from U.S. government and commercial resources in response to International Charter requests
- Provides project manager (PM) services as needed
- Supports training and coordination of the global PM network
- Supports training and coordination of nonmember authorized users
- Provides U.S. representation on International Charter operational and governance bodies

For more information, see:

- *www.disasterscharter.org/web/guest/home;jsessionid=644D24FA6A58EC56F2F36A5079D98531.jvm1*
- *https://www.usgs.gov/emergency-operations-portal*

7.11 SUMMARY

Today, Earth observation satellites are widely used throughout the world, supporting environmental monitoring, agriculture, forestry, water quality, defense and national security, urban systems, and many other contributions. Since the launch of Landsat 1, multiple new systems have been launched by at least 34 sovereign states to help address Earth system questions of national and global concern. Belward and Skøien (2015) offers a more complete assessment of the launches of civilian Earth observation statellites. Public knowledge of and interest in remote sensing has increased. Digital data for satellite images have contributed greatly to the growth of image processing, pattern recognition, and image analysis (Chapters 11–15). Satellite observation systems have increased international cooperation and collaboration through joint construction and operation of ground-receiving stations and through collaboration in the training of scientists.

The history of the U.S. Landsat system, as well as histories of comparable systems of other nations, illustrate the continuing difficulties experienced in defining structures for financing the development, operation, and distribution costs required for satellite imaging systems. The large initial investments and high continuing operational costs (quite unlike those of aircraft systems) resemble those of a public utility. With Landsat data having been freely available to the public since 2008, the use of these data has increased

exponentially, significantly increasing the public value and allowing for new types of multitemporal analysis. In 2015, the U.S. National Geospatial Advisory Committee reported that "Landsat is widely considered to be a crucial national asset, comparable to the satellite-based GPS system and National Weather Service satellites" (*www.usgs.gov/news/landsat-seen-stunning-return-public-investment*). Other systems have adopted a similar philosophy (e.g., Sentinel), and data are freely explored and available to users. Because of the long-term records of some of the data (i.e., Landsat, MODIS, AVHRR, and VIIRS), data continuity is a major concern (and benefit), and long-term funding for these systems is an important challenge.

Another concern that has been attracting more and more public attention is the issue of personal privacy. As systems provide more or less routine availability of imagery with submeter resolution, governmental agencies and private corporations have direct access to data that could provide detailed information about specific individuals and their property. Although this imagery does not necessarily provide information that is not also available through aerial photography, the ease of access and the standardized format of satellite data open new avenues for use of such information. Combined with significant advancements in "big data" analytics, the real concern should focus not so much on the imagery itself as on the effects of combining such imagery with other data from marketing information, census information, and the like. The combination of these several forms of information, each in itself rather benign, could develop capabilities that many people would consider to be objectionable or even dangerous. Land observation systems have benefited from the coordination and collaboration of government and corporate partners. In recent years, programs have increased collaborations, seeking consistency and continuity.

Technological developments have taken Earth observation from space into new directions. The focus on miniaturization in the form of SmallSat constellations has enabled cheaper, more agile roles for smaller enterprises and tailored Earth observation questions. Smallsats provide great benefit to certain applications but as yet cannot substitute for the value of Landsat-like or broad-scale systems that provide continuity in coverage with carefully calibrated systems to answer many important questions about Earth system dynamics.

This chapter forms an important part of the foundation necessary to develop topics presented in subsequent chapters. The specific systems described here are significant in their own right, but they also provide the foundation for understanding other satellite systems that operate in the microwave (Chapter 8) and far-infrared (Chapter 10) regions of the spectrum. Finally, it can be noted that the discussion thus far has emphasized acquisition of satellite data. Little has been said about analysis of these data and their applications to specific fields of study. Both topics will be covered in subsequent chapters (Chapters 11–15 and 16–21, respectively).

REVIEW QUESTIONS

1. Several satellite systems have a long-term series of global data that are freely available to the public (e.g., Landsat, MODIS, VIIRS, AVHRR). What are some science questions that can now be investigated using these long-term data that would not be possible with only a few dates?

2. What are some advantages of using satellite imagery, relative to the use of aerial photography? Can you identify disadvantages?
3. What are some pros and cons of private ownership of high-resolution satellite systems?
4. In some instances, it may be necessary to form a mosaic of several satellite scenes by matching images together at the edges. List some of the problems you expect to encounter as you prepare such a mosaic.
5. Why are orbits of land observation satellites so low relative to those of communications satellites?
6. What are some advantages of SmallSat constellations over more traditional Landsat-like or broad-scale systems? What are some disadvantages? Can you list some questions that might be better addressed with SmallSat systems?
7. Explain why a satellite image and an aerial mosaic of the same ground area are not equally useful, even though image scale might be the same.

REFERENCES

Baker, J. C., K. M. O'Connell, and R. Williamson (eds.). 2001. *Commercial Observation Satellites at the Leading Edge of Global Transparency*. Santa Monica, CA: RAND Corporation. *www.rand.org/pubs/monograph_reports/MR1229.html*.

Belward, A. S., and J. O. Skøien. 2015. Who Launched What, When, and Why; Trends in Global Land-Cover Observation Capacity from Civilian Earth Observation Satellites. *ISPRS Journal of Photogrammetry and Remote Sensing*, Vol. 103, pp. 115–128.

Campbell, J. B., and V. V. Salomonson. 2010. Remote Sensing—A Look to the Future. Chapter 25 in *Manual of Geospatial Science and Technology* (2nd ed.) (J. D. Bossler, ed.). Boca Raton, FL: CRC Press, pp. 487–509.

Claverie, M., J. Ju, J. G. Masek, J. L., Dungan, E. F. Vermote, J. C. Roger, S. V. Skakun, and C. Justice. 2018. The Harmonized Landsat and Sentinel-2 Surface Reflectance Data Set. *Remote Sensing of Environment*, Vol. 219, pp. 145–161.

Cowan, G. 2016. It's a New Space Age for Satellite Builders. *Wall Street Journal*. October 2, *https://twitter.com/cears_vt/status/782934092904075264*.

Curran, P. J. 1985. *Principles of Remote Sensing*. New York: Longman, 282 pp.

Day, D. A., J. M. Logsdon, and B. Latell (eds.). 1998. *Eye in the Sky: The Story of the Corona Spy Satellites*. Washington, DC: Smithsonian Institution Press, 303 pp.

David, J. E. 2015. *Spies and Shuttles*. Smithsonian National Air and Space Museum, Washington, DC, 355 pp.

Drusch, M., U. Del Bello, S. Carlier, O. Colin, V. Fernandez, F. Gascon, . . . and P. Bargellini. 2012. Sentinel-2: Optical High-Resolution Mission for GMES Operational Services. *Remote Sensing of Environment*, Vol. 120, pp. 25–36.

Hand, E. 2016. Startup Liftoff. *Science*, Vol. 348, pp. 172–177.

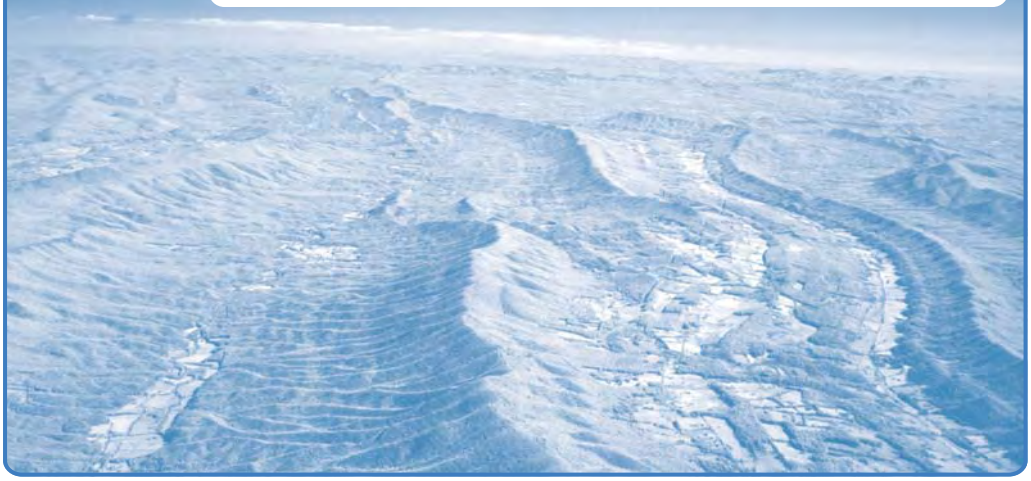
Justice, C. O., and C. J. Tucker. 2009. Coarse Resolution Optical Sensors. Chapter 10 in *The Sage Handbook of Remote Sensing* (T. A. Warner, M. D. Nellis, and G. M. Foody, eds.). Washington, DC: SAGE, pp. 139–150.

Lauer, D. T., S. A. Morain, and V. V. Salomonson. 1997. The Landsat Program: Its Origins, Evolution, and Impacts. *Photogrammetric Engineering and Remote Sensing*, Vol. 63, pp. 831–838.

Landsat Legacy Project Team. 2017. *Landsat's Enduring Legacy: Pioneering Global Land Obser-*

vations from Space. Bethesda, MD: American Society for Photogrammetry and Remote Sensing.
Mack, P. 1990. *Viewing the Earth: The Social Constitution of the Landsat Satellite System*. Cambridge, MA: MIT Press, 270 pp.
Markham, B. L. 2004. Landsat Sensor Performance: History and Current Status. *IEEE Transactions on Geoscience and Remote Sensing,* Vol. 42, pp. 2691–2694.
Mather, P. M., and I. A. Newman. 1995. U.K. Global Change Federal Metadata Network. Chapter 9 in *TERRA-2: Understanding the Terrestrial Environment* (P. M. Mather, ed.). New York: Wiley, pp. 103–111.
Morain, S. A., and A. M. Budge. 1995. *Earth Observing Platforms and Sensors CD-ROM*. Bethesda, MD: American Society for Photogrammetry and Remote Sensing.
National Geospatial Advisory Committee. 2015. Landsat Seen as Stunning Return on Public Investment. Available at *https://landsat.gsfc.nasa.gov/article/landsat-seen-as-stunning-return-on-public-investment*
O'Conner, J. 2015. *NPIC: Seeing the Secrets and Growing the Leaders*. Alexandria VA: Acumensa Solutions, 272 pp.
Roy, D. P., M. A. Wulder, T. R. Loveland, C. E. Woodcock, R. G. Allen, M. C. Anderson, . . . and Z. Zhu. 2014. Landsat-8: Science and Product Vision for Terrestrial Global Change Research. *Remote Sensing of Environment,* Vol. 145, pp. 154–172.
Sivanpillai, R., B. K. Jones, and R. M. Lamb. 2017. Accessing Satellite Imagery for Disaster Response through the International Charter: Lessons Learned from the 2011 US Midwestern Floods. *Space Policy,* Vol. 42, pp. 54–61.
Stoney, W. E. 2008. *ASPRS Guide to Land Imaging Satellites*. Bethesda, MD: American Society for Photogrammetry and Remote Sensing.
U.S. Geological Survey, Department of Interior. 2016. *Landsat 8 (L8) Data Users Handbook*. LSDS-1574, Version 2.0, 98 pp.
Van Aardt, J., M. Luck, W. Luck, and J. D. Anderson. 2010. Remote Sensing Systems for Operational and Research Use. Chapter 18 in *Manual of Geospatial Science and Technology* (J. D. Bossler, ed.). Boca Raton, FL: CRC Press, pp. 319–361.

8 Active Microwave



MAJOR TOPICS TO UNDERSTAND

- Active Microwave
- Geometry of the Radar Image
- Look Direction and Look Angle
- Wavelength
- Penetration of the Radar Signal
- Polarization
- Interpreting Brightness Values
- Interferometric SAR

8.1 INTRODUCTION

This chapter introduces active microwave sensor systems or radars. An active microwave sensor forms an example of an *active sensor*—a sensor that illuminates the Earth's surface using its own energy. More specifically, an active microwave system broadcasts a directed pattern of microwave energy to illuminate a portion of the Earth's surface, then receives the portion scattered back to the instrument. This energy forms the basis for

the imagery we interpret. Because passive sensors (e.g., photography) form images from reflected solar radiation, their use is constrained by weather and time of day. In contrast, active sensors generate their own energy, so their use is subject to fewer constraints, and they can be used under a wider range of operational conditions. Some of the unique characteristics of active microwave systems include the following:

- All-weather capability, constrained only by extreme weather events
- Ability to operate free of atmospheric effects and time of day
- Ability to record precise details of the transmitted energy they generate (such as wavelength, phase, and polarization), which can be matched with characteristics of the returned signal
- Ability to acquire detailed imagery at great distances

The term *radar* originated as an acronym for *Radio Detection and Ranging* and was used in U.S. Navy terminology as radar technology assumed a significant role in World War II. (*Ranging* refers to the ability of radars to measure the time required for a transmitted signal to return to the transmitting antenna—basically, a way to measure distance with microwave energy.)

Radar imagery, first available for civil applications in the mid-1960s, holds a place in the history of remote sensing as the first widely used imaging system that could provide a synoptic perspective of large regions (Simpson, 1966). At the time, aerial photography of the day was largely acquired at low altitudes because aerial films suitable for high-altitude photography were not in widespread use; atmospheric scattering inhibited photography from altitude, and uneven illumination of aerial photography limited the use of aerial mosaics. Thus, the availability of active microwave systems greatly expanded the scope of remote sensing.

Active microwave systems *transmit* microwave signals and then capture the portion of the signal returned to the radar antenna (**Figure 8.1**, left). There are also *passive* microwave and thermal systems that capture microwave or thermal energy *emitted* from the Earth's surface (**Figure 8.1**, right). Thermal passive systems are discussed in Chapter 10.

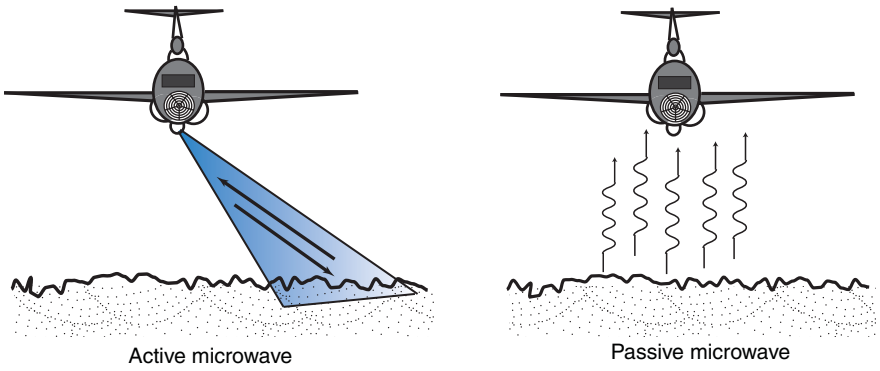


FIGURE 8.1 Active and passive remote sensing. Active microwave sensing (left) uses energy emitted by the sensor. Passive microwave systems (right) capture energy emitted from the surface at microwave wavelengths. Image by Susmita Sen.

8.2 ACTIVE MICROWAVE

The microwave region of the electromagnetic spectrum extends from wavelengths of about 1 mm to about 1 m. This region is, of course, far removed from those in and near the visible spectrum, where our direct sensory experience can assist in our interpretation of images and data. Thus, a formal understanding of the concepts of remote sensing is vital to understanding imagery acquired in the microwave region. Because many aspects of the behavior are outside our everyday experience, the study of microwave imagery is often a difficult subject for beginning students and requires more attention than is necessary for study of other regions of the spectrum.

Active Microwave Sensors

Active microwave sensors are *radar* devices—instruments that transmit a microwave signal and then receive its reflection as the basis for forming images of the Earth's surface. A typical imaging radar system includes a transmitter, a switching mechanism, an antenna, a receiver, and a data recorder. Two key components include the *transmitter,* which generates powerful repetitive pulses (electromagnetic waves at microwave wavelengths), and the *antenna,* which emits microwave radiation to illuminate a small region of the Earth's surface, and receives the returned signal.

The "ranging" capability of active microwave systems is achieved by measuring the time delay between the time a signal is transmitted toward the terrain and the time its echo is received. Through its ranging capability (possible only with active sensors), radar can accurately measure distances from the antenna to features on the ground. A second unique capability, also a result of radar's status as an active sensor, is its ability to detect differences between the transmitted and received signals with respect to frequency, polarization, and phase. Because the sensor transmits a signal of known characteristics, it is possible to compare the received signal with the transmitted signal. From such comparisons, imaging radars can detect changes in frequencies that provide capabilities that are not possible with other sensors.

Origins and History

Foundations for imaging radars were laid by scientists who first investigated the nature and properties of microwave and radio energy. Scottish physicist James Clerk Maxwell (1831–1879) first defined essential characteristics of electromagnetic radiation; his mathematical descriptions of the properties of magnetic and electrical fields prepared the way for further theoretical and practical work. In Germany, Heinrich R. Hertz (1857–1894) confirmed much of Maxwell's work and further studied properties and propagation of electromagnetic energy in microwave and radio portions of the spectrum. (The *hertz,* the unit for designating frequencies [Chapter 2], is named in his honor.) Hertz was among the first to demonstrate the reflection of radio waves from metallic surfaces and thereby began research that led to the development of modern radios and radars.

Subsequent research improved the electronics required to produce high-power transmissions over narrow wavelength intervals, to carefully time short pulses of energy, and to amplify reflected signals. These and other developments led to the rapid evolution of radar systems in the years prior to and during World War II. Due to the profound military significance of radar technology, World War II provided the context for rapid inno-

vations in radar capabilities. Thus, the development and refinement of imaging radars is linked to military reconnaissance, even though many current applications now focus on civil applications.

Characteristics of Active Microwave Systems

An imaging radar operates by transmitting electromagnetic radiation, then receiving the radiation reflected/scattered from the Earth's surface. (We might visualize returned radiation as an "echo" of the original transmitted energy.) It acquires data using an antenna array directed laterally to the side of the path of the aircraft or satellite, so that it images a strip of land parallel to, and usually at some distance from, the ground track. The key to this process relies on an *antenna*, a metallic structure that can transmit and receive a narrow beam of timed electromagnetic pulses.

Initially, imaging radars used externally mounted antennae, aimed to laterally view the region to the side of the flight path (hence, the name for these early systems—*side-looking airborne radar* [SLAR]). Real aperture SLAR systems (sometimes referred to as *brute force* systems) are the oldest, simplest, and least expensive of imaging radar systems. For these systems, the ability to acquire fine detail in the along-track axis derives from its ability to focus the radar beam to illuminate a small area. A long antenna, relative to wavelength, permits the system to focus energy on a small strip of ground, improving details recorded in the along-track dimension of the image (**Figure 8.2**).

By the 1960s, release of selected advanced military radar technologies for civil applications greatly improved the caliber of imagery available for everyday applications. These new systems, *synthetic aperture radar* (SAR), replaced earlier real aperture imaging radar technologies. SAR systems are designed to mimic the effect of a longer antenna using the motion of the sensor, allowing for higher resolution data with smaller antennas.

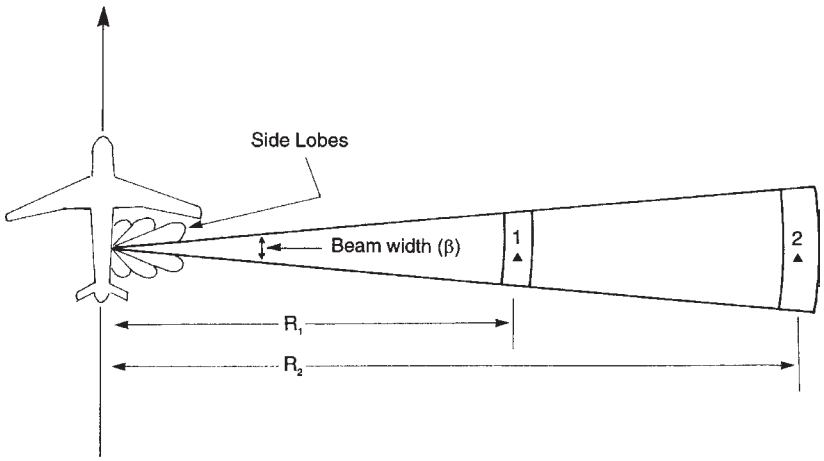


FIGURE 8.2 Azimuth resolution of a real aperture side-looking airborne radar system. The *beam width* (β) measures this quality of an imaging radar. Beam width, in relation to range (R), determines detail—region 1 at range R_1 will be imaged in greater detail than region 2 at greater range R_2. Also illustrated here are side lobes, smaller beams of microwave energy created because the antenna cannot be perfectly effective in transmitting a single beam of energy.

Although SAR systems have greater complexity and are more expensive to manufacture and operate than are real aperture systems, they can overcome some of the limitations inherent to real aperture systems and are therefore applied in a wider variety of applications, including observation from Earth-orbiting satellites. Today, because of their many advantages, SAR systems form the standard for operational imaging radars.

Here we provide an example of a SAR antenna from the Shuttle Imaging Radar-C/X-Band Synthetic Aperture Radar (SIR-C/X-SAR) to illustrate some basic features (Figure 8.3). This antenna, far larger than most SAR systems, is designed to provide a kind of multispectral radar image by collecting images at three different wavelengths. One antenna operates at L-band (23.5-cm wavelength), one at C-band (5.8-cm wavelength), and the third at X-band (3-cm wavelength). Energy transmitted by the L- and C-band antennae can be electronically aimed away from the aircraft to view different regions within a landscape. (The X-band antenna must be physically [mechanically] aimed to orient the X-band beam at the objects or terrain of interest.) For more information, see *https://earth.esa.int/web/eoportal/satellite-missions/s/sir-c*.

The same antenna serves both to transmit the radar signal and to receive its echo from the terrain, a capability known as *duplexing*. The *switch* initially directs the pulse to the antenna, and the echo from the terrain to the receiver. The *receiver* accepts the reflected signal from the antenna, then filters and amplifies as required. Returned signals undergo the A-to-D conversion process (as discussed in Chapter 4), to create a digital record.

The antenna's function is to transmit microwave energy and receive returned signals (the "echo" mentioned previously) from the landscape. The radar engineer's challenge is to design a narrowly focused beam of microwave energy to the Earth's surface, focused to illuminate a small area on the Earth's surface to provide detailed spatial resolution. SAR uses the motion of the antenna to achieve this (Figure 8.4).

Consider a SAR system that images the landscape as depicted in Figure 8.4. At *1*, the aircraft is positioned so that a specific region of the landscape is just barely outside the region illuminated by the SAR. At *2*, it is fully within the area of illumination. At *3*, it is just at the trailing edge of the illuminated area. Finally, at *4*, the aircraft moves so that the region falls just outside the area illuminated by the radar beam. SAR operates on the principal that objects within a scene are illuminated by the radar over an interval

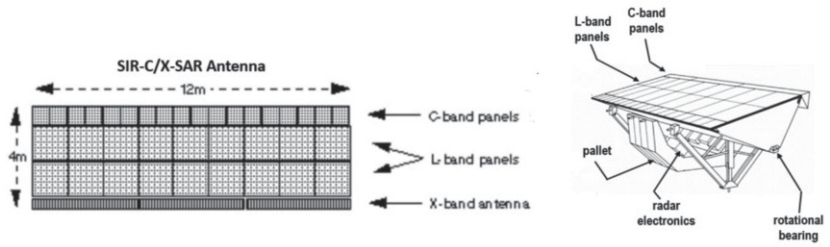


FIGURE 8.3 Synthetic aperture radar (SAR) antenna: large antenna systems (about 12 m in length) can be mounted in aircraft (or, in this instance, the Space Shuttle Cargo Bay, designed to transmit microwave radiation to the Earth's surface and to receive echoes from the Earth's surface). This is a depiction of an antenna from the SIR-C/X-SAR Project (NASA, with Italian Space Agency, and the German Space Agency).

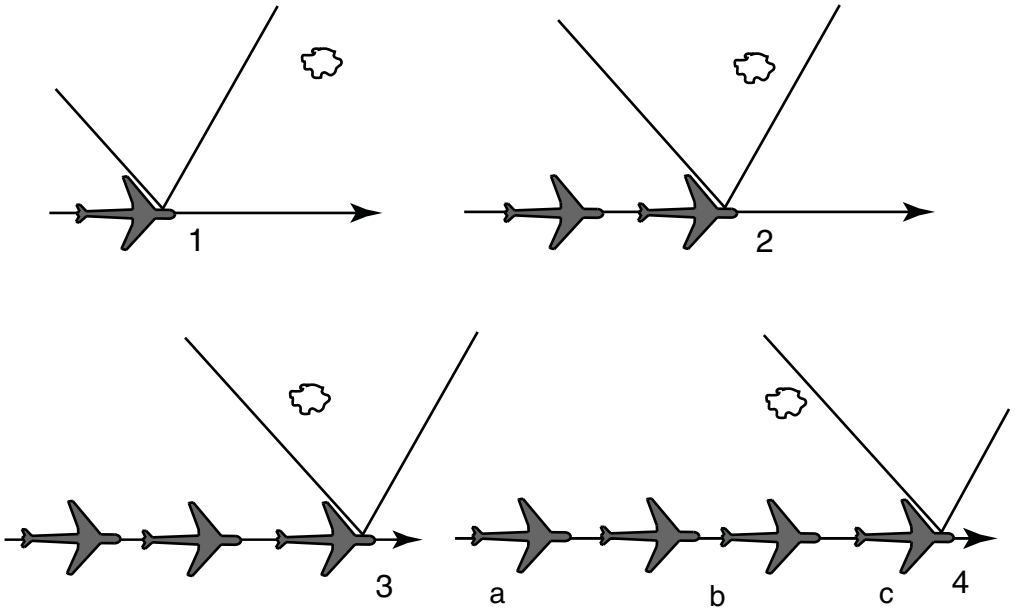


FIGURE 8.4 SAR systems accumulating a history of backscattered signals from the landscape as the antenna moves along the path a-b-c. Image by Susmita Sen.

of time, as the aircraft moves along its flight path. A SAR system receives the signal scattered from the landscape during this interval and saves the complete history of reflections from each object. Knowledge of this history permits reconstruction of the reflected signals as though they were received by a single antenna occupying physical space *abc*, even though they were in fact received by a much shorter antenna that was moved in a path along distance *1234*. Thus, the term *synthetic aperture* denotes the artificial length of the antenna, in contrast to the "real" aperture based on the physical length of the antenna used with real aperture systems.

In order to implement this strategy, it is necessary to define a practical means of assigning separate components of the reflected signal to their correct positions as the spatial representation of the landscape is re-created on the image. This process is, of course, extraordinarily complicated if each such assignment must be considered an individual problem in unraveling the complex history of the radar signal at each of a multitude of antenna positions. Fortunately, this problem can be solved in a practical manner because of the systematic changes in frequency experienced by the radar signal as it is scattered from the landscape. Objects within the landscape experience different frequency shifts in relation to their distances from the aircraft track. At a given instant, objects at the leading edge of the beam reflect a pulse with an increase in frequency (relative to the transmitted frequency) due to their position ahead of the aircraft, and those at the trailing edge of the antenna experience a decrease in frequency (Figure 8.5). This is the *Doppler effect*, often explained by analogy to the change in pitch of a train whistle heard by a stationary observer as a train passes by at high speed. As the train approaches, the pitch appears higher than that of a stationary whistle due to the increase in frequency of sound waves.

As the train passes the observer, then recedes into the distance, the pitch appears lower, due to the decrease in frequency. Radar, as an active remote sensing system, is operated with full knowledge of the frequency of the transmitted signal. As a result, it is possible to compare the frequencies of transmitted and reflected signals to determine the nature and amount of frequency shift. Knowledge of frequency shift permits the system to assign reflections to their correct positions on the image and to synthesize the effect of a long antenna.

One of SAR's unique and most useful characteristics is its ability to function at night or during inclement weather. SAR is often said to possess an "all-weather" capability, meaning that it can acquire imagery in all but the most severe weather conditions. The microwave energy used for SAR imagery is characterized by wavelengths long enough to escape interference from clouds and light rain. Because SAR systems are independent of solar illumination, missions using SAR can be scheduled at night or during morning or evening hours when solar illumination might be unsatisfactory for acquiring optical imagery. This capability is especially important for imaging radars carried by the Earth-orbiting satellites described later in this chapter because, for example, they can routinely capture conditions in polar regions that experience long periods without solar illumination.

SAR images typically provide crisp, clear representations of topography and drainage (**Figure 8.6**). Despite the presence of the geometric errors described in the following sections, radar images typically provide good positional accuracy, so they can provide the basis for accurate maps. Some of the most successful operational applications of SAR imagery have occurred in tropical climates, where persistent cloud cover has prevented acquisition of aerial photography and satellite imagery. Another important characteristic of SAR imagery is its synoptic view of the landscape. SAR's ability to clearly represent the major topographic and drainage features within relatively large regions at moderate image scales makes it a valuable addition to our repertoire of remote sensing imagery. Furthermore, because it acquires images in the microwave spectrum, SAR may show

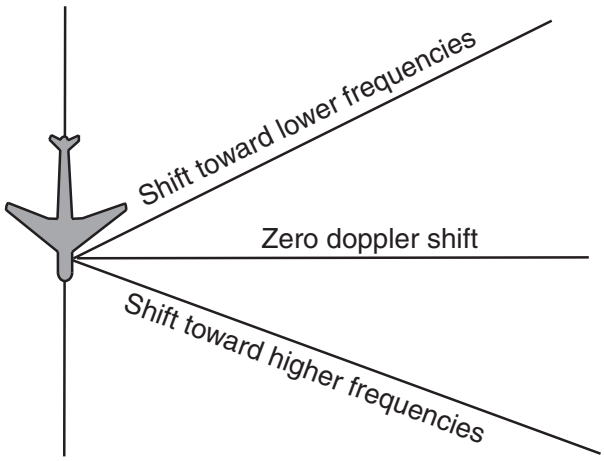


FIGURE 8.5 The Doppler effect: frequency shifts experienced by features within the field of view of the radar system. Image by Susmita Sen.

FIGURE 8.6 SAR image of a landscape near Chattanooga, Tennessee, September 1985 (X-band, horizontal–horizontal [HH] polarization). The shadowing to the west indicates that the aircraft is illuminating the landscape from the eastern edge of the image, casting terrain shadows in the mountain valleys. The image shows topographic relief, open water, dammed river channels, electric power generation, topographic relief, urban structures, transportation systems, forested terrain, open farmland, bridges, and drainage systems. From USGS.

detail and information that differ greatly from that of sensors operating in the visible and near-infrared spectra.

8.3 GEOMETRY OF THE RADAR IMAGE

The side-looking character of SAR imagery produces an image geometry that differs greatly from that of other remotely sensed images; this geometry establishes radar imagery as a distinctive form of remote sensing imagery. The basics of the geometry of a SAR image are illustrated in Figure 8.7. Here the radar beam is represented in vertical cross section as the fan-shaped figure at the side of the aircraft. The upper edge of the beam forms an angle with a horizontal line extended from the aircraft; this angle is designated as the *depression angle* of the far edge of the image. The upper and lower edges of the beam, as they intersect with the ground surface, define the edges of the radar image; the forward motion of the aircraft forms what is usually the "long" dimension of the strip of radar imagery. The smallest depression angle forms the *far-range* side of the image. The near-range region is the edge nearest to the aircraft. Intermediate regions between the two edges are sometimes referred to as *midrange* portions of the image. Steep terrain may hide areas of the imaged region from illumination by the radar beam, causing *radar shadow*. Note that radar shadow depends on topographic relief and the direction of the flight path in relation to topography. Within an image, radar shadow also depends on

depression angle so that, given equivalent topographic relief, radar shadow will be more severe in the far-range portion of the image, where depression angles are smallest, or for those radar systems that use shallow depression angles. A specific radar system is often characterized by a fixed range of depression angles.

Radar systems measure distance to a target by timing the delay between a transmitted signal and its return to the antenna (the *ranging* capability mentioned previously). Because the speed of electromagnetic energy is a known constant, the measure of time translates directly to a measure of distance from the antenna. Microwave energy travels in a straight path from the aircraft to the ground—a path that defines the *slant range* distance, as if one were to stretch a length of string from the aircraft to a specific point on the ground as a measure of distance. Image analysts prefer images to be presented in *ground-range* format, with distances portrayed in their correct relative positions on the Earth's surface. Because radars collect all information in the slant-range domain, radar images inherently contain geometric artifacts, even though the image display may ostensibly appear to match a ground-range presentation.

One such artifact is *radar layover* (Figure 8.8). At near range, the top of a tall object is closer to the antenna than its base. As a result, the echo from the top of the object reaches the antenna before the echo from the base. Because radar measures all distances with respect to time elapsed between transmission of a signal and the reception of its echo, the top of the object appears, in the slant-range domain, to be closer to the antenna than does its base. Indeed, it is closer, if only the slant-range domain is considered. However, in the ground-range domain (the context for correct positional representation and

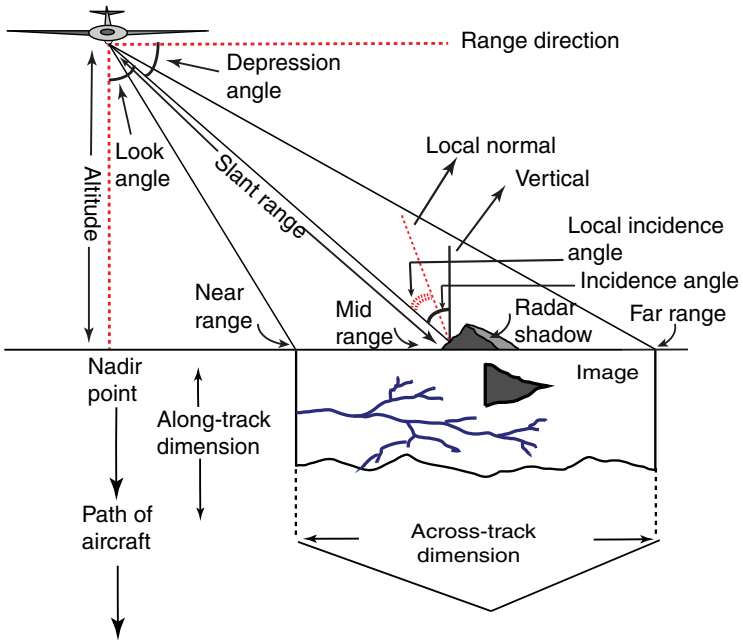


FIGURE 8.7 Geometry of a SAR system. The radar beam illuminates a strip of ground parallel to the flight of the aircraft; the reflection and scattering of the microwave signal from the ground forms the basis for the image. Image by Susmita Sen.

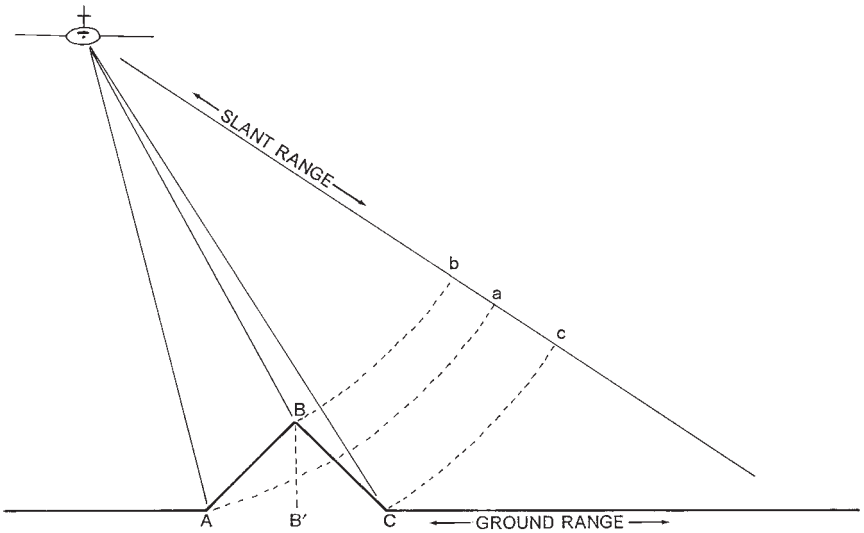


FIGURE 8.8 Radar layover. In the ground-range domain, AB and BC are equal. Because the radar can measure only slant-range distances, AB and BC are projected onto the slant-range domain, represented by the line *bac*. The three points are not shown in their correct relationship because the slant-range distance from the antenna to the points does not match their ground-range distances. Point B is closer to the antenna than is point A, so it is depicted on the image as closer to the edge of the image.

for accurate measurement), both the top and the base of the object occupy the same geographic position. In the slant-range domain of the radar image, they occupy different image positions—a geometric error perhaps analogous to relief displacement in the context of aerial photography.

Radar layover is depicted in Figure 8.8. Here the topographic feature *ABC* is shown with *AB* = *BC* in the ground-range representation. However, because the radar can position *A*, *B*, and *C* only by the time delay with relation to the antenna, it must perceive the relationships between *A*, *B*, and *C* as shown in the slant range (image plane). Here *A* and *B* are reversed from their ground-range relationships, so that *ABC* is now *bac* because the echo from *B* must be received before the echo from *A*.

A second form of geometric error, *radar foreshortening*, occurs in terrain of modest to high relief depicted in the mid- to far-range portion of an image (Figure 8.9). Here the slant-range representation depicts *ABC* in their correct relationships *abc*, but the distances between them are not accurately shown. Whereas *AB* = *BC* in the ground-range domain, *ab* < *bc* when they are projected into the slant range. Radar foreshortening tends to cause images of terrain features to appear to have steeper slopes than they do in nature on the near-range side of the image and to have shallower slopes than they do in nature on the far-range side of the feature (Figure 8.10). Thus, a terrain feature with equal fore- and backslopes may be imaged to have shorter, steeper, and brighter slopes than it would in a correct representation, and the image of the backslope would appear to be longer, shallower, and darker than it would in a correct representation. Because depression angle varies with position on the image, the amount of radar foreshortening in the image of a terrain feature depends not only on the steepness of its slopes, but also on its position on

the radar image. As a result, apparent terrain slope and shape on radar images are not necessarily accurate representations of their correct character in nature. Care should be taken when interpreting these features.

Figures 8.10 and 8.11 also show the sensitivity of radar signals to surface roughness, with rough areas depicted as bright surfaces and smoother areas appearing as dark features. In Figure 8.10, the mountains appear bright, while the surrounding basins and valleys (Death Valley, California) are dark and smooth. The green areas in the far right of the image are the Furnace Creek alluvial fan—gravel deposits that wash down from the mountains over time. Several other semicircular alluvial fans are visible along the mountain fronts. The sand dunes near Stovepipe Wells are near the center. A smooth floodplain enclosing Cottonball Basin appears dark (as a wrench-shaped feature).

Similarly, Figure 8.11 is a multispectral SAR image showing a region of the Mojave Desert in the vicinity of Barstow, California. Barstow is the lighter colored area in the lower left quadrant of the image, adjacent and below the sharp bend in the Mojave River, which is the blue thread that extends from the lower left to the upper edge of the image. At the center of the image, the V-shaped blue area is the Manix Basin, which includes the bed of the Mojave River. Within the basin, the orange circular and rectangular patches mark irrigated agricultural fields. Sparsely vegetated areas of gravel appear as blue; hills and rough gravel deposits appear mostly as shades of orange and brown. The distinctive dark patch at the upper left marks the smooth surface of Coyote Dry Lake. A set of electrical power line towers extends from the top of the image to the center. At the left side of the image, variations in color on the hills are caused by differences in surface roughness related to rock composition.

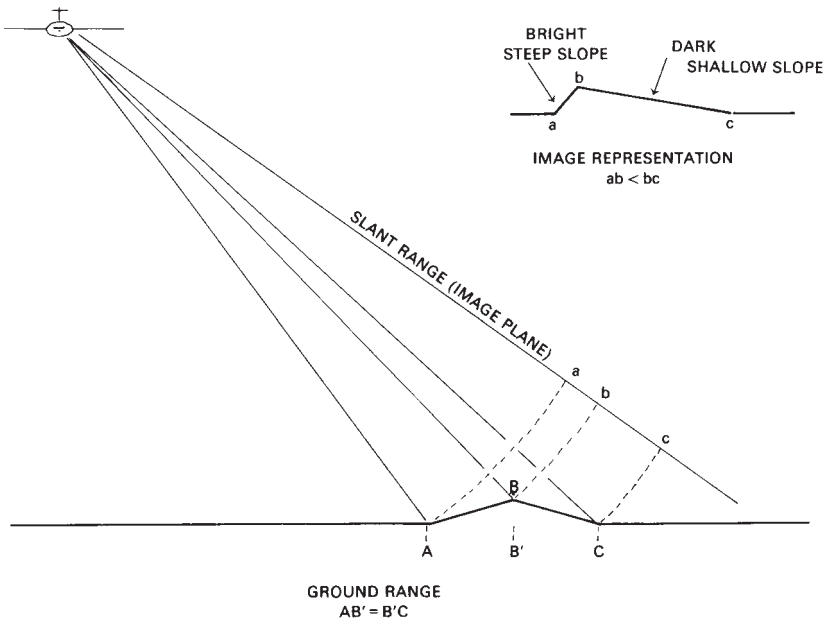


FIGURE 8.9 Radar foreshortening. Projection of *A, B,* and *C* into the slant-range domain distorts the representation of *AB* and *BC,* so that *ab* appears shorter, steeper, and brighter than it should be in a faithful rendition, and *bc* appears longer, shallower in slope, and darker than it should be.

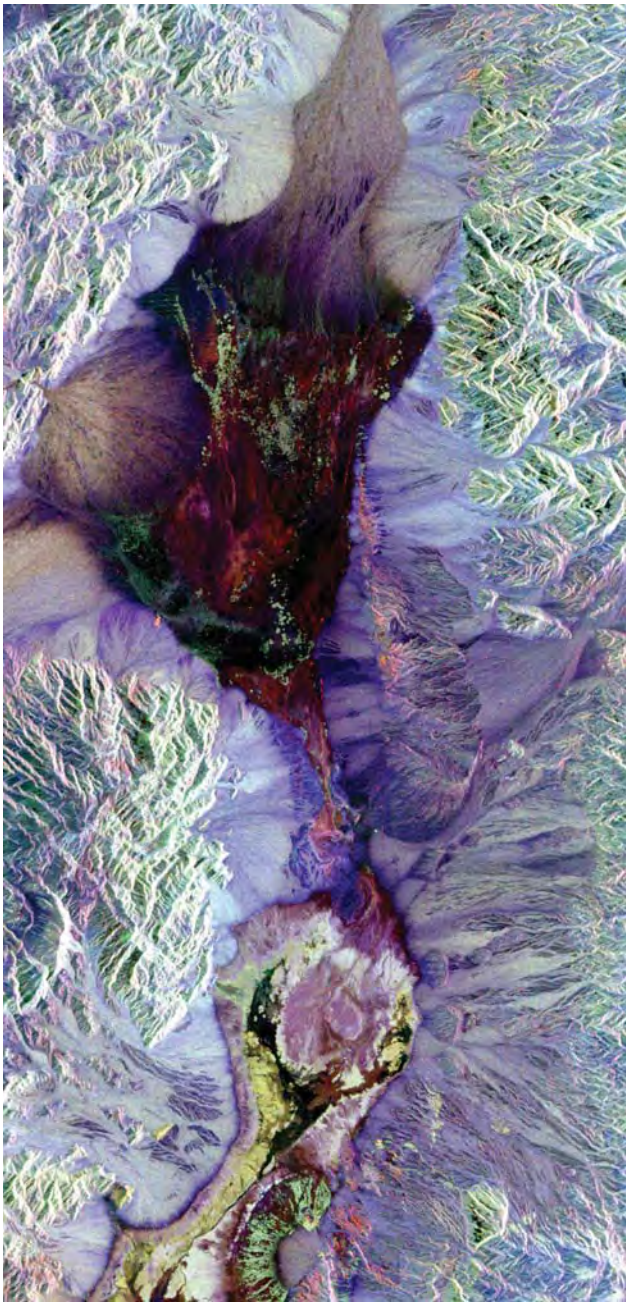


FIGURE 8.10 Radar foreshortening, SAR imagery, Death Valley, California. The unnatural appearance of the steep terrain illustrates the effect of radar foreshortening, when a radar system observes high, steep topography at steep depression angles. The radar observes this terrain from the right; radar foreshortening creates the compressed appearance of the mountainous terrain in this scene. Elevations range from a low of 70 m (230 ft) in the Valley to a mountaintop high of more than 3,300 m (10,800 ft) above sea level. In this image, colors represent varied radar channels: *red* = L band horizontally transmitted–horizontally received (LHH) polarization; *green* = L band horizontally transmitted–vertically received (LHV); and *blue* = C band horizontally transmitted–vertically received (CHV). From NASA, Jet Propulsion Laboratory.

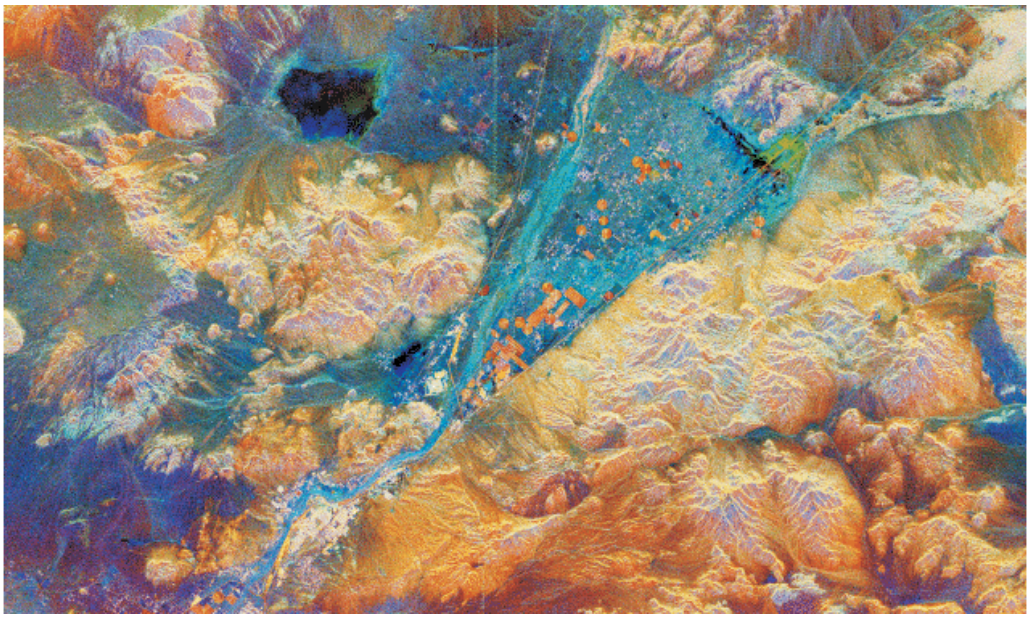


FIGURE 8.11 Multispectral SAR image near Barstow, California, in the Mojave Desert in April 1994. (*Note:* north is toward the upper left.) This is from NASA's Spaceborne Imaging Radar-C/X-Band SAR (SIR-C/X-SAR) system that was on the space shuttle Endeavor. Colors are assigned according to wavelength and polarization as in **Figure 8.10**. Red is the C-band (6 cm) horizontally transmitted–vertically received (HV); green is the L-band (24 cm) (HH); and blue is the ratio of C-band to L-band (HH). From NASA, Jet Propulsion Laboratory.

Scientists are using these radar data to address questions about the Earth's geology, including processes that form alluvial fans and their changes through time in response to climatic changes and earthquakes. As noted previously, radars illuminate the Earth with microwaves, allowing detailed observations at any time, regardless of weather or sunlight conditions.

8.4 LOOK DIRECTION AND LOOK ANGLE

Look Direction

Look direction, the direction at which the radar signal strikes the landscape, is important in both natural and human-made landscapes. In natural landscapes, look direction is especially important when terrain features display a preferential alignment. Look directions perpendicular to topographic alignment will tend to maximize radar shadow, whereas look directions parallel to topographic orientation will tend to minimize radar shadow. In regions of small or modest topographic relief, radar shadow may be desirable as a means of enhancing microtopography or revealing the fundamental structure of the regional terrain. The extent of radar shadow depends not only on local relief, but also on orientations of features relative to the flight path; those features positioned

in the near-range portion (other factors being equal) will have the smallest shadows, whereas those at the far-range edge of the image will cast larger shadows (Figure 8.12). In areas of high relief, radar shadow is usually undesirable, as it masks large areas from observation.

In landscapes that have been heavily altered by human activities, the orientation of structures and land-use patterns are often a significant influence on the character of the radar return, and therefore on the manner in which given landscapes appear on radar imagery. For instance, if an urban area is viewed at a look direction that maximizes the scattering of the radar signal from structures aligned along a specific axis, it will have an appearance quite different from that of an image acquired at a look direction that tends to minimize reflection from such features.

Look Angle

Look angle, the depression angle of the radar, varies across an image, from relatively steep at the near-range side of the image to relatively shallow at the far-range side (Figure 8.13). The exact values of the look angle vary with the design of specific radar systems, but some broad generalizations are possible concerning the effects of varied look angles. First, the basic geometry of a radar image ensures that the resolution of the image must vary with look angle; at steeper depression angles, a radar signal illuminates a smaller area than does the same signal at shallow depression angles. Therefore, the spatial resolution, at least in the across-track direction, varies with respect to depression angle. It has been shown that the sensitivity of the signal to ground moisture is increased as depression angle becomes steeper. Furthermore, the slant-range geometry of a radar image means that all landscapes are viewed by the radar at oblique angles. As a result, the image tends to record reflections from the sides of features. The obliqueness, and therefore the degree to which we view sides rather than tops of features, vary with look angle. In some landscapes, the oblique view may be very different than the overhead view to which we are accustomed in the use of other remotely sensed imagery (Figure 8.13). Such variations in viewing angle may contribute to variations in the appearance on radar imagery of otherwise similar landscapes.

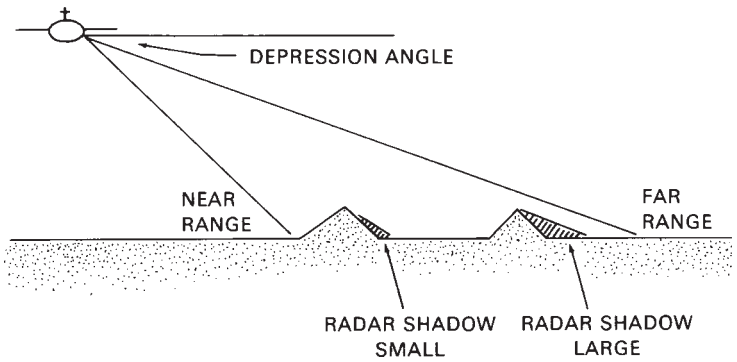


FIGURE 8.12 Radar shadow. Radar shadow increases as terrain relief increases and depression angle decreases.

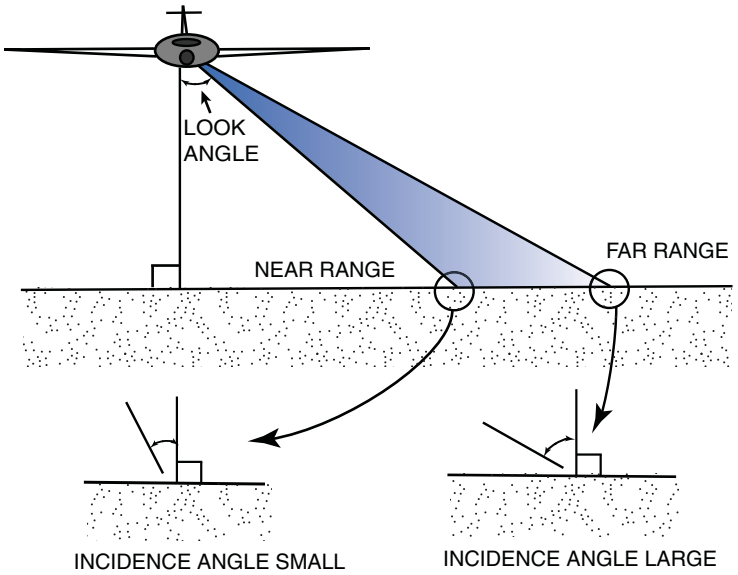


FIGURE 8.13 Look angle and incidence angle. Image by Susmita Sen.

8.5 WAVELENGTH

Imaging radars normally operate within a small range of wavelengths within the rather broad interval defined at the beginning of this chapter. Table 8.1 lists primary subdivisions of the active microwave region, as commonly defined in the United States. These divisions and their designations have an arbitrary, illogical flavor that is the consequence of their origin during the development of military radars, when it was important to conceal the

TABLE 8.1 Radar Band Wavelengths

	Wavelengths
P	30–107 cm
L	15–30 cm
S	7.5–15 cm
C	3.75–75 cm
X	2.40–3.75 cm
Ku	1.67–2.40 cm
K	1.18–1.67 cm
Ka	0.75–1.18 cm
VHF	1–10 m
UHF	10 cm–1 m

use of specific frequencies for given purposes. To preserve military security, the designations were designed as much to confuse unauthorized parties as to provide convenience for authorized personnel. Eventually, these designations became established in everyday usage, and they continue to be used even though there is no longer a secrecy concern. Although experimental radars can often change frequency, or sometimes even use several frequencies (for "multispectral radar"), operational systems are generally designed to use a single wavelength, or a specific set of predesignated wavelengths. Recent operational SAR systems have frequently used C, L, P, and X bands. The choice of a specific microwave band has several implications for the nature of the radar image. Penetration of the signal into the soil is, in part, a function of wavelength; for given moisture conditions, penetration is greatest at longer wavelengths. The longer wavelengths of microwave radiation (relative to visible radiation) mean that imaging radars are insensitive to the usual problems of atmospheric attenuation; usually only very heavy rain will interfere with transmission of microwave energy.

8.6 PENETRATION OF THE RADAR SIGNAL

In principle, radar signals are capable of penetrating what would normally be considered solid features, including vegetative cover and the soil surface. In practice, it is very difficult to assess the existence or amount of radar penetration in the interpretation of specific images. Penetration is assessed by specifying the *skin depth,* the depth to which the strength of a signal is reduced to 1/*e* of its surface magnitude, or about 37%. Separate features are subject to differing degrees of penetration; specification of the skin depth, measured in standard units of length, provides a means of designating variations in the ability of radar signals to penetrate various substances. In the absence of moisture, skin depth increases with increasing wavelength. Thus, optimum conditions for observing high penetration would be in arid regions, using long-wavelength radar systems. Penetration is also related to surface roughness and to incidence angle; penetration is greater at steeper angles and decreases as incidence angle increases. We should therefore expect maximum penetration at the near-range edge of the image and minimum penetration at the far-range portion of the image.

In a practical sense, the effect of varying SAR wavelengths is to highlight different features for different environments. Shorter wavelengths tend to scatter from the first surface they encounter, while longer wavelengths may penetrate some surfaces. For example, in heavily forested regions, images collected at shorter wavelengths may resemble a panchromatic aerial photograph; the X-band microwave signal is scattered from the first surface it encounters (the forest canopy). Longer wavelengths (such as P-band images) penetrate through the canopy to the reveal features below the vegetation canopy, and possibly in some instances below the soil surface, depicting terrain and land-use features not visible in imagery collected in the visible region of the spectrum.

Figure 8.14 illustrates an example over the Sahara Desert, collected from Shuttle Imaging Radar-A (the first imaging radar in space, launched in 1981) at L-band wavelengths (15–30 cm, Table 8.1). This relatively long wavelength, combined with the arid environment, favored deep penetration of the radar signal, and its return to the antenna from the ancient subsurface drainage system formed when this region of the Sahara experienced a more humid environment. To illustrate the contrast between the two environments, the graphic shows the elongated radar image, depicting the subsurface drainage

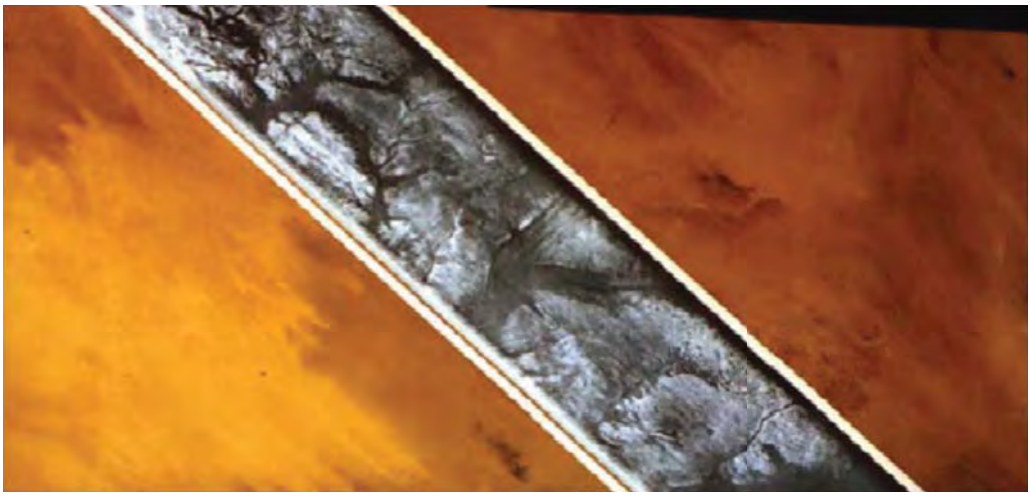


FIGURE 8.14 L-band Shuttle Imaging Radar-A image superimposed on an optical satellite image of the Sahara, from November 12, 1981. The L-band signal can penetrate through the arid desert sands, revealing the same region as in its ancient, humid environment. The image pair illustrates the ability of the L-band signal to penetrate the dry, sandy, surface and reveal the ancient drainage pattern of an earlier humid climate. From NASA, Jet Propulsion Laboratory.

pattern, superimposed over satellite imagery of the surface as seen in the visible spectrum. This image, and others like it, revealed the existence of an ancient drainage system and the possibility of archaeological evidence of human settlement. In subsequent years, similar imagery has provided the basis for archeological studies.

8.7 POLARIZATION

The *polarization* of a radar signal denotes the orientation of the field of electromagnetic energy emitted and received by the antenna. Radar systems can be configured to transmit either horizontally or vertically polarized energy and to receive either horizontally or vertically polarized energy as it is scattered from the ground. Unless otherwise specified, an imaging radar usually transmits horizontally polarized energy and receives a horizontally polarized echo from the terrain. However, some radars are designed to transmit horizontally polarized signals, but to separately receive the horizontally and vertically polarized reflections from the landscape. Such systems produce two images of the same landscape (**Figure 8.15**). One is the image formed by the transmission of a horizontally polarized signal and the reception of a horizontally polarized return signal. This is often referred to as the *HH image* or the *like-polarized* mode. A second image is formed by the transmission of a horizontally polarized signal and the reception of the vertically polarized return; this is the *HV image* or the *cross-polarized* mode.

By comparing the two images, the interpreter can identify features and areas that represent regions on the landscape that tend to depolarize the signal. Such areas will reflect the incident horizontally polarized signal back to the antenna as vertically polarized energy; that is, they change the polarization of the incident microwave energy. Such areas

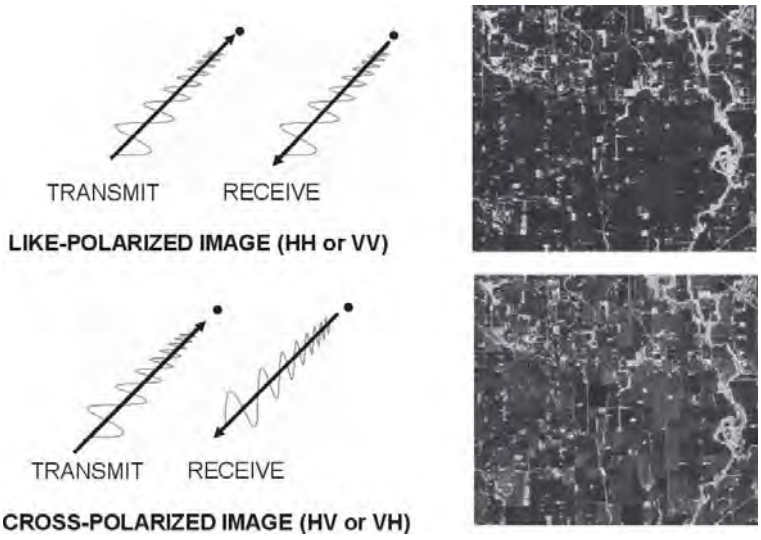


FIGURE 8.15 Radar polarization. Many imaging radars can transmit and receive signals in both horizontally and vertically polarized modes. By comparing the like-polarized and cross-polarized images, analysts can learn about the characteristics of terrain surfaces. From NASA, Jet Propulsion Laboratory.

can be identified as bright regions on the HV image and as dark or dark gray regions on the corresponding HH image. Their appearance on the HV image is much brighter due to the effect of depolarization; the polarization of the energy that would have contributed to the brightness of the HH image has been changed, so it creates a bright area on the HV image instead. Comparison of the two images therefore permits detection of those areas that are good depolarizers. This same information can be restated in a different way. A surface that is an ineffective depolarizer will tend to scatter energy in the same polarization in which it was transmitted; such areas will appear bright on the HH image and dark on the HV image. In contrast, a surface that is a "good" depolarizer will tend to scatter energy in a polarization different from that of the incident signal. Such areas will appear dark on the HH image and bright on the HV image. Causes of depolarization are related to physical and electrical properties of the ground surface. A rough surface, with respect to the wavelength of the signal, may depolarize the signal.

Another cause of depolarization is volume scattering from an inhomogeneous medium. Such scatter might occur if the radar signal is capable of penetrating beneath the soil surface (as might conceivably be possible in some desert areas where vegetation is sparse and the soil is dry enough for significant penetration to occur), where it might encounter subsurface inhomogeneities, such as buried rocks or indurated horizons.

8.8 INTERPRETING BRIGHTNESS VALUES

Each radar image is composed of many image elements of varying brightness (Figure 8.16). Variations in image brightness correspond, at least in part, to place-to-place changes within the landscape; through knowledge of this correspondence, the image

interpreter has a basis for making predictions, or inferences, concerning landscape properties. Unlike passive remote sensing systems, active systems illuminate the land with radiation of known and carefully controlled properties. Therefore, in principle, the interpreter should have a firm foundation for deciphering the meaning of the image because the only "unknowns" of the many variables that influence image appearances are the ground conditions—the object of study.

In practice, the interpreter of a radar image faces many difficult obstacles in making a rigorous interpretation of a radar image. First, most imaging radars are uncalibrated in the sense that brightness values on an image cannot be quantitatively matched to backscattering values in the landscape. Typically, returned signals from a terrain span a broad range of magnitudes from very low to very high; the ranges in values are often so large that they exceed the ability of display systems to accurately portray the actual range of values, so the full dynamic range of the brightnesses cannot be viewed. Furthermore, the features that compose even the simplest landscapes have complex shapes and arrangements, and are formed from diverse materials of contrasting electrical properties. As a result, there are often few detailed models of the kinds of backscattering that should in principle be expected from separate classes of surface materials. Direct experience and intuition are not always reliable guides to interpretation of images acquired outside the visible spectrum. In addition, many images are acquired at very shallow depression angles. Since interpreters gain experience from their observations at ground level, or from studying overhead aerial views, they may find the oblique radar view from only a few degrees above the horizon difficult to interpret.

Speckle

SAR images are subject to fine-textured effects that can create a grainy salt-and-pepper effect when inspected in detail. Speckle (formally known as *random fading noise*) is

FIGURE 8.16 Two examples of radar images illustrating their ability to convey detailed information about quite different landscapes. Left: Agricultural fields of the Maricopa Agricultural Experiment Station, Phoenix, Arizona (Ku band, spatial resolution about 1 m); varied tones and textures distinguish separate crops and growth states. Right: Structures near athletic fields, University of New Mexico, Albuquerque (Ku band, spatial information about 1 m). Varied brightness and tones convey information about diffuse surfaces, specular reflection, and corner reflectors, each of which carries specific meaning within the context of different landscapes. From Sandia National Laboratories. Used by permission.

created by illumination of individual scattering facets that are too small relative to the wavelength of the system to be individually resolved. Because the radar signal is coherent (transmitted at very narrow range of wavelengths, in phase), the maxima and minima of energy scattered by small features tend either to reinforce each other (*constructive interference*) or to suppress each other (*destructive interference*). When high-amplitude peaks of the waveform coincide, they create bright specks. Alternatively, when the high-amplitude peaks match the low-amplitude values of the waveforms, they tend to cancel each other, creating a dark speck. Because speckle constitutes a form of noise (it does not convey useful information), it is usually processed either by *multi-look* processes that illuminate the scene with slightly differing frequencies that produce independent returns, which then can be averaged to reduce the effect of the speckle, or by local averaging that smooths the speckled effect. The alternative filtering strategies are characterized either as *nonadaptive* or *adaptive*. Nonadaptive filters require less computation: they apply a single filter to the entire image uniformly, whereas the more complex adaptive filters adjust to match local properties of the terrain, thereby preserving natural edges and boundaries. All strategies, though intended to extract accurate estimates of backscatter, run the risk of eliminating genuine high-frequency information within the scene, so they must always balance benefits and losses within each image.

The Radar Equation

The fundamental variables influencing the brightness of a region on a radar image are formally given by the *radar equation*:

$$P_r = \frac{\sigma G^2 P_t \lambda^2}{(4\pi)^3 R^4} \qquad \text{(EQ. 8.1)}$$

Here, P_r designates the power returned to the antenna from the ground surface; R specifies the range to the target from the antenna; P_t is the transmitted power; λ is the wavelength of the energy; and G is the antenna gain (a measure of the system's ability to focus the transmitted energy). All of these variables are determined by the design of the radar system and are therefore known or controlled quantities. The one variable in the equation that is not thus far identified is σ, the backscattering coefficient; σ is, of course, not controlled by the radar system, but by the specific characteristics of the terrain surface represented by a specific region on the image. Whereas σ is often an incidental factor for the radar engineer, it is the primary focus of study for the image interpreter, as it is this quantity that carries information about the landscape. The value of σ conveys information concerning the amount of energy scattered from a specific region on the landscape as measured by σ^o, the *radar cross section*. It specifies the corresponding area of an isotropic scatterer that would return the same power as does the observed signal. The backscattering coefficient (σ^o) expresses the observed scattering from a large surface area as a dimensionless ratio between two areal surfaces; it measures the average radar cross section per unit area. σ^o varies over such wide values that it must be expressed as a ratio rather than as an absolute value.

Ideally, radar images should be interpreted with the objective of relating observed σ^o (varied brightnesses) to properties within the landscape. It is known that backscattering is related to specific *system variables,* including wavelength, polarization, and azimuth, in relation to landscape orientation and depression angle. In addition, *landscape parameters* are important, including surface roughness, soil moisture, vegetative cover, and micro-

topography. Because so many of these characteristics are interrelated, making detailed interpretations of individual variables is usually very difficult. This is in part due the extreme complexity of landscapes, which normally are intricate compositions of diverse natural and human-made features. Often many of the most useful landscape interpretations of radar images have attempted to recognize integrated units defined by assemblages of several variables rather than to separate individual components. The notion of "spectral signatures" is very difficult to apply in the context of radar imagery because of the high degree of variation in image tone as incidence angle and look direction change.

Moisture

Moisture in the landscape influences the backscattering coefficient through changes in the dielectric constant of landscape materials. (The *dielectric constant* is a measure of the ability of a substance to conduct electrical energy, an important variable determining the response of a substance that is illuminated with microwave energy.) Although natural soils and minerals vary in their ability to conduct electrical energy, these properties are difficult to exploit as the basis for remote sensing because the differences between dielectric properties of separate rocks and minerals in the landscape are overshadowed by the effects of even very small amounts of moisture, which greatly change the dielectric constant. As a result, the radar signal is sensitive to the presence of moisture both in the soil and in vegetative tissue; this sensitivity appears to be greatest at steep depression angles. The presence of moisture also influences effective skin depth; as the moisture content of surface soil increases, the signal tends to scatter from the surface. As moisture content decreases, skin depth increases, and the signal may be scattered from a greater thickness of soil.

Roughness

A radar signal that strikes a surface will be reflected in a manner that depends both on characteristics of the surface and properties of the radar wave, as determined by the radar system and the conditions under which it is operated. The *incidence angle* (θ) is defined as the angle between the axis of the incident radar signal and perpendicular to the surface that the signal strikes (Figure 8.17). If the surface is homogeneous with respect to its electrical properties, and "smooth" with respect to the wavelength of the signal, then the reflected signal will be reflected at an angle equal to the incidence angle, with most of the energy directed in a single direction (specular reflection).

For "rough" surfaces, reflection will not depend as much on incidence angle, and the signal will be scattered more or less equally in all directions (diffuse, or isotropic, scattering). For radar systems, the notion of a rough surface is defined in a manner considerably more complex than that familiar from everyday experience, as roughness depends not only on the physical configuration of the surface, but also on the wavelength of the signal and its incidence angle (Table 8.2). Consider the physical configuration of the surface to be expressed by the standard deviation of the heights of individual facets (Figure 8.17). Although definitions of surface roughness vary, one common definition describes a rough surface as one in which the standard deviation of surface height (h) exceeds one-eighth of the wavelength (λ) divided by the cosine of the incidence angle ($\cos\theta$):

$$h > \frac{\lambda}{(8cos\theta)} \qquad \text{(EQ. 8.2)}$$

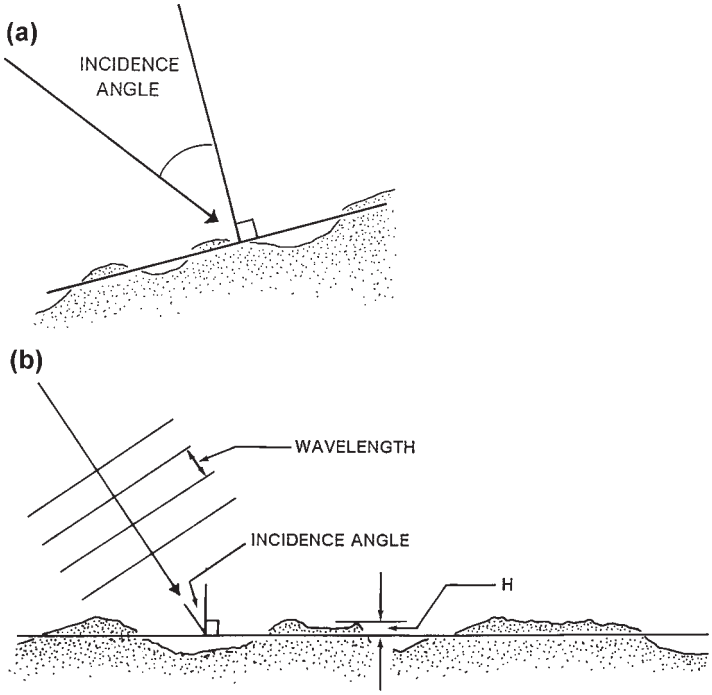


FIGURE 8.17 Measurement of incidence angle (a) and surface roughness (b).

TABLE 8.2 Surface Roughness Defined for Several Wavelengths

Roughness category	K-band (λ = 0.86 cm)	X-band (λ = 3 cm)	L-band (λ = 25 cm)
Smooth	$h < 0.05$ cm	$h < 0.17$ cm	$h < 1.41$ cm
Intermediate	$h = 0.05–0.28$ cm	$h = 0.17–0.96$ cm	$h = 1.41–8.04$ cm

Data from Jet Propulsion Laboratory (1982).

where h is the average height of the irregularities. In practice, this definition means that a given surface appears rougher as wavelengths become shorter. Also, for a given wavelength, surfaces will act as smooth scatterers as incidence angle becomes greater; that is, equal terrain slopes will appear as smooth surfaces as depression angle becomes smaller, as occurs in the far-range portions of radar images.

Corner Reflectors

The return of the radar signal to the antenna can be influenced not only by moisture and roughness, but also by the broader geometric configuration of targets. Objects that have complex geometric shapes, such as those encountered in an urban landscape, can create radar returns that are much brighter than would be expected based on size alone. This effect is caused by the complex reflection of the radar signal directly back to the antenna in a manner analogous to a ball that bounces from the corner of a pool table directly back to the player. This behavior is caused by objects classified as *corner reflec-*

tors, which often are in fact corner-shaped features (such as the corners of buildings and the alleyways between them in a dense urban landscape), but are also formed by other objects of complex shape. Corner reflectors are common in urban areas due to the abundance of concrete, masonry, and metal surfaces constructed in complex angular shapes (Figure 8.18). Corner reflectors can also be found in rural areas, formed sometimes by natural surfaces, but more commonly by metallic roofs of farm buildings, agricultural equipment, and items such as powerline pylons and guardrails along divided highways.

Corner reflectors are important in interpretation of the radar image. They form a characteristic feature of the radar signatures of urban regions, and they identify other features such as powerlines, highways, and railroads (Figure 8.19). It is important to remember that the image of a corner reflector is not shown in proportion to its actual size: the returned energy forms a star-like burst of brightness that is proportionately much larger than the size of the object that caused it. Thus, they can convey important information but do not appear on the image in their correct relative sizes.

8.9 INTERFEROMETRIC SAR

Interferometric SAR (*InSAR* or *IfSAR*) is a procedure that can extract elevation or change information from multiple SAR images of the same area taken from different positions. This capability is based on the unique status of SAR as an active sensor that can compare, in detail, its transmitted signal with its echo from the terrain. Thus far, our examination of SAR imagery has focused on interpretation of the intensities of energy scattered by the terrain. InSAR exploits another characteristic of the SAR energy—its phase information.

Imagine two SAR images of the same region acquired simultaneously from slightly different flight (or orbital) tracks. Because SAR is an active sensor, the characteristics

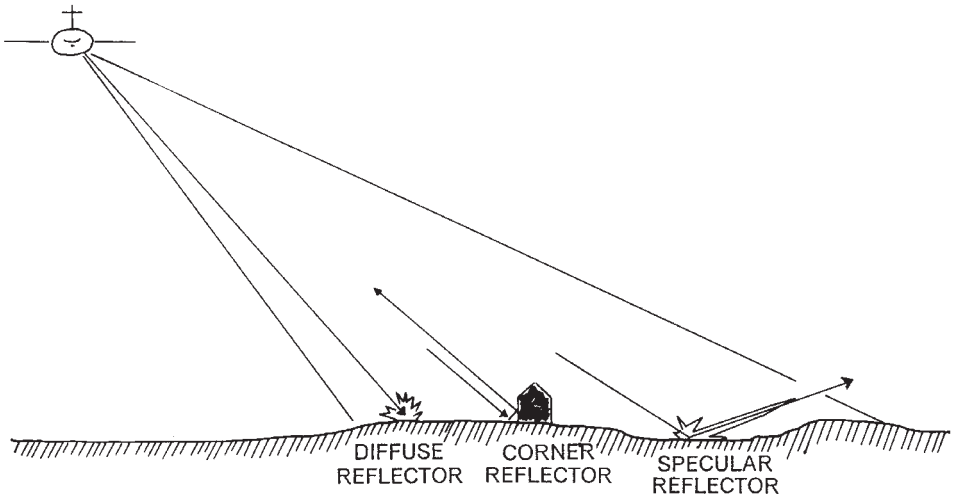


FIGURE 8.18 Three classes of reflectors important for interpretation of radar imagery. Very smooth surfaces cause specular reflection, typically away from the sensor, and will appear dark on the image (as in **Figure 8.16**). Corner reflectors, caused by abrupt geometric features on the landscape, will cause the signal to reflect back toward the sensor and appear very bright in the SAR image.

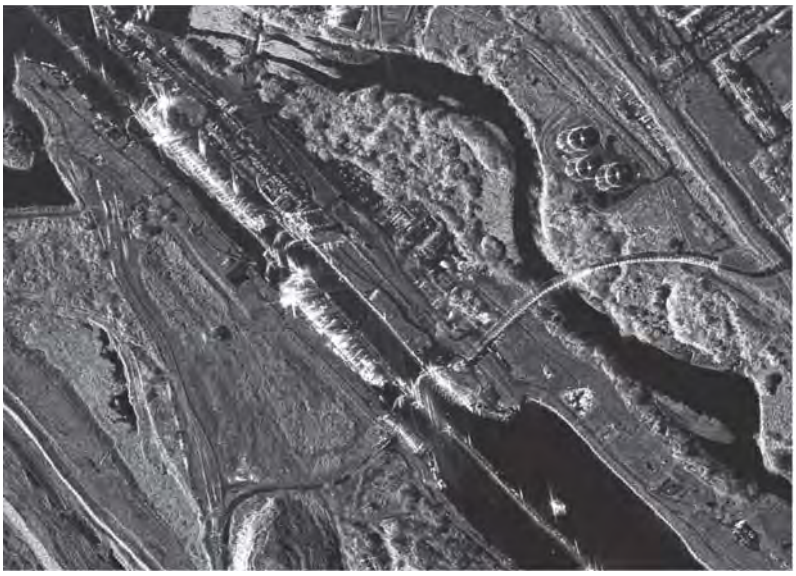


FIGURE 8.19 Corner reflectors seen as a subset of an X-band SAR (TerraSAR-X, 3.1-cm wavelength) in the Panama Canal. Ships and lock features can be seen as bright corner reflectors, while the water acts as a specular reflector (dark and smooth). Vertical displacement can also be clearly seen with the three circular towers on shore in the upper right quadrant of the image. From Airbus Defense and Space. Used by permission.

of the transmitted signal are known in detail. Since SAR illuminates the terrain with coherent radiation of known wavelengths, it is possible to evaluate not only variations in brightnesses of the returned signals, but also variations in the *phases* of the scattered signal—that is, the extent to which the peaks of the transmitted waveforms align with those of the scattered signal from the terrain. The composite formed by the interaction between phases of the transmitted and scattered signals is known as an *interferogram*, which shows the differences in phase between the transmitted and scattered signals. Because the positions of the two antennae are known, differences in phase can be translated to differences in terrain elevation.

The interferogram can be processed to reveal differences in topographic elevation that have generated the phase differences, thereby providing an accurate representation of the terrain. This analytical process, known as phase unwrapping, generates elevation differences expressed in multiples of the SAR wavelength (as the process relies on the mathematical foundations of measuring phase differences) that create contour-like fringes, known as *wrapped color* (**Figure 8.20**). The most notable example of use of the across-track configuration for the two antennas was the Shuttle Radar Topography Mission (SRTM), in which a second antenna was extended from the U.S. Space Shuttle orbiter to complement the primary antenna in the shuttle cargo bay. This system permits accurate mapping of a large portion of the Earth's terrain. For example, other variations of this strategy can yield other kinds of information. If the two images are collected at different times, such as different orbital passes (*repeat-pass interferometry*), phase differences can reveal such processes as tectonic uplift or subsistence, movement of glaciers, or changes in vegetative cover.

If two images are acquired from the same track at different times (e.g., if two antennas are mounted fore and aft in the same aircraft, or if images are acquired at different times within the same orbital track), it is possible to establish a *temporal baseline,* which provides an image pair that can reveal changes that occurred during the interval between the acquisition of the two images. In a military context, the objects in motion might be vehicles; in a scientific context, the motion might record ocean currents, ice flow in glaciers, or ice floes in polar oceans. The sensitivity of such analyses depends on the nature of the temporal baseline, so that very short temporal baselines can record rather rapid motion, such as vehicular motion, whereas longer baselines can detect slower speeds, such as movement of surface ice in glaciers. Paired antennas operated from a single platform are sensitive to velocities of centimeters per second (e.g., suitable for observing moving vehicles or ocean waves). Longer temporal baselines (e.g., separate passes within the same orbital path) are effective in recording slower speeds, perhaps centimeters per day, such as the motion of glacial ice.

A more widespread use of SAR interferometry depends on the acquisition of pairs of images of the same region acquired by imaging radars following separate tracks. The separation in distance of the two instruments establishes a *spatial baseline,* which permits measurement of topographic relief. The most favorable case arises when two images are acquired on parallel tracks, although analysis is possible with nonparallel tracks, provided that the angle of intersection is small.

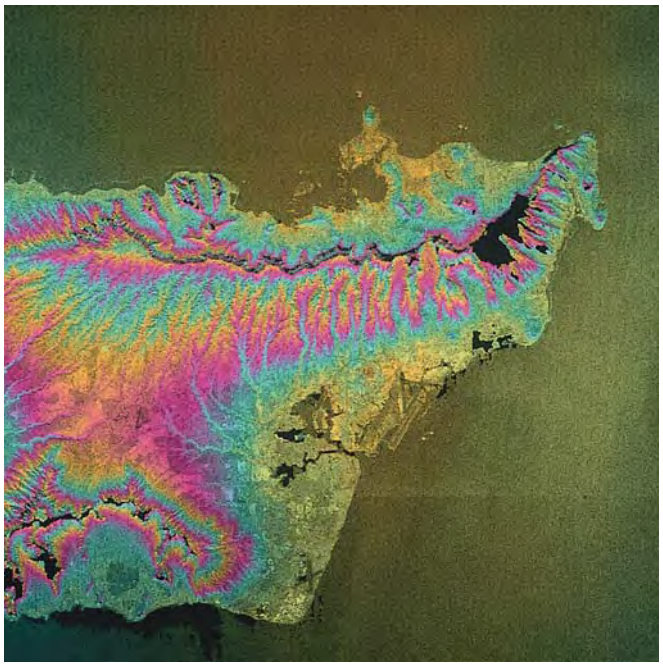


FIGURE 8.20 Radar interferometry, Honolulu, Hawaii, February 18, 2000. The Shuttle Radar Topography Mission (SRTM) uses two radar antennae to observe the same topography from separate positions. Processing of the two sets of microwave returns generates topographic data, shown here using the "wrapped color" bands to signify elevation classes. SRTM data have been processed to provide elevation data for a large proportion of the Earth's surface. From NASA, Jet Propulsion Laboratory.

8.10 SUMMARY

Radar imagery is especially useful because it complements the characteristics of images acquired in other portions of the spectrum. Aerial photography, for example, provides excellent information concerning the distribution and status of the Earth's vegetation cover. The information it conveys is derived from biologic components of plant tissues. However, aerial photography gives us little direct information about the physical structure of the vegetation. In contrast, although active microwave imagery provides no data about the biologic component of the plant cover, it does yield detailed information concerning the physical structure of plant communities (Figure 8.21).

Scientists working with radar remote sensing have been interested for years in the possibility of observing the Earth by means of imaging radars carried by Earth satellites. Whereas real aperture systems cannot be operated at satellite altitudes without unacceptably coarse spatial resolution or the use of impractically large antennas, the synthetic aperture principle permits compact radar systems to acquire imagery of fine spatial detail

FIGURE 8.21 Contrast between aerial (top) and X-band SAR (bottom) images of the terrain bordering the Rio Grande Valley, south of Albuquerque near Los Lunas, New Mexico. The wooded region borders the floodplain, buffering nearby landscapes from flooding. The images also show the levee/canal that directs the flow of water to the landscapes bordering the river channel. The image shows agricultural lands bordering the channel. At distances farther from the river channel, the fertile terrain visible here forms an abrupt edge with arid landscapes without the structures, roadways, and agricultural lands visible here. Note these images are oriented with the west at the top. Top: Google Earth image, from Landsat/Copernicus. Bottom: X-band SAR, from Sandia National Laboratories.

at very high altitudes. This capability, combined with the ability of imaging radars to acquire imagery in darkness, through cloud cover, and during inclement weather, provides the opportunity for development of a powerful remote sensing capability, with the potential to observe large areas of the Earth's ocean and land areas that might otherwise be unobservable because of remoteness and atmospheric conditions.

We would not expect to replace aerial photography or optical satellite data with radar imagery, but we could expect to be able to combine information from microwave imagery to acquire a more complete understanding of the character of the vegetative cover. Thus, the value of any sensor must be assessed not only in the context of its specific capabilities, but also in the context of its characteristics relative to other sensors.

REVIEW QUESTIONS

1. List advantages of the use of radar images relative to images from aerial photography and Landsat MSS. Can you identify disadvantages?
2. Imaging radars may not be equally useful in all regions of the Earth. Can you suggest certain geographic regions where they might be most effective? Are there other geographic zones where imaging radars might be less effective?
3. Radar imagery has been combined with data from other imaging systems, such as the Landsat MSS, to produce composite images. Because these composites are formed from data from two widely separated portions of the spectrum, together they convey much more information than either image can alone. Perhaps you can suggest (from information already given in Chapters 3 and 6) some of the problems encountered in forming and interpreting such composites.
4. Why might radar images be more useful in many less developed nations than in industrialized nations? Can you think of situations in which radar images might be especially useful in the industrialized regions of the world?
5. A given object or feature will not necessarily have the same appearance on all radar images. List some of the factors that will determine the texture and tone of an object as it is represented on a radar image.
6. Why are the steep depression angles of SAR inappropriate for many land areas but well suited to oceans? Can you think of advantages for use of steep depression angles in some land regions?
7. What problems would you expect to encounter if you attempted to prepare a mosaic from several radar images?
8. Why are synthetic aperture radars required for radar observation of the Earth by satellite?
9. Why is the shuttle imaging radar so important in developing a more complete understanding of interpretation of radar imagery?

REFERENCES

Born, G. H., J. A. Dunne, and D. B. Lane. 1979. Seasat Mission Overview. *Science,* Vol. 204, pp. 1405–1406.

Brown, W. M., and L. J. Porcello. 1969. An Introduction to Synthetic Aperture Radar. *IEEE Spectrum,* Vol. 6, No. 9, pp. 52–62.

Chan, Y. K., and V. C. Koo. 2008. An Introduction to Synthetic Aperture Radar (SAR). *Progress in Electromagnetics Research B,* Vol. 2, pp. 27–60.

Elachi, C., W. E. Brown, J. B. Cimino, T. H. Dixon, D. L. Evans, J. P. Ford, and R. S. Saunders. 1982. Shuttle Imaging Radar Experiment. *Science,* Vol. 218, No. 4576, pp. 996–1003.

Elachi, C. 1982. Radar Images from Space. *Scientific American,* Vol. 247, pp. 54–61.

Evans, D. L., J. J. Plant, and E. R. Stofan. 1997. Overview of the Spaceborne Imaging Radar-C/X band Synthetic Aperture Radar (SIR-C/X-SAR) Missions. *Remote Sensing of Environment,* Vol. 59, pp. 135–140.

Ford, J. P., J. B. Cimino, and C. Elachi. 1983. *Space Shuttle Columbia Views the World with Imaging Radar: The SIR-A Experiment.* JPL Publication 82–95. Pasadena, CA: Jet Propulsion Laboratory, 179 pp.

Gens, R. 1999. Quality Assessment of Interferometrically Derived Digital Elevation Models. *International Journal of Applied Earth Observation and Geoinformation,* Vol. 1, pp. 102–108.

Gens, R., and J. L. Vangenderen. 1996. SAR Interferometry: Issues, Techniques, Applications. *International Journal of Remote Sensing,* Vol. 17, pp. 1803–1835.

Henderson, F. M., and A. J. Lewis (eds.). 1998. *Principles and Applications of Imaging Radars* (*Manual of Remote Sensing,* Vol. 2, 3rd ed.). New York: John Wiley, 866 pp.

Jensen, H., L. C. Graham, L. J. Porcello, and E. M. Leith. 1977. Side-Looking Airborne Radar. *Scientific American,* Vol. 237, pp. 84–95.

Jet Propulsion Laboratory. 1982. *The SIR-B Science Plan.* JPL Publication 82–78. Pasadena, CA: Jet Propulsion Laboratory, 90 pp.

Jordan, R. L., B. L. Honeycutt, and M. Werner. 1991. The SIR-C/X-SAR Synthetic Aperture Radar System. *Proceedings of IEEE,* Vol. 79, pp. 827–838.

Kellendorfer, J. M., and K. C. MacDonald. 2009. Active and Passive Microwave Systems. Chapter 13 in the *SAGE Handbook of Remote Sensing.* (Warner, T. A., M. D. Nellis, and G. M. Foody, eds.). Washington, DC: SAGE, pp. 179–198.

Kimura, H., and Y. Yamaguchi. 2000. Detection of Landslide Areas Using Satellite Radar Technology. *Photogrammetric Engineering and Remote Sensing,* Vol. 66, pp. 337–344.

Maden, S., and H. A. Zebker. 1998. Imaging Radar Interferometry. Chapter 6 in *Principles and Applications of Imaging Radars* (*Manual of Remote Sensing,* Vol. 2, 3rd ed.) (F. M. Henderson and A. J. Lewis, eds.). New York: John Wiley, pp. 359–380.

McCauley, J. F., G. G. Schaber, C. S. Breed, M. J. Grolier, C. V. Haynes, B. Issawi, . . . and R. Blom. 1982, December 3. Subsurface Valleys and Geoarcheology of the Eastern Sahara. *Science,* Vol. 218, No. 4576, pp. 1004–1007.

Page, R. M. 1962. *The Origin of Radar.* New York: Doubleday, 169 pp.

Raney, R. K., A. P. Luscombe, E. J. Langham, and S. Ahmed. 1991. RADASAT. *Proceedings of IEEE,* Vol. 79, pp. 839–849.

Sabins, F. F. 1983. Geologic Interpretation of Space Shuttle Radar Images of Indonesia. *AAPG Bulletin,* Vol. 67, pp. 2076–2099.

Settle, M., and J. V. Taranick. 1982. Use of the Space Shuttle for Remote Sensing Research: Recent Results and Future Prospects. *Science,* Vol. 218, pp. 993–995.

Simpson, R. B. 1966. Radar, Geographic Tool. *Annals, Association of American Geographers,* Vol. 56, pp. 80–96.

Skolnik, M. I. (ed.). 2008. *Radar Handbook* (3rd edition). New York: McGraw-Hill, 1328 pp.

Ulaby, F. T., R. K. Moore, and A. K. Fung. 1981. *Microwave Remote Sensing: Active and Passive,* Vol. 1. Norwood, MA: Artech House, 456 pp.

9 Lidar

MAJOR TOPICS TO UNDERSTAND

- Profiling Lasers
- Scanning (Imaging) Lidars
- Types of Lidar
- Lidar Data
- Selected Lidar Applications
- Lidar Data Formats

9.1 INTRODUCTION

Lidar (*light detection and ranging*) is an active ***remote sensing*** system (i.e., a system that can generate energy [light] to assess ground features). Lidars generate pulsed laser light that can measure distances and generate precise, three-dimensional data describing the Earth and its surface features.

Functional components of a lidar system include (1) a laser scanner, (2) a Global Positioning System (GPS) with its associated highly accurate clock, and (3) an inertial navigation system (INS), typically mounted on aircraft. The laser scanner transmits brief

laser pulses to the surface (up to about 300,000 per second), which are scattered back to the laser scanner. As the system receives returning pulses from the surface, it records the time interval required to reach the surface and return. The system can then calculate, for each pulse, the distance between the laser scanner and the surface. Lidars can use ultraviolet, visible, or near-infrared light to scan objects. Near-infrared and green wavelengths are most commonly used for the systems discussed in this chapter.

Lidar can be considered as a technology analogous to radar imagery, in the sense that both families of sensors are designed to transmit energy in a narrow range of frequencies, then receive backscattered energy to form images of the Earth's surface. Both classes of instruments are active sensors; that is, they provide their own sources of energy, which means they are independent of solar illumination. More importantly, both have characteristics of transmitted and returned energy (i.e., the timing of pulses, wavelengths, and angles), so that they can be used to assess not only the brightness of backscattered energy, but also its angular position, changes in frequency, and timing of reflected pulses. Such knowledge means that lidar data, much like data acquired by active microwave sensors, permit extraction of information describing terrain, structures, vegetative features, and other features not recorded by optical sensors.

Lidars are based on an application of *lasers,* using a form of *coherent* light—light that is composed of a very narrow band of wavelengths—very "pure" with respect to color. Whereas ordinary light (even if it transmits a specific color) is composed of many wavelengths, with a diverse assemblage of waveforms, a laser produces light that is in phase ("coherent"), comprised of a narrow range of wavelengths ("monochromatic") (**Figure 9.1**). Such light can be transmitted over large distances as narrow beams that will diverge only slightly, in contrast to light we observe in our everyday experience that disperses over distance.

The laser—an acronym for light amplification by stimulated emission of radiation—is an instrument that applies a strong electrical current to a "lasable" material, usually crystals or gases, such as rubies, CO_2, helium–neon, argon, and other less familiar materials. Such *lasable* materials have atoms, molecules, or ions that emit light as they return to a normal ground state after excitement by a stimulus, such as electricity or light. The emitted light forms the coherent beam described above. Each separate material provides a specific laser with its distinctive characteristics with respect to wavelength.

The laser provides an intense beam that does not diverge as it travels from the transmitter, a property that can favor applications involving heating, cutting, etching, or illu-

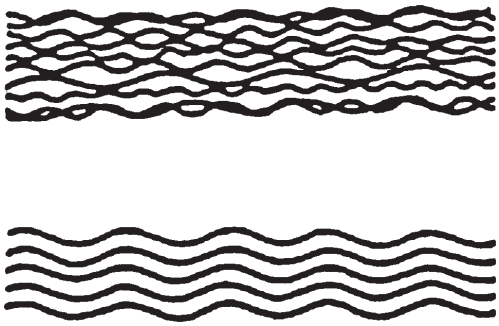


FIGURE 9.1 Normal (top) and coherent (bottom) light.

mination. Laser pointers, laser printers, CD players, scanners, bar code readers, and many other everyday consumer items are based on laser technology. Although imaging lasers do not use intense beams, they do exploit the focused, coherent nature of the beam to produce focused light. A laser uses mirrored surfaces to accumulate many pulses to increase the intensity of the light before it leaves the laser (Figure 9.2).

9.2 PROFILING LASERS

Lasers were invented in the late 1950s and were initially used for scientific inquiry and industrial applications. The first environmental uses of lidars were principally for *atmospheric profiling:* static lasers were mounted to point upward into the atmosphere to assess atmospheric aerosols. Solid particles suspended in the atmosphere directed a portion of the laser beam back to the ground, where its density indicated the abundance of atmospheric particles. Because lasers can measure the time delay of the backscatter, they can assess the clarity of the atmosphere over a depth of several kilometers, providing a measure of atmospheric quality.

The first airborne lasers were designed as *profiling lasers*—lasers aimed directly beneath the aircraft to illuminate a single region in the nadir position. (When used primarily to acquire topographic data, such instruments are known as *airborne laser altimeters.*) Forward motion of the aircraft carries the illuminated region forward to view a single track directly beneath the aircraft. Echoes from repetitive lidar pulses provide an elevation profile of the narrow region immediately beneath the aircraft (Figure 9.3). Although lidar profilers do not provide the image formats that we now expect, they provide a high density of observations and are used as investigative tools for researchers investigating topography, vegetation structure, hydrography, and atmospheric studies, among many applications.

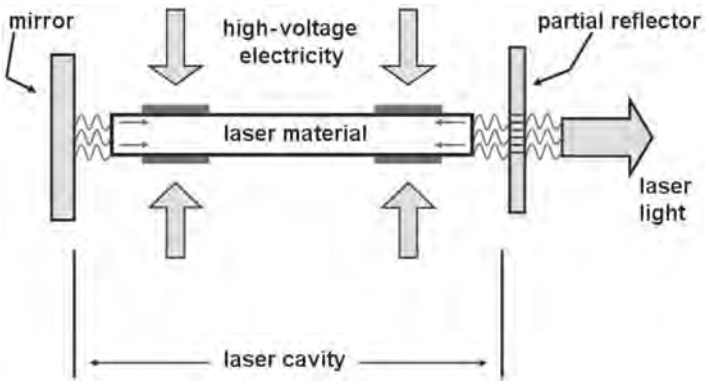


FIGURE 9.2 Schematic diagram of a simple laser. Energy, such as electricity, is applied to a substance, such as lasable gases (e.g., nitrogen, helium, neon) or materials (e.g., ruby crystals). When the materials return to their normal state, they emit coherent light, which is intensified before release by multiple reflections between the mirrored surfaces. Intensified light can then pass through the semitransparent mirror to form the beam of coherent light that is emitted by the instrument.

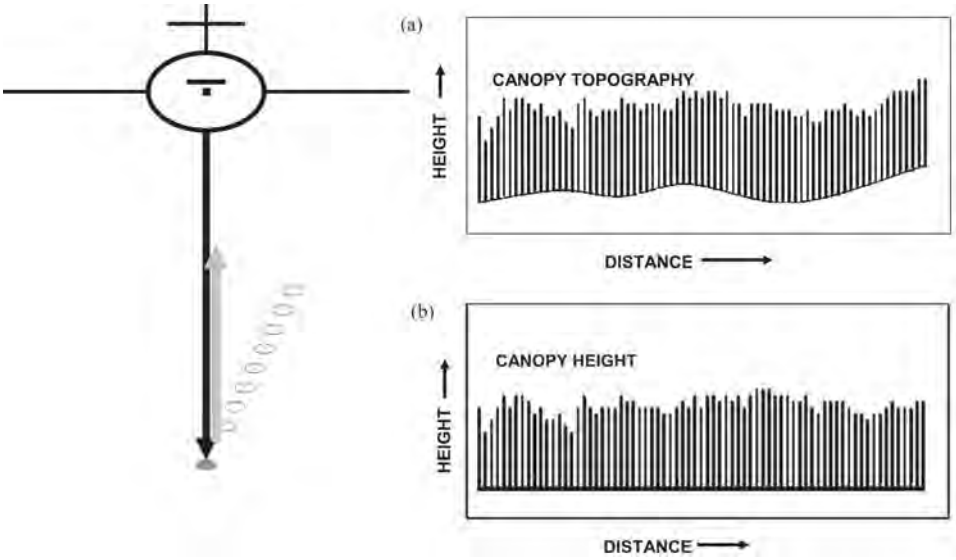


FIGURE 9.3 Schematic representation of an airborne laser profiler. (a) Acquisition of laser profile; (b) sample data gathered by a laser profiler, illustrating extraction of canopy height from the raw profile data.

9.3 SCANNING (IMAGING) LIDARS

By the late 1980s, several technologies matured and converged to create the context for development of precision scanning lidar systems that we now know. In this context, lidars assumed their current role as remote sensing instruments tailored for collection of imagery of the Earth's surface. *Inertial measurement units* (IMUs) enabled precise control and recording of orientation of aircraft (roll, pitch, and yaw). GPS could provide accurate records of geographic location of an aircraft as it acquired data. Furthermore, development of highly accurate clocks permitted the precise timing of lidar pulses required to create high-performance lidar scanning systems.

A lidar scanner can transmit up to 300,000 pulses per second, depending on the specific design and application. A scanning mirror directs the pulses back and forth across the image swath beneath the aircraft. The width of the swath is determined by the instrument's design and the operating conditions of the aircraft. Most imaging lidars use wavelengths in the visible or near-infrared regions of the electromagnetic spectrum. Common near-infrared wavelengths include 1,024 or 1,064 nm, which are sensitive to vegetation, relatively free from atmospheric scattering, and are absorbed by open water. Green wavelengths (e.g., 532 nm) are more common for applications with ice and water.

Several alternative designs for imaging lidar instruments are in use (Habib, 2010). Figure 9.4 presents a schematic representation of a typical lidar system. (1) The system's laser (coordinated by the electronic component) generates a beam of coherent light, transmitted by a fiber optic cable to (2) a rotating mirror, offset to provide a scanning motion. The laser light is directed to fiber optic cables that can be twisted to transmit the light as a linear beam (3). The oscillating motion of the mirror scans the laser beam from side to side along the cross-track axis of the image, recording many thousands of returns each

second. Because a lidar scanner is well integrated with GPS, IMUs, and timing systems, these pulses can be associated with specific points on the Earth's surface. As the reflected portion of the laser beam reaches the lidar aperture, it is received by another system of lenses (4) and channeled through fiber optic cables to another scanning lens. (5) It is then directed through an optical system to filter light before it is directed to (6) a receiving system to accept and direct the signal to the electronics components. The electronics coordinate timing of the pulses and permit matching of the signal with data from the inertial navigation system and GPS.

Together, such components permit the system to place each returned signal in its correct geographic position. Typically, two fiberglass bundles are configured to view the ground along a linear path. One transmits laser pulses, and an identical bundle receives the echoes. The system operates at such high speed that a large collection of pulses is received from each square meter on the terrain. The timing capability of the lidar scan permits accurate assessment of distance and elevation, which enables formation of an image with detailed and accurate representation of features in the scene.

9.4 TYPES OF LIDAR

There are two main types of lidar technology: analog and photon-counting technology. The vast majority of lidar data currently available comes from analog systems designed for terrain analysis. Both of these technologies can also be used for water-related analysis, referred to as bathymetric lidar. Most of the lidar systems designed for terrain analysis use laser energy from the NIR wavelengths (1,024 nm or 1,064 nm are common), although some do use green wavelengths.

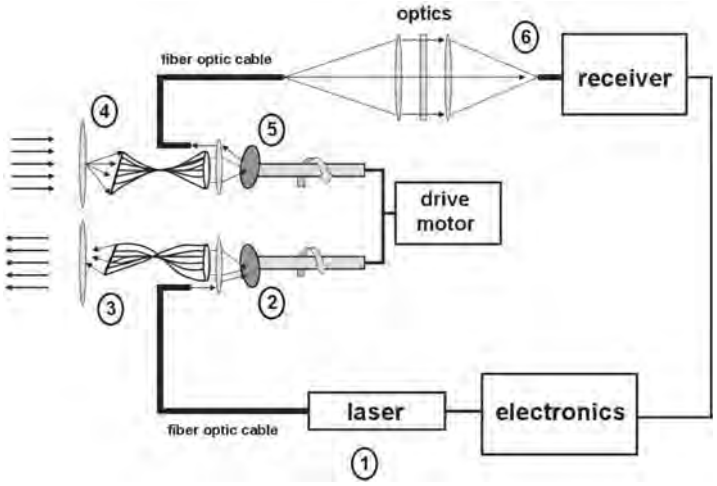


FIGURE 9.4 Schematic diagram of a lidar scanner. (1) The system's laser scanner (coordinated by the electronic component) generates a beam of light, transmitted by fiber optic cable to (2) a rotating mirror, offset to provide a scanning motion. The laser light is directed to a bundle of fiber optic cables that are twisted to provide a linear beam and then directed through a system of lenses toward the ground. The energy received back from the terrain is received by another system of lenses.

Analog Lidar

By far the most common type of lidar is analog lidar, where the returned lidar pulse is recorded as the amount of energy returned over time. Through processing, information about the speed of light, the scan angle, and the movement and orientation of the aircraft (made possible because of the GPS and IMU systems on the vehicle) are used to convert the data into elevation above the Earth's ellipsoid and the amount of energy returned from that elevation. These analog systems store the data as either the entire returned waveform, referred to as either *waveform lidar systems* or discrete returns (1, 2, or multiple returns) extracted at specific points within that waveform, referred to as *discrete lidar systems*. Most available lidar data today are from these analog systems, either waveform or discrete. The data are primarily collected from airborne systems, including UAVs, but they may also be collected from other vehicles, from backpack systems, or systems mounted on a tripod. These systems can collect profiles (typical of earlier lidar technology) or scans (most common today). Most of our discussion in this chapter refers to scanning analog system technology.

Photon-Counting Lidar

The second type of lidar system, referred to as photon-counting lidar, or quantum lid, uses quantum sensors to detect individual photons of light that are returned from the initial pulse. Photon-counting lidar has numerous technological advantages, with increased vertical sensitivity and the possibility of smaller instrumentation. However, the technology is relatively new, and there are far fewer systems with these types of data. One notable exception is the ATLAS instrument on the ICESat-2 satellite, which is designed to monitor ice sheet elevation and sea ice thickness. There are also a few airborne systems with this technology, which can provide data in incredible detail. However, because of the sensitivity of quantum sensors, the data tend to be very noisy, especially when collected from space, with significant returns from the atmosphere. This makes it difficult to use the data for the same types of land surface analysis that have become routine with analog lidar. Numerous ongoing research efforts are ongoing to remove the "noise" (unwanted returns) from the signal for specific applications. Because of the increased vertical sensitivity of these systems and the possibility for smaller instrumentation and other technological advantages, this type of lidar system will likely become prevalent in the coming years.

Bathymetric Lidar

Although bathymetric lidar is essentially the same technology we have described, the applications are very different. As discussed in Chapter 2, water typically absorbs most of the energy at the NIR wavelengths, so there are very few lidar returns from water if these wavelengths are used. However, as noted above, some systems use green light for the lidar energy (532 nm is common) or have multiple beams and capture both the green and NIR wavelength ranges. At green wavelengths, lidar can penetrate water, and so it may be used for bathymetric applications that capture the floor of the water body. Having both green and NIR wavelengths (typical for systems intended specifically for bathymetry) allows for the capture of the surrounding shoreline terrain and water surface (NIR) and

bottom of the water body (green). Note that water penetration by bathymetric lidar is further discussed in Chapter 20.

9.5 LIDAR DATA

Lidar systems do not provide imagery in the same sense that we have been discussing with aerial photography, optical, or SAR sensors. Rather, they provide locations and ancillary data at each location, usually as discrete points or waveforms that describe some portion of the returned laser pulse recorded by the lidar system. In the case of scanning lidar systems, the data are so dense that it is common to convert either the height information or some of the ancillary information (e.g., the intensity of the return energy) into a high-resolution raster product and examine it with the same types of software programs used for other raster data. These products may be referred to as lidar imagery or a specific lidar-derived product.

Figure 9.5 provides a schematic sketch of a lidar system in flight, scanning side-to-side as the aircraft flies forward along its planned (linear) flight path. The noticeable V-shaped gaps occur because of the simultaneous forward motion of the aircraft (Figure 9.5, top). In reality, actual scan lines are positioned quite close together and frequently

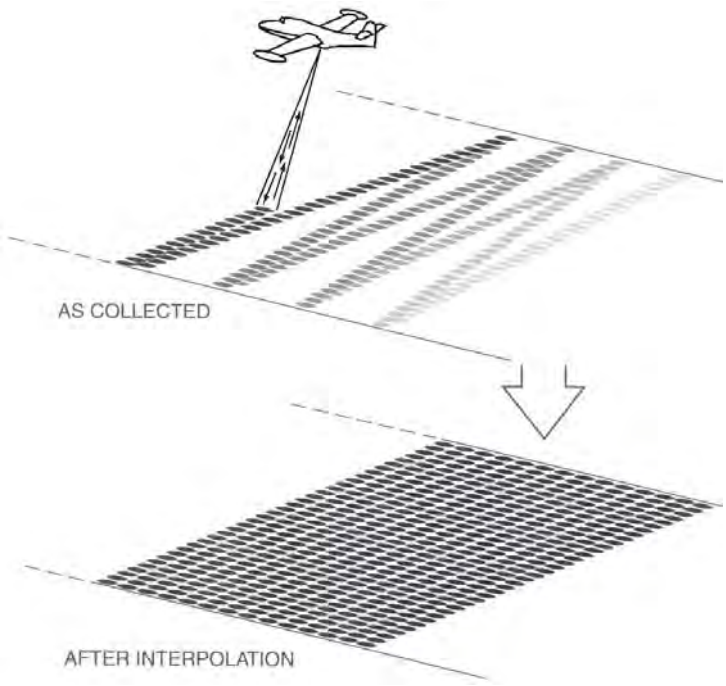


FIGURE 9.5 Acquisition of lidar data. Lidar systems acquire data by scanning in the patterns suggested by the top diagram; details vary according to specific systems. The pattern of returns can then be interpolated to generate a systematic array from the lidar data.

overlap with those of adjacent flight lines; they can be interpolated to form a dense, systematic array of pixels, as depicted in the lower section of Figure 9.5.

Discrete Analog Lidar Data

In the case of discrete lidar systems, each recorded lidar return can be precisely positioned in *xyz* space to provide a three-dimensional point cloud of position and associated attributes (intensity, scan angle, GPS time, etc.). The accuracy of these points will vary depending on the specifications of the system, which can be obtained from the manufacturer and the conditions on the ground. (*Note:* Forest cover and other conditions that cause GPS multipath errors can decrease locational point accuracy.) For small-footprint lidars (the most common type of lidar system), horizontal accuracy might be in the range of 20–30 cm, and vertical accuracy in the range of 15–20 cm. This enables the derivation of products from lidar data, such as digital elevation models of the ground surface, at comparable detail and positional accuracy to those acquired by photogrammetric analysis of aerial photographs.

The available terrain detail will vary based on the lidar pulse density. The lidar pulse rate refers to the number of laser pulses emitted by the lidar per second, which can be in the hundreds of thousands. This is dependent on the sensor specifications and has tended to increase over time, as technology improves. The pulse density refers to the number of lidar returns per unit of ground area and is dependent on many factors, including the pulse rate, altitude and speed of the plane, and the local terrain (Figure 9.6).

Within the lidar point cloud, *primary returns* (or *first returns*) originate from the first objects a lidar pulse encounters—which could be the upper surface of a vegetation canopy (Figure 9.6), buildings, objects, or the ground. In addition, portions of a pulse pass through gaps in the canopy or other pervious surfaces. Some of this energy may be returned from features within the encountered object (such as branches in a tree), of which some energy may eventually reach the ground. This energy creates echoes known as *secondary returns* (or *partial* or *multiple returns*). Therefore, for complex surfaces such as forests with multiple canopies, some portions of a pulse might be reflected from upper and middle portions of the canopy and other portions from the ground surface at the base (Figures 9.7 and 9.8).

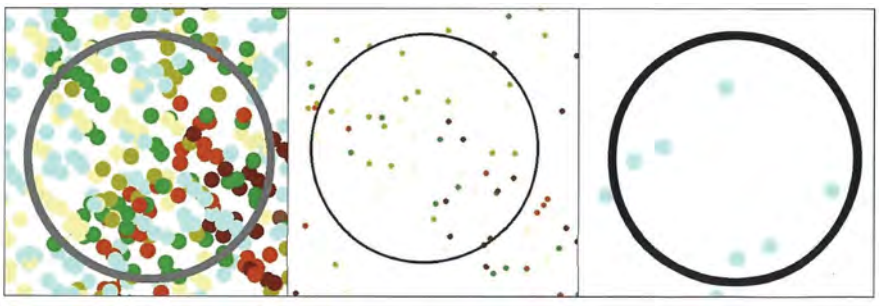


FIGURE 9.6 Representations of differing lidar pulse densities within a 1-m circle, illustrating variations in the objectives of the lidar mission. From Parece et al. (2016). Used by permission of *VirginiaView.net.*

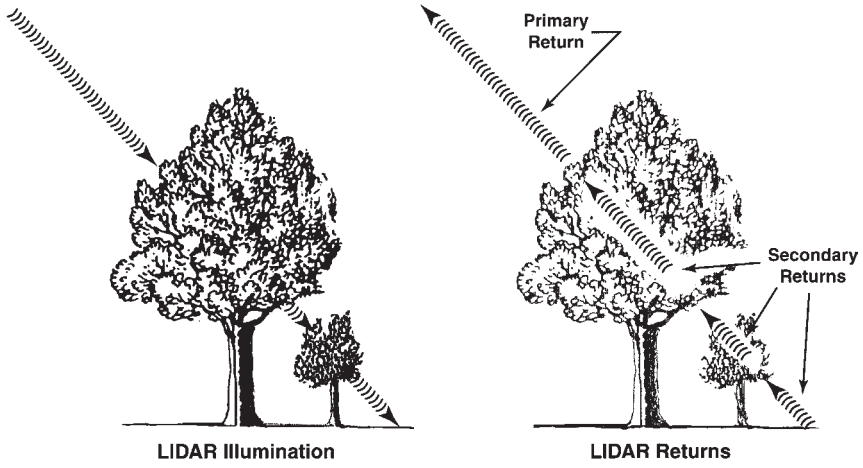


FIGURE 9.7 Schematic diagrams of primary and secondary lidar returns.

The resulting data is a three-dimensional point cloud (Figure 9.9). For some applications, particularly those related to vegetation analysis, researchers will analyze the point clouds themselves, to better understand the vertical distribution of points under different vegetation conditions (see Chapter 18). In the case of forest analysis, information such as canopy height, canopy openness, amount of leaf area, aboveground biomass, forest carbon, and other information about the physical shape and arrangement of the trees can be calculated from lidar point clouds.

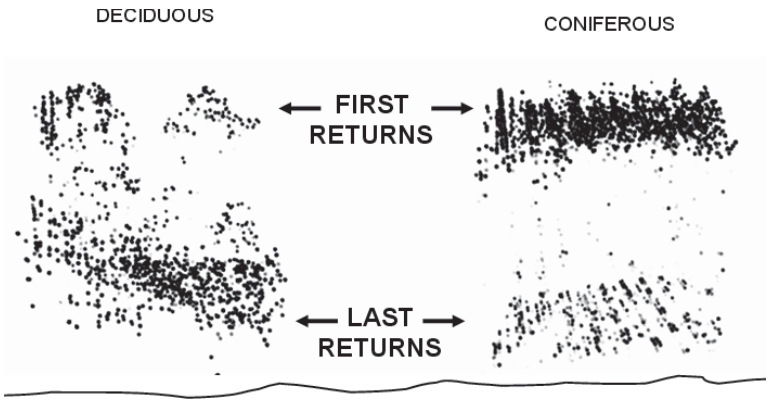


FIGURE 9.8 First and last lidar returns from a deciduous forest canopy (left) and a coniferous forest canopy (right), shown in a two-dimensional profile. Dots near the tops of the diagram represent returns that are received first (*primary returns*), when the pulse encounters the top or near the top of the forest canopy, and dots at the lower and central portions of the diagram represent returns received later (*secondary returns*). Note the contrast between dome-shaped canopies formed by crowns of the deciduous forest (left) and peaked crowns of the coniferous canopy (right). The coniferous forest has only sparse undergrowth, while the deciduous forest is characterized by abundant undergrowth. From Sorin Popescu. Used by permission.

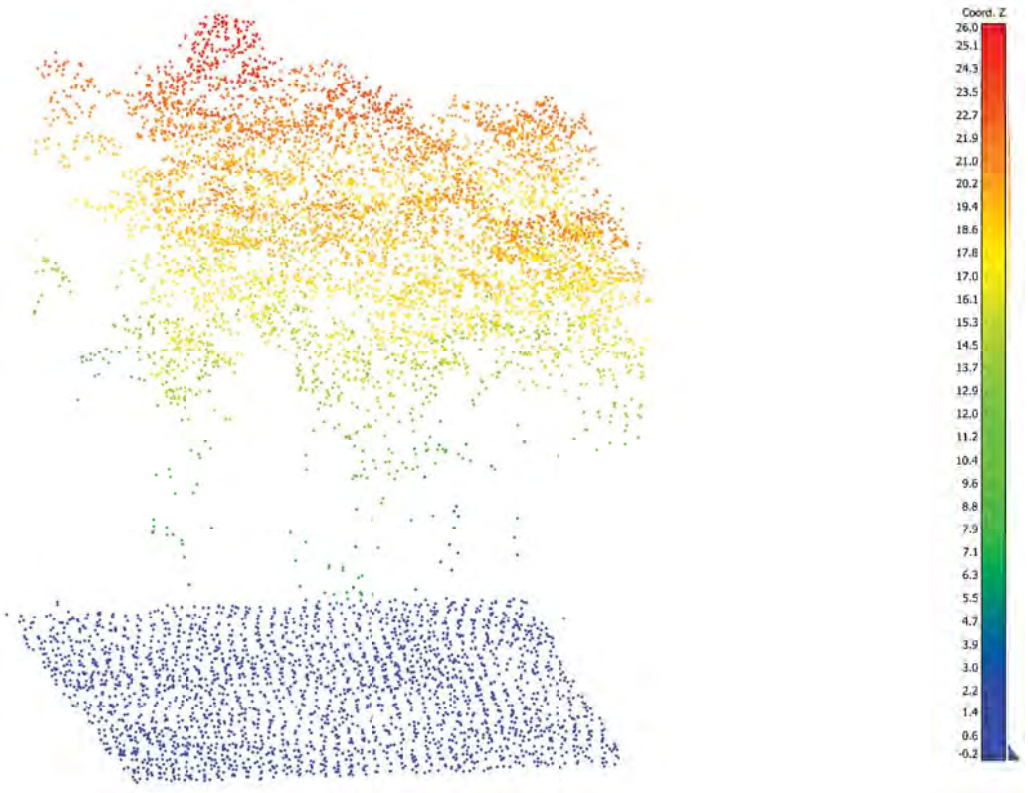


FIGURE 9.9 Normalized lidar point cloud for a deciduous forest. Ground level can be seen as dark blue points. Points at the top of the canopy are red. Normalization of the point cloud means that the ground elevation has been subtracted from all points in the cloud to remove the effect of terrain. The lidar is sourced from the National Ecological Observatory Network Mountain Lake Biological Station in Virginia. From Elizabeth M. Prior. Used by permission.

Figure 9.10 provides an example of a lidar profile (a slice extracted from the point cloud) of lidar data that contains multiple features of interest and illustrates the level of detail contained in the data. Details can be seen about the shape of individual tree crowns, the height and slope of building roofs, and the terrain along the profile. If desired, users can extract individual features and explore their shape in three dimensions, with measurements. In this example, the ground returns have been identified with a ground detection algorithm, and they are shown in brown. All other returns are white or gray.

Although the point cloud is useful for vegetation analysis, most applications of lidar data use raster-based products that are derived from the points through data processing. By far the most common product derived and used from lidar data is an interpolation of the ground returns to generate a digital elevation model (DEM) (Figure 9.11). Because lidar passes through vegetation and has high point density, lidar-derived DEMs have significant detail about the terrain surface and higher accuracy than other data sources that have been used in the past.

FIGURE 9.10 Example of a lidar profile. This lateral view depicts tree canopies, dwellings, and a gentle downhill slope left to right, revealing a slight variation in terrain relief. From Parece et al. (2016, pp. 159, 160). Used by permission of *VirginiaView.net*.

A closer view of the lidar-derived DEM from **Figure 9.11** illustrates the detailed terrain information that is available with lidar data (**Figure 9.12**). Numerous details can be observed on the shaded relief of the surface, including a four-lane highway, nearby quarry, forested areas, and agricultural lands with a corn crop. Note that the data are detailed enough to show individual trees, including along a fence line (left side of the hillshade model). The parallel strips visible near the upper right depict mature cornfields—another example of the detail recorded by this technology.

Many applications use digital elevation models and topographic maps that existed long before lidar data became common. Some of these applications use ancillary information, such as contour lines, to provide information about topographic relief, slope, and aspect (**Figure 9.13**). There are industry standards with respect to the required accuracy of these products for topographic mapping. It is common practice to derive this ancillary information from lidar data during the data preprocessing stage and provide it as a product with the lidar data (at or exceeding required accuracy standards), even though the lidar data and the lidar-derived DEMs contain significantly more detail.

For urban applications, it is also common to extract information such as building footprints and their associated heights from the lidar first returns, in addition to providing the ground elevation in the DEM (**Figure 9.14**). It is also possible to generate an elevation model of the surface, which includes the heights of objects; these products are

FIGURE 9.11 Top: Lidar-derived digital elevation model depicting terrain near Wytheville, Virginia, where brightness indicates relative elevation. Bottom: Lidar-derived shaded relief of the same area, generated using a hill-shading algorithm. This enables a view of the surface with shadow effects. From Virginia Department of Transportation. Copyright © 2003 Commonwealth of Virginia. Used by permission.

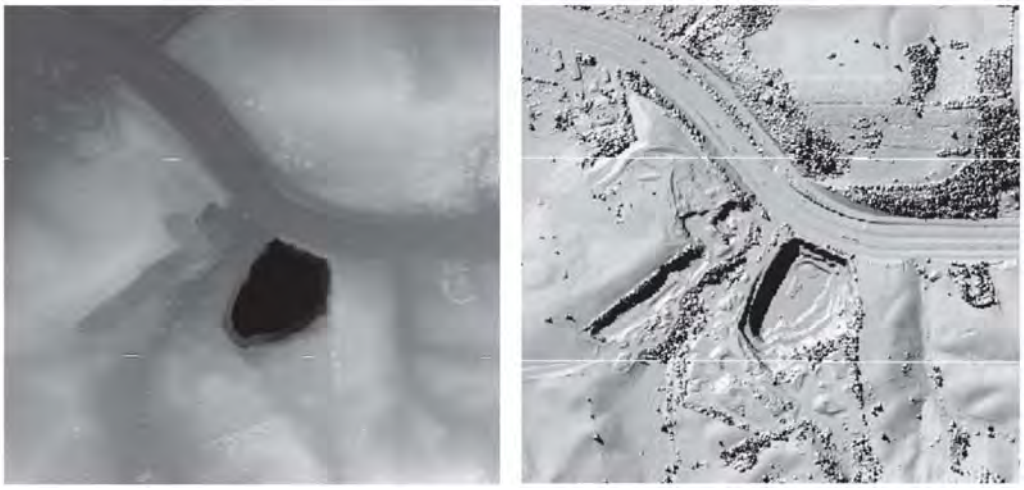
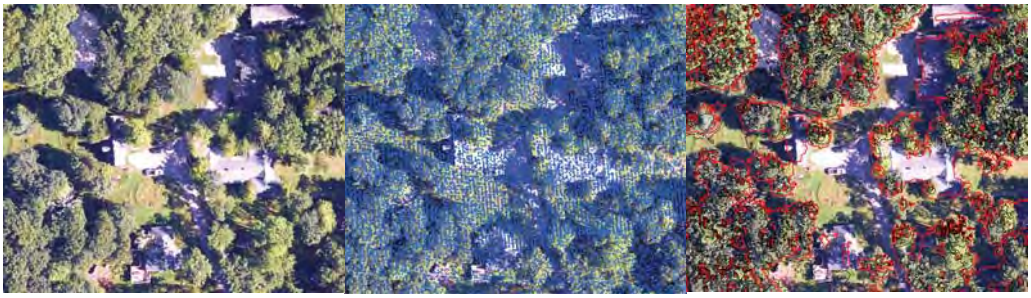


FIGURE 9.12 Portions of the lidar-derived DEM and hillshade model from **Figure 9.11** enlarged to depict detail. Left: The digital elevation model, with darker areas corresponding to lower elevation. Right: The hillshade model (or shaded relief) of the same area with shadow effects. Note the very fine spatial detail of the terrain, including individual trees and microtopography. From Virginia Department of Transportation. Copyright © 2003 Commonwealth of Virginia. Used by permission.



(a) (b) (c)

FIGURE 9.13 (a) High-resolution orthorectified camera imagery from 2015 of buildings and trees at the Mountain Lake Biological Station in Virginia. (b) Lidar last returns from 2015 (blue) displayed over the same imagery. Note the high density of points, which provides significant detail regarding three-dimensional structure. (c) 5-m contours derived from the last return points. Users can generate contours at intervals of their choice. Provisional data from National Ecological Observatory Network (2021a, 2021b).

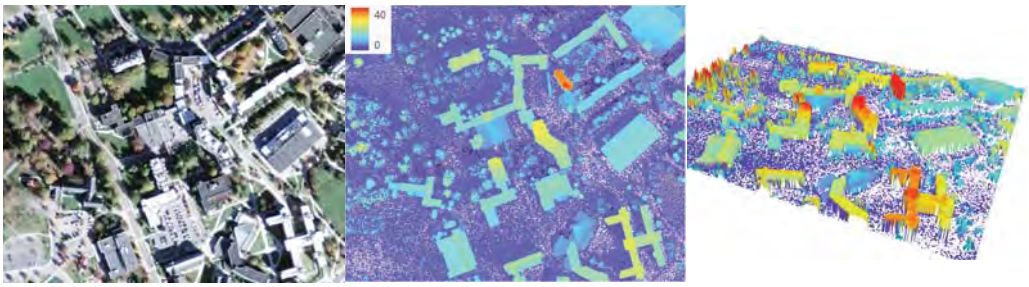


FIGURE 9.14 Left: Portion of a 2014 image from the U.S. Department of Agriculture National Agriculture Imagery Program (NAIP) over buildings on the Virginia Tech campus in Blacksburg, Virginia. Middle: A model of heights above ground (m) derived from lidar data. This model was created by subtracting a lidar-derived DSM from a lidar-derived DEM. The effects of terrain are removed, enabling the determination of feature heights, which include trees and buildings. Right: A three-dimensional viewing perspective of the same height model. From NAIP (left image). Middle and right image derived from 2017 lidar data provided by the USGS.

referred to as digital surface models (DSMs). Both DEMs and DSMs are raster products derived from lidar (similar to images) and are elevations relative to the Earth's ellipsoid. If they are subtracted, a third raster product will be generated that contains the heights above the ground.

Waveform Data

Waveform data provide significantly more information about the interaction of the lidar pulse with features on the surface. Although many systems collect waveform data, as a practical matter the applications for the waveform itself are relatively few other than for research. Instead, scientists have developed algorithms to process the raw lidar waveform and extract relative heights. Figure 9.15 shows a waveform of the boreal forest in northern Ontario, Canada, collected by the Global Ecosystem Dynamics Investigation (GEDI) profiling lidar that is currently operating from the International Space Station (ISS). The large spike in the waveform, near the bottom orange dashed line, represents the interaction of the waveform with the ground. The other nodes between the dashed lines characterize the vertical structure of the canopy (i.e., a representation of the trunks, branches, and foliage). Readers should keep in mind that this level of detail is available for all 470 samples that make up the graph in Figure 9.16. This information is valuable for scientists, particularly those who are interested in forest vertical structure and how it is changing.

Algorithms have been developed to identify heights that indicate a more significant return from a feature, referred to as relative heights (RH), where RH100 would represent the height of 100% aboveground energy return—in other words, the height of the top of a feature such as a forest canopy. RH0 would represent the ground, and RH25, RH50, and RH75 would represent some intermediate height, conceptually similar to multiple returns in the discrete systems described above. The algorithm used to process the waveform in Figure 9.15 identified the following relative heights, which are seen at sample number 442 in Figure 9.16: RH100 = 13.7 m (top orange dashed line in Figure 9.15);

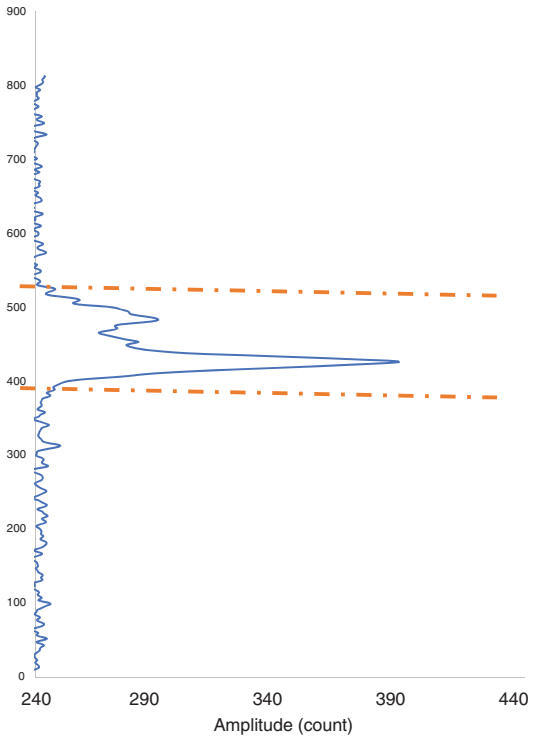


FIGURE 9.15 GEDI Level 1B RX waveform for sample 442 in **Figure 9.16**. The two dashed orange lines represent the approximate portion of the waveform that interacted with the boreal forest. The top orange dashed line represents the top of the canopy, at approximately 13.7 m above the ground. The largest spike in the waveform is caused by the ground, which causes the largest amount of energy return to the sensor. The signal above and below the dashed lines are considered background noise and are filtered out by the data processing algorithm.

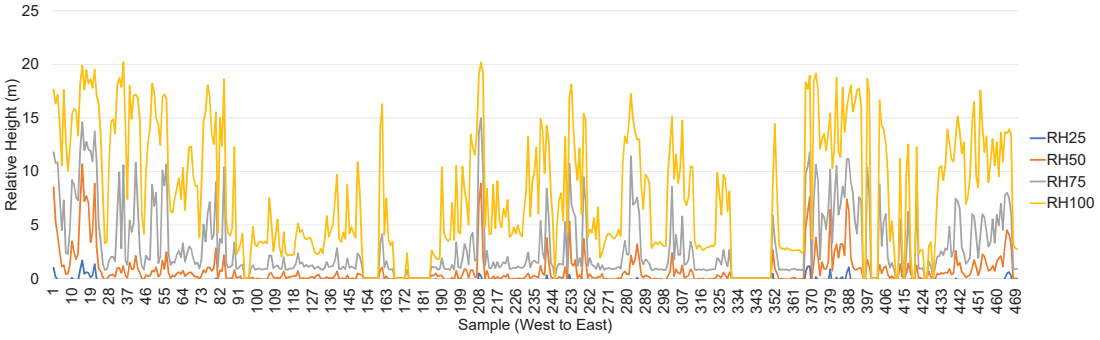


FIGURE 9.16 Example of relative heights extracted from waveform lidar collected by the GEDI system from the International Space Station (ISS) on August 7, 2019. The orbital path of the ISS passed over the boreal forest in northern Ontario, Canada. The graph shows the relative heights along a short transect, where each sample location on the graph can be linked to a full waveform, such as that in **Figure 9.15**. Available for public download in the GEDI level 2b data.

RH75 = 6.5 m; RH25 = 1.46 m. Users can obtain either the waveforms themselves or the processed relative heights, which are much easier to explore. In this example, RH25, RH50, RH75, and RH100 are shown along the profile, covering 470 individual waveform samples along the ISS ground track over northern Ontario. You can see that the top of the forest canopy is at or below 20 m (typical at this latitude), and there are areas along the transect that are relatively open, with much lower heights.

9.6 SELECTED LIDAR APPLICATIONS

From the beginning, lidar systems have been successfully applied to address an abundance of useful applications—essentially any application that benefits from height information, including detailed information about ground surface. Lidars have been effective in their ability to support varied functions under a range of conditions, including unfavorable weather, day/night conditions, and rough terrain. Common uses of lidar data include surveying, mapping of anthropogenic features, particularly urban, transportation, piping, and communication infrastructure, floodplain and flood risk mapping, archaeology, forestry (Chapter 18), coastal resilience (Chapter 20), agriculture, and bathymetry. There are many more uses of lidar data—too many to describe here. Below we expand upon four common terrain applications of lidar.

Forestry. Lidar data is used extensively in forestry, as it provides a method for characterizing tree size and shape, as well as the ground. This issue is discussed further in Chapter 18. Lidar is used to estimate canopy height, aboveground biomass, leaf area, and other forest inventory variables. Among other uses, this is important for forest industry, carbon monitoring, understanding biodiversity and wildlife habitat, mapping forested wetlands, predicting fire risk, and understanding hydrological processes in forests.

Agriculture. Monitoring agricultural lands is a critical component for a farmer's maintenance of agricultural productivity. Lidar observations provide an effective and inexpensive resource for monitoring crops during the growing season, supporting farmers' ability to closely monitor their crops, their status, and their yield. Lidar observations can also provide the basis for preparing topographic maps of fields in order to detect slopes, ponding, and erosion. Other types of remote sensing used for agriculture are described in more detail in Chapter 17.

Geology. Geoscientists have applied lidar imagery, often with GPS, to detect and interpret structural geology and geophysics to detect and study faults, as well as to measure uplift and other changes in terrain.

Archaeology. Lidar has made significant contributions to the detection and mapping of archaeological remains by revealing structures beneath forest canopies as well as the subtle outlines of building foundations, which are often buried but cause slight elevation changes detectable in the microtopography of a lidar-derived DEM. There have been dramatic archaeology finds in the rainforest, which is often too dense to penetrate with other forms of remote sensing. An example is the Mayan archaeology of Guatemala and Belize, where lidar revealed a much larger extent of civilization than had been previously known (e.g., Chase et al., 2014; Canuto et al., 2018).

9.7 LIDAR DATA FORMATS

As lidar technology has evolved and the data have become more available and heavily used, the way the data are stored has become more standardized. In early years, the data were often either in ASCII text format or in a proprietary binary format from the vendor. However, the ASCII format was impractical and computationally inefficient, and the proprietary formats were difficult to access. Over time, the binary LAS (LASer) file format became an industry standard, and it is now common to store the point and waveform in this format. The standards have evolved over time, as technology and data improved, and are specified by the American Society for Photogrammetry and Remote Sensing (ASPRS). Through each evolution (currently, version ASPRS LAS 1.4), consideration is given to legacy compatibility such that earlier lidar data can still be accessed. The format is designed to contain the data and the associated attributes. There is also the ability to store a reference class for each point, which can be assigned in postprocessing after the data are collected (ground, water, vegetation, etc.). The idea is that binary data are stored in a computationally efficient yet open and common format, around which software and analysis tools can be designed while still maintaining all-important attribute data for each point or waveform.

9.8 SUMMARY

Lidar provides a highly accurate, detailed representation of terrain and permits the separation of features (such as vegetation) from the terrain, a capability unique among remote sensing instruments. Its status as an active sensor permits convenience and flexibility in flight planning due to its insensitivity to variations in weather and solar illumination—both important constraints on aerial photography.

Lidar data provide detailed, spatial data of high accuracy and precision. Lidar gives a direct measurement of surface elevation, with detail and accuracy usually associated only with photogrammetric surveys. Some lidar applications replace photogrammetric applications of aerial photography. Many applications have focused on urban regions, which experience continuing needs for detailed information concerning building densities, urban structures, and building footprints. Lidar data are used for highway planning, pipeline routing, and design of wireless communication systems in urban regions. Although wall-to-wall lidar coverage of the United States does not exist, several states have completed or are planning to acquire statewide lidar coverage to support floodplain mapping programs and other efforts with broader geographic reach. Lidar data have also been used to study forest structure, as the detailed and accurate information describing canopy configuration and structure may permit accurate mapping of timber volume. There are now a broad range of environmental applications, in which lidar's detailed representations of terrain have opened avenues of inquiry that are not practical with coarser data. As lidar archives acquire increasing geographic scope and temporal depth, the field will be able to expand its reach to examine sequential changes in vegetation cover, land use, and geomorphology with a precision and accuracy not previously feasible.

SOME TEACHING AND LEARNING RESOURCES

Introductory Resources

- Introduction to Light Detection and Ranging (LiDAR)—Explore Point Clouds and Work with LiDAR Raster Data in R; National Ecological Observatory Network
  www.neonscience.org/intro-lidar-r-series
- Lidar: Light Detection and Ranging
  www.youtube.com/watch?v=hxiRkTtBQp8&fmt=22
- Parece, Tammy E., John A. McGee, and James B. Campbell. 2016. *Working with Lidar Using ArcGIS Desktop*. Blacksburg, VA: VirginiaView, 333 pp.
- 3D Elevation Program—United States Geological Survey (for publicly available lidar data)
  www.usgs.gov/core-science-systems/ngp/3dep

Examples of Lidar Data, Applications, and Acquisitions

- Lidar Surface Shadow Model for Boston's Back Bay
  www.youtube.com/watch?v=s4OhzaIXMhg&NR=1
- Pylon Lidar Survey
  www.youtube.com/watch?v=Dv6a0KgTbiw
- Terrapoint Aerial Services—Lidar Flight Simulation
  www.youtube.com/watch?v=GSPcyhSAgTQ&NR=1
- LandXplorer: Lidar Scan of London
  www.youtube.com/watch?v=F2xy-US46PQ&NR=1
- Lidar Survey
  www.youtube.com/watch?v=f1P42oQHN_M&feature=related
- eCognition Image Analysis: Extracting Tree Canopy from Lidar
  www.youtube.com/watch?v=OR1Se18Zd4E

REVIEW QUESTIONS

1. Review some of the strengths of lidar data relative to other forms of remotely sensed data discussed thus far.
2. Identify some reasons that lidar might be effectively used in combination with other data. What difficulties might be encountered in bringing together lidar data and, for example, fine-resolution optical satellite imagery?
3. Many observers believe that the increasing availability of lidar will displace the current role of aerial photography for many applications. What are some of the reasons that might lead people to believe that lidar could replace many of the remote sensing tasks now performed by aerial photography?
4. Can you identify reasons that aerial photography might yet retain a role, even in competition with the strengths of lidar data?

5. If aerial photography is largely replaced by lidar, do you believe that there will still be a role for teaching aerial photography as a topic in a university remote sensing course? Explain.
6. Assume for the moment that lidar data become much cheaper, easier to use, and, in general, more widely available to remote sensing practitioners. What kinds of new remote sensing analyses might become possible, or what existing analyses might become more widespread?
7. The text discusses how lidar data is based on the convergence of several technologies. Review your notes to list these technologies. Think about the technologic, scientific, social, and economic contexts that foster the merging of these separate capabilities. How do you think we can prepare to encourage future convergences of other technologies (now unknown) that might lead to advances in remote sensing instruments?
8. Lidar imagery may not be equally useful in all regions of the Earth. Can you suggest certain geographic regions or environments in which lidar data might not be effective?
9. Discuss some of the considerations that might be significant in deciding the season in which to acquire lidar data for your region.
10. Identify some of the special considerations that might be significant in planning acquisition of lidar data in urban regions.

REFERENCES

Blair, J. B., D. L. Rabine, and M. A. Hofton. 1999. The Laser Vegetation Imaging Sensor: A Medium-Altitude, Digitization-Only, Airborne Laser Altimeter for Mapping Vegetation and Topography. *ISPRS Journal of Photogrammetry and Remote Sensing*, Vol. 54, pp. 115–122.

Canuto, M. A., F. Estrada-Belli, T. G. Garrison, S. Houston, M. J. Acuña, M. Kováč, and D. Chatelain. (2018). Ancient Lowland Maya Complexity as Revealed by Airborne Laser Scanning of Northern Guatemala. *Science*, Vol. 361, No. 6409, p. eaau0137.

Chase, A. F., D. Z. Chase, J. J. Awe, J. F. Weishampel, G. Iannone, H. Moyes, . . . and M. K. Brown. (2014). The Use of LiDAR in Usnderstanding the Ancient Maya Landscape: Caracol and Western Belize. *Advances in Archaeological Practice*, Vol. 2, No. 3, pp. 208–221.

DeLoach, S. R., and J. Leonard. 2000. Making Photogrammetric History. *Professional Surveyor*, Vol. 20, No. 4, pp. 6–11.

Flood, M. 2001. Laser Altimetry: From Science to Commercial LiDAR Mapping. *Photogrammetric Engineering and Remote Sensing*, Vol. 67, pp. 1209–1217.

Flood, M. 2002. Product Definitions and Guidelines for Use in Specifying LiDAR Deliverables. *Photogrammetric Engineering and Remote Sensing*, Vol. 68, No. 12, pp. 1230–1234.

Habib, A. F. 2010. Airborne LIDAR Mapping. Chapter 23 in *Manual of Geospatial Science and Technology* (2nd ed.). (J. D. Bossler, ed.). Boca Raton, FL: CRC Press, pp. 439–465.

Hill, J. M., L. A. Graham, and R. J. Henry. 2000. Wide-Area Topographic Mapping Using Airborne Light Detection and Ranging (LIDAR) Technology. *Photogrammetric Engineering and Remote Sensing*, Vol. 66, pp. 908–914, 927, 960.

Hyyppä, J., W. Wagner, M. Hollaus, and H. Hyyppä. 2009. Airborne Laser Mapping. Chapter 14 in the *Sage Handbook of Remote Sensing* (T. A. Warner, M. D. Nellis, and G. M. Foody, eds.). London: SAGE, pp. 199–211.

Lefsky, M. A., D. Harding, W. B. Cohen, G. Parker, and H. H. Shugart. 1999. Surface Lidar Remote Sensing of Basal Area and Biomass in Deciduous Forests of Eastern Maryland, USA. *Remote Sensing of Environment*, Vol. 67, pp. 83–98.

National Ecological Observatory Network.2021a, July 28. Data Product DP1.30003.001, Discrete Return Lidar Point Cloud. Provisional data retrieved from *https://data.neonscience.org*.

National Ecological Observatory Network. 2021b, July 28. Data Product DP1.30010.001, High-Resolution Orthorectified Camera Imagery and Data Product DP1.30003.001, Discrete Return Lidar Point Cloud. Provisional data retrieved from *https://data.neonscience.org*.

Nelson, R., W. Krabill, and J. Tonelli. 1988a. Estimating Forest Biomass and Volume Using Airborne Laser Data. *Remote Sensing of Environment,* Vol. 24, pp. 247–267.

Nelson, R., R. Swift, and W. Krabill, W. 1988. Using Airborne Lasers to Estimate Forest Canopy and Stand Characteristics. *Journal of Forestry,* Vol. 86, pp. 31–38.

Nilsson, M. 1996. Estimation of Tree Heights and Stand Volume Using an Airborne Lidar System. *Remote Sensing of Environment,* Vol. 56, pp. 1–7.

Parece, T. E., J. A. McGee, and J. B. Campbell. 2016. *Working with Lidar Using ArcGIS Desktop*. Blacksburg, VA: VirginiaView, 333 pp.

Popescu, S. C. 2002. *Estimating Plot-Level Forest Biophysical Parameters Using Small-Footprint Airborne Lidar Measurements*. PhD Dissertation, Virginia Tech, Blacksburg, 144 pp.

Romano, M. E. 2004. Innovation in Lidar Processing. *Photogrammetric Engineering and Remote Sensing*, Vol. 70, pp. 1201–1206.

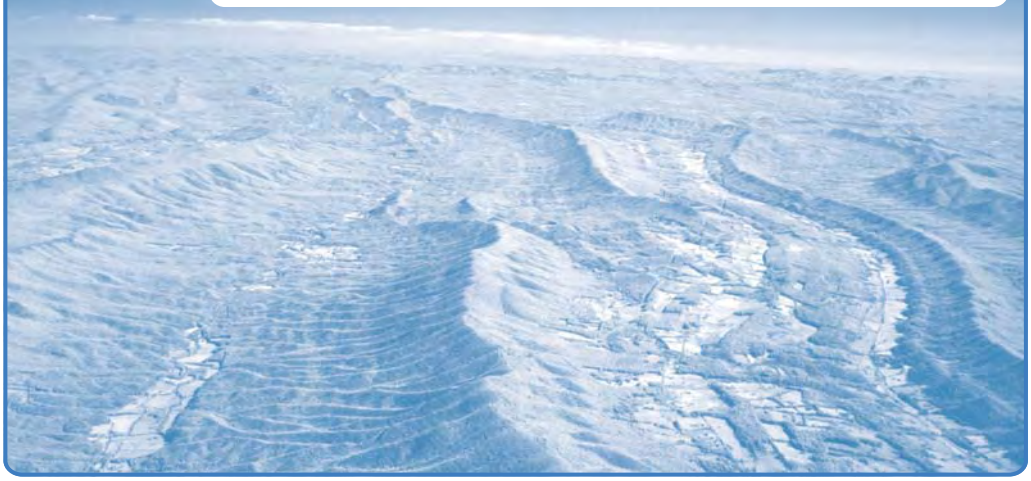
Sapeta, K. 2000. Have You Seen the Light? LIDAR Technology Is Creating Believers. *GEOWorld,* Vol. 13, No. 10, pp. 32–36.

Wehr, A., and U. Lohr (eds.). 1999. Airborne Laser Scanning: An Introduction and Overview. *ISPRS Journal of Photogrammetry and Remote Sensing,* Vol. 54, Nos. 2–3, pp. 68–82.

White, S. A., and Y. Wang. 2003. Utilizing DEMs Derived from LIDAR Data to Analyze Morphologic Change in the North Carolina Coastline. *Remote Sensing of Environment,* Vol. 85, pp. 39–47.

Zhang, K., and D. Whitman. 2005. Comparison of Three Algorithms for Filtering Airborne Lidar Data. *Photogrammetric Engineering and Remote Sensing,* Vol. 71, pp. 313–324.

10 Thermal Imagery



MAJOR TOPICS TO UNDERSTAND

- Thermal Detectors
- Thermal Radiometry
- Microwave Radiometers
- Infrared Thermography
- Thermal Properties of Objects
- Land Surface Temperature
- Geometry of Thermal Images
- The Thermal Image and Its Interpretation

10.1 INTRODUCTION

This chapter discusses the uses of radiation from about 7 to 18 μm for remote sensing of landscapes, using *passive remote sensing*, as defined previously in Chapter 2. The infrared spectrum was discovered in 1800 by Sir William Herschel (1738–1822), a British

astronomer who recognized that surfaces illuminated with wavelengths beyond the red region produced temperatures higher than those observed within the visible spectrum. Later, he found that these longer wavelengths could be reflected, refracted, absorbed, and transmitted in a manner similar to that of visible radiation. His research not only advanced the field of astronomy, but has also opened new dimensions for fields such as meteorology, medicine, fire suppression, and geosciences.

The infrared region of the spectrum begins at the longwave edge of the visible region (at about 0.72 μm), extending to the shortwave edge of the microwave region (at about 1 mm). The infrared region is notable because of the wide range of wavelengths, its diverse range of applications for remote sensing, and the range of sensors.

Six subdivisions of the infrared spectrum are recognized:

- The *near-infrared (NIR)* region (about 0.72–2.5 μm), already introduced in Chapter 2, is noted for its ability to (1) assess the health and vigor of living plant tissue; (2) avoid the effects of atmospheric haze and scattering; and (3) sharply define the boundary between land and open water. In addition, NIR radiation is used with other bands to define several band ratios to detect vegetation health, as discussed in later chapters.
- The *reflective infrared region* (0.72–3.0 μm) behaves much like visible light in the sense that it is reflected from, or absorbed by, objects. Many of the same kinds of films, filters, lenses, and cameras that we use in the visible portion of the spectrum can also be used, with modest variations, for imaging in the shortwave infrared region.
- Within the reflective infrared region, there is sometimes a distinction between the *photographic infrared* (0.72–1.2 μm) and the broader region extending to 3 μm. The photographic infrared region identifies wavelengths that can be captured with photographic films.
- The *shortwave infrared (SWIR)* region (1.40–3.0 μm) behaves much as visible light does, in the sense that it is reflected, which can allow for relatively fine spatial resolution. The SWIR region can penetrate atmospheric smoke and haze, so it is effective in monitoring forest and range fires, and in defining regions of active burning, detecting hot spots, and estimating severity of burned regions. SWIR is significant for its role in detecting the presence of minerals exposed at the soil surface, so it is especially useful in searching for minerals.
- The *mid-infrared (mid-IR)* region (3–8 μm), also known as the "thermal infrared" or "emissive infrared" region, does not reflect radiation that forms imagery, but rather generates and emits it.
- The *thermal infrared (TIR)* region (3–14 μm) also includes longer IR wavelengths *(far-IR)* and transmits heat generated by processes such as geothermal warming, forest and range fires, and solar warming of impervious surfaces.

In a remote sensing context, the infrared region's spectrum presents challenges because it encompasses such a broad range of wavelengths, with varied subdivisions, applications, and sensing technologies. Landsat and Sentinel sensors have bands specifically targeted within the SWIR and TIR (for Landsat) regions to capture wavelengths that will provide useful information at wavelengths that can minimize atmospheric inter-

ference. In Figure 10.1, "SWIR" identifies regions for two common SWIR bands. Table 10.1 provides more specific definitions of the Landsat SWIR and thermal bands.

Remote sensing in the mid- and far-IR is based on a family of imaging devices that differ greatly from the cameras and films used in the visible and near-infrared regions. The interactions of the mid- and far-IR regions with the atmosphere are also quite different than those of shorter wavelengths. The far-IR regions are free from the scattering that is so important in the ultraviolet and visible regions, but absorption by atmospheric gases restricts uses of the mid- and far-IR spectrum to specific atmospheric windows. Also, information acquired by sensing the far-IR spectrum differs from that acquired in the visible and near-infrared regions. Variations in emitted energy in the far-infrared region provide information concerning surface temperature and thermal properties of soils, rocks, vegetation, and anthropogenic structures.

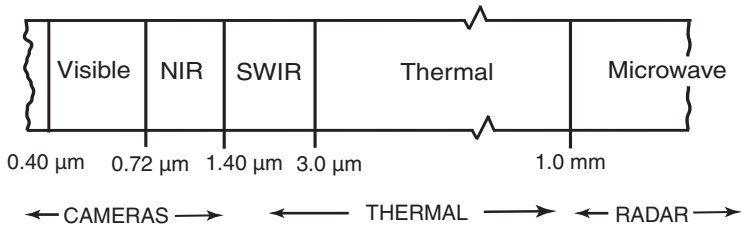


FIGURE 10.1 Infrared and thermal spectrum. This schematic view of the infrared spectrum identifies principal designations within the infrared region. Labeled regions are not always contiguous because some portions of the infrared spectrum are unavailable for remote sensing because of atmospheric effects. Names and spectral limits vary within different disciplines, so details may not agree with other references. (This diagram is not drawn to scale.) Image by Susmita Sen.

TABLE 10.1 Landsat SWIR and TIRS Summary

	Band	Wavelength (µm)	Principal application
Landsat 4, 5, & 7 (TM, ETM+)	SWIR1, b5	1.55–1.75	Soil and vegetation moisture; penetrates thin clouds
	SWIR2, b7	2.09–2.35	Hydrothermally altered rocks associated with mineral deposits
	TIR	10.40–12.5	Thermal mapping, soil moisture
Landsat 8 OLI	SWIR1, b6	1.566–1.651	Soil and vegetation moisture; penetrates thin clouds
	SWIR2, b7	2.107–2.294	
Landsat 8 TIRS (100 m)	TIRS1, b10	10.60–11.19	Thermal mapping, soil moisture
	TIRS2, b11	11.50–12.51	

Note: ETM+ = Enhanced Thematic Mapper Plus; OLI = Operational Land Imager; TIRS = thermal infrared sensor; TM = Thematic Mapper. Data from Barsi et al. (2014).

10.2 THERMAL DETECTORS

Before the 1940s, the absence of suitable instruments limited the use of thermal infrared radiation for aerial reconnaissance. Aerial mapping of thermal energy depends on use of a sensor that is sufficiently sensitive to thermal radiation such that variations in apparent temperature can be detected by an aircraft moving at considerable speed high above the ground. Early instruments for thermographic measurements examined differences in electrical resistance caused by changes in temperature. But such instruments could function only when in close proximity to the objects of interest. Although such instruments could be useful in an industrial or laboratory setting, they are not sufficiently sensitive for use in the context of remote sensing; they respond slowly to changes in temperature and cannot be used at distances required for remote sensing applications.

During the late 1800s and early 1900s, the development of *photon detectors* (sometimes called *thermal photon detectors*) provided a practical technology for use of the thermal portion of the spectrum in remote sensing. Such detectors are capable of responding directly to incident photons by reacting to changes in electrical resistance, providing a sensitivity and speed of response suitable for use in reconnaissance instruments. By the 1940s, a family of photon detectors had been developed to provide the basis for electro-optical instruments used in several portions of the TIR spectrum.

Detectors are devices formed from substances known to respond to energy over a defined wavelength interval, generating a weak electrical signal with a strength related to radiances of features in the field of view of the sensor. Often, sensitivity of such materials increases to practical levels when they are cooled to very low temperatures to increase sensitivity and reduce noise. The electrical current is amplified, then used to generate a digital signal that can be employed to form a pictorial image, roughly similar in overall form to an aerial photograph.

Detectors have been designed with sensitivities for many of the spectral intervals of interest in remote sensing, including regions of the visible, near-infrared, and ultraviolet spectra. Detectors sensitive in the thermal portion of the spectrum are formed from rather exotic materials, such as *indium antimonide* (InSb) and mercury-doped germanium (Ge:Hg). InSb has a peak sensitivity near 5 μm in the mid-infrared spectrum, and Ge:Hg has a peak sensitivity near 10 μm in the far-infrared spectrum (Figure 10.2). Mercury cadmium telluride (MCT) is sensitive over the range 8–14 μm. To maintain maximum sensitivity, such detectors must be cooled to very low temperatures (–196°C or –243°C) using liquid nitrogen or liquid helium.

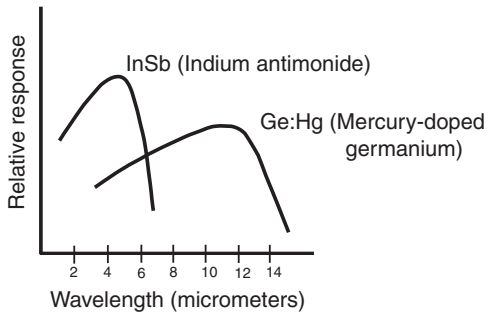


FIGURE 10.2 Sensitivity of some common thermal detectors. Image by Susmita Sen.

The sensitivity of the detector is a significant variable in the design and operation of the system. Low sensitivity means that only large differences in brightness are recorded ("coarse radiometric resolution") and most of the finer detail in the scene is lost. High sensitivity means that finer differences in scene brightness are recorded ("fine radiometric resolution"). The *signal-to-noise ratio* (SNR or S/N ratio) expresses this concept (Chapter 5). The "signal" in this context refers to differences in image brightness caused by actual variations in scene brightness. "Noise" designates variations unrelated to scene brightness. Such variations may be the result of unpredictable variations in the performance of the system. (There may also be random elements contributed by the landscape and the atmosphere, but here "noise" refers specifically to that contributed by the sensor.) If noise is large relative to the signal, the image does not provide a reliable representation of the feature of interest. Clearly, high noise levels will prevent imaging of subtle features. Even if noise levels are low, there must be minimum contrast between a feature and its background (i.e., a minimum magnitude for the signal) for the feature to be imaged. Also, note that increasing fineness of spatial resolution decreases the energy incident on the detector, with the effect of decreasing the strength of the signal. For many detectors, noise levels may remain constant even though the level of incident radiation decreases; if so, the increase in spatial resolution may be accompanied by decreases in radiometric resolution, as suggested in Chapter 5.

10.3 THERMAL RADIOMETRY

A *radiometer* is a sensor that measures the intensity of radiation received within a specified wavelength interval and within a specific field of view, as shown by the schematic view in Figure 10.3. A lens, or mirror, collects radiation from the ground and then focuses it on a detector positioned in the focal plane. A *field stop* may restrict the field of view, and a filter may be used to restrict the wavelength interval that reaches the detector. The *dichroic mirror* in Figure 10.3 serves this purpose by separating specific wavelengths.

A characteristic feature of radiometers is that radiation received from the ground is compared with a *reference source* of known radiometric qualities. This process can be implemented through the design sketched in Figure 10.3. The slotted rotating mirror depicted in Figure 10.3 (sometimes known as a *chopper*) is capable of interrupting radiation reaching the detector. The chopper consists of a slotted disk, or similar device, rotated by an electrical motor so that as the disk rotates, it causes the detector to alternately view the target and the reference source of radiation. Because the chopper rotates very fast, the signal from the detector consists of a stream of data that alternately measures the radiance of the reference source, then radiation from the ground. This signal can be used to determine the radiance difference between the reference and the target. Because the reference source has known radiance, radiance of the target can then be estimated.

Although there are many variations on this design, this description outlines the most important components of radiometers. Related instruments include *photometers,* which operate at shorter wavelengths and often lack the internal reference source, and *spectrometers,* which examine radiance over a range of wavelengths. Radiometers can be designed to operate at different wavelength intervals, including portions of the infrared and ultraviolet spectra. By carefully tailoring the sensitivity of radiometers, scientists

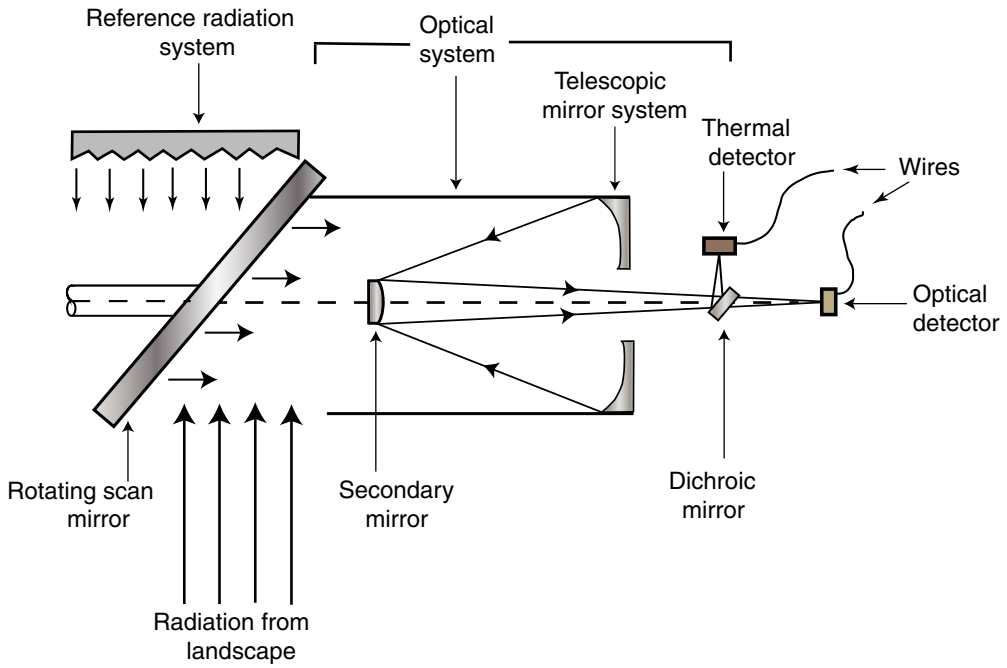


FIGURE 10.3 Schematic sketch of an optical–mechanical thermal scanner. The instrument collects radiation from the Earth's surface. As the rotating mirror (shown here in cross section) rotates, it alternatively projects radiation from the Earth's surface and from the reference source. The lens system directs radiation to the dichroic mirror (which splits incoming light by wavelength) to provide both thermal and optical data. Image by Susmita Sen.

have been able to design instruments that are very useful in studying atmospheric gases and cloud temperatures. Radiometers used for Earth resources studies are often configured to view only a single trace along the flight path; the output signal then consists of a single stream of data that varies in response to differences in radiances of features along the flight line. A *scanning radiometer* can gather data from a corridor beneath the aircraft; output from such a system resembles that from some of the scanning sensors discussed in earlier chapters.

Instantaneous Field of View

Spatial resolution of a radiometer is determined by an *instantaneous field of view* (IFOV), which is in turn controlled by the sensor's optical system, the detector, and flying altitude. Radiometers often have relatively coarse spatial resolution. For example, satellite-borne radiometers may have spatial resolutions of 60–100 km or more—in part because of the desirability of maintaining high radiometric resolution. To ensure that the sensor receives enough energy to make reliable measurements of radiance, the IFOV is defined to be rather large; a smaller IFOV would mean that less energy would reach the detector, the signal would be much too small with respect to system noise, and the measure of radiance would be much less reliable.

The IFOV can be informally defined as the area viewed by the sensor if the motion of the instrument were to be suspended so that it records radiation from only a single patch of ground. The IFOV can be more formally expressed as the angular field of view (β) of the optical system (Figure 10.4). The projection of this field of view onto the ground surface defines the circular area that contributes radiance to the sensor. Usually, for a particular sensor it is expressed in radians (r); to determine the IFOV for a particular image, it is necessary to know the flying altitude (H) and to calculate the size of the circular area viewed by the detector.

From elementary trigonometry it can be seen that the diameter of this area (D) is given as

$$D = H\beta \qquad \text{(EQ. 10.1)}$$

as illustrated in Figure 10.4. Thus, for example, if the angular field of view is 1.0 milliradians (mr) (1 mr = 0.001 r) and the flying altitude (H) is 400 m above the terrain, then:

$$D = H\beta$$
$$D = 400 \times 1.0 \times 0.001$$
$$D = 0.40 \text{ m}$$

Because a thermal scanner views a landscape over a range of angles as it scans from side to side, the IFOV varies in size depending on the angle of observation (θ). Near the nadir (ground track of the aircraft), the IFOV is relatively small; near the edge of the image, the

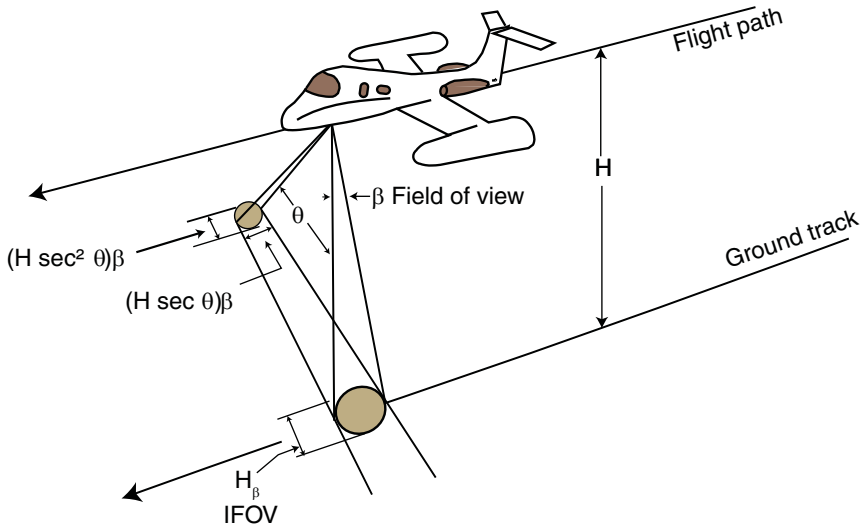


FIGURE 10.4 Sketch of the geometry of an aerial thermal scanner. The circular shape beneath the aircraft depicts the sensor's instantaneous field of view (IFOV) (representing the size and shape of the area viewed by the sensor as it views the nadir). At nadir, the diameter of the IFOV is determined by the flying altitude and the instrument's field of view (H_β). As the instrument scans side to side, it observes at an angle (θ) in off-nadir positions along the scan path. Image by Susmita Sen.

IFOV is large. This effect is beneficial in one sense because it compensates for the effects of the increased distance from the sensor to the landscape, thereby providing consistent radiometric sensitivity across the image.

Other effects are more troublesome. Equation 10.1 defines the IFOV at nadir as $H\beta$. Most thermal scanners scan side to side at a constant angular velocity, which means that, in a given interval of time, they scan a larger cross-track distance near the sides of the image than at the nadir. At angle θ, the IFOV measures $H \sec \theta \beta$ in the direction of flight and $H \sec^2 \theta \beta$ along the scan axis. Thus, near the nadir, the IFOV is small and symmetrical; near the edge of the image, it is larger and elongated in the direction of flight. The variation in the shape of the IFOV creates geometric errors in the representations of features—a problem discussed in subsequent sections. The variation in size means that radiance from the scene is averaged over a larger area and can be influenced by the presence of small features of contrasting temperature. Although small features can influence data for IFOVs of any size, the impact is more severe when the IFOV is large.

10.4 MICROWAVE RADIOMETERS

Microwave emissions from the Earth convey some of the same information carried by thermal (far-IR) radiation. Even though their wavelengths are much longer than those of thermal radiation, microwave emissions are related to temperature and emissivity in much the same manner as is thermal radiation. *Microwave radiometers* are sensitive instruments tailored to receive and record radiation in the range from about 0.1 mm to 3 cm. Whereas imaging radars discussed in Chapter 8 are active sensors that illuminate the terrain with their own energy, microwave radiometers are passive sensors that receive microwave radiation naturally emitted by the environment. The strength and wavelength of such radiation are largely functions of the temperature and emissivity of the target. Thus, although microwave radiometers, like radars, use the microwave region of the spectrum, they are functionally most closely related to the thermal sensors discussed in this chapter.

In the present context, we are concerned with microwave emissions from Earth, which indirectly provide information pertaining to vegetation cover, soil moisture status, and surface materials. Other kinds of studies, peripheral to the field of remote sensing, derive information from microwave emissions from the Earth's atmosphere or from extraterrestrial objects. In fact, the field of microwave radiometry originated with radio astronomy, and some of its most dramatic achievements have been in the reconnaissance of extraterrestrial objects.

A *microwave radiometer* consists of a sensitive receiving instrument typically in the form of a horn- or dish-shaped antenna that observes a path directly beneath an aircraft or satellite. The signal gathered by the antenna is electronically filtered and amplified and then displayed as a stream of digital data, or, in the instance of scanning radiometers, as an image. As with thermal radiometers, microwave radiometers have a reference signal from an object of known temperature. The received signal is compared with the reference signal as a means of deriving the radiance of the target.

Examination of data from a microwave radiometer can be complex due to the many factors that contribute to a given observation. The component of primary interest is usually energy radiated by the features within the IFOV; of course, variations within the IFOV are lost, as the sensor can detect only the average radiance within this area. The

atmosphere also radiates energy, so it contributes radiance, depending on moisture content and temperatures. In addition, solar radiation in the microwave radiation can be reflected from the surface to the antenna.

10.5 INFRARED THERMOGRAPHY

Infrared thermography is the science of recording imagery at the longer wavelengths of the infrared spectrum (about 9–14 μm). Thermal scanners collect overhead lateral scans that can be assembled to form composite vertical images (Section 10.3). In contrast, thermographic cameras collect snapshot-like images (often with an oblique perspective) in the manner of the framing cameras discussed in Chapter 4.

A principal design difference between optical cameras and thermal cameras is composition of the lens. Glass lenses designed for optical cameras will reflect thermal radiation rather than allow radiation to pass through the lens to form an image. Therefore, thermal cameras require designs tailored to accommodate thermal radiation—specifically, to use lenses manufactured of materials such as germanium (Ge), chalcogenide glass, zinc selenide (ZnSe), and zinc sulfide (ZnS). Such materials are expensive, especially in the case of germanium, and are fragile, so protective coatings are required. The high costs of these special lenses form the principal source of the expense of thermographic cameras relative to that of more familiar optical cameras.

Although a broad range of thermographic cameras is in use, a specific design is widely used for aerial thermal imagery. *Forward-looking infrared (FLIR)* cameras form an important means for collecting aerial thermography. FLIR cameras employed by both military and civilian aircraft, as well as other thermal imaging cameras, detect infrared radiation typically emitted from a heat source. The "forward-looking" designation distinguishes between fixed, forward-looking, thermal imaging systems and the lateral-scanning infrared systems used for the aerial scanners and satellite systems previously discussed in this chapter. The FLIR camera is typically mounted to view forward to acquire an oblique view along the flight path. Since systems using lateral scans are not practical for real-time imaging, FLIR technologies expand the range of capabilities for thermal imaging. They also provide a real-time capability that is especially important for military, law enforcement, and search and rescue applications. Other important applications include heat loss surveys, monitoring forest and range fires, and security monitoring (Figure 10.5). FLIR cameras facilitate aerial navigation at night and in fog, haze, and total darkness, through light fog, light rain, and snow. The distance they can see is affected by these atmospheric conditions.

In many applications discussed here, it is not feasible to calibrate instruments to measure the temperatures of surfaces. The purpose of many thermal surveys is not so much to precisely measure temperatures, but to gauge relative temperatures to identify anomalies or abnormalities to highlight features that require further examination using other technologies. FLIR images can be displayed in a black-and-white/panchromatic format, with lighter tones indicating warmer temperatures and darker grays and blacks signifying cooler surfaces. It is also common to display temperature variations in FLIR imagery as a density-sliced image (Chapter 5), using a more or less standardized color scale (whites, reds, yellows signifying warmer darker shades of blues, then black for cooler/cold colors.)

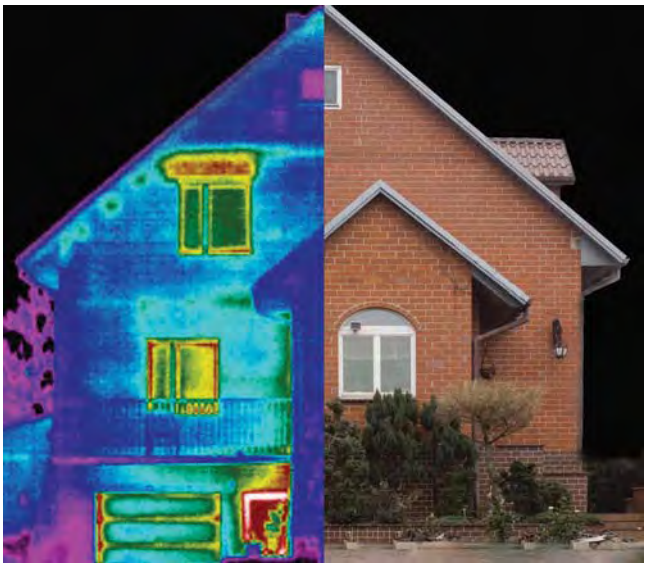


FIGURE 10.5 A residential structure with varied degrees of thermal protection. Dark blue sections indicate a favorable degree of energy retention. In contrast, light blue tones signify energy losses at this region of the structure. In winter, this section releases energy through this relatively transparent section. Yellow and red colors at window frames indicate higher energy losses through windows. The dark blue color at the roof suggests a careful effort to reduce energy losses at the roof. Image by Explorer1001 (2013).

10.6 THERMAL PROPERTIES OF OBJECTS

All objects at temperatures above absolute zero emit thermal radiation, although the intensity and peak wavelength of such radiation vary with the temperature of the object, as specified by the radiation laws outlined in Chapter 2. For remote sensing in the visible and NIR spectra, we examine contrasts in the abilities of objects to reflect direct solar radiation to the sensor. For remote sensing in the far-IR spectrum, we sense differences in the abilities of objects and landscape features to absorb shortwave visible and near-infrared radiation, then to emit this energy as longer wavelengths in the far-infrared region.

Thus, except for geothermal energy, anthropogenic thermal sources, and range and forest fires, the immediate source of emitted thermal infrared radiation is shortwave solar energy. Direct solar radiation (with a peak at about 0.5 μm in the visible spectrum) is received and absorbed by the landscape (Chapter 2). The amount and spectral distribution of energy emitted by landscape features depend on the thermal properties of these features, as discussed below. Contrasts in thermal brightness, observed as varied gray tones on the image, can form the basis for identifying specific features.

A *blackbody* is a theoretical object that acts as a perfect absorber and emitter of radiation; it absorbs and reemits all energy that it receives. Although the blackbody is a theoretical concept, it is useful in describing and modeling the thermal behavior of actual objects. Moreover, it is possible to approximate the behavior of blackbodies in laboratory experiments. As explained in Chapter 2, as the temperature of a blackbody increases,

the wavelength of peak emission decreases in accordance with Wien's displacement law (Equation 2.5). The Stefan–Boltzmann law (Equation 2.4) describes mathematically the increase in total radiation emitted (over a range of wavelengths) as the temperature of a blackbody increases.

Emissivity (ε_λ) is the ratio of emittance of an object to emittance of a blackbody at the same temperature:

$$\varepsilon_\lambda = \frac{\text{Radiant emittance of an object}}{\text{Radiant emittance of a blackbody at the same temperature}} \quad \text{(EQ. 10.2)}$$

(See also Equation 2.3.) The subscript (λ) sometimes used with ε signifies that λ has been measured for specific wavelengths. Emissivity therefore varies from 0 to 1, with 1 signifying a substance with a thermal behavior identical to that of a blackbody. Table 10.2 lists emissivities for some common materials. Note that many of the substances commonly present in the landscape (e.g., soil, water) have emissivities rather close to 1. Note, however, that emissivity can vary with temperature, wavelength, and angle of observation.

TABLE 10.2 Emissivities of Some Common Materials

Material	Temperature (°C)	Emissivity[a]
Polished copper	50–100	0.02
Polished brass	200	0.03
Polished silver	100	0.03
Steel alloy	500	0.35
Graphite	0–3,600	0.7–0.8
Lubricating oil (thick film on nickel base)	20	0.82
Snow	–10	0.85
Sand	20	0.90
Wood (planed oak)	20	0.90
Concrete	20	0.92
Dry soil	20	0.92
Brick (red common)	20	0.93
Glass (polished plate)	20	0.94
Wet soil (saturated)	20	0.95
Distilled water	20	0.96
Ice	–10	0.96
Carbon lamp black	20–400	0.96
Lacquer (matte black)	100	0.97

Note: Data from Hudson (1969) and Weast (1986).
[a]Measured at normal incidence over a range of wavelengths.

Graybodies

An object that has an emissivity less than 1.0 but constant emissivity over all wavelengths is known as a *graybody* (Figure 10.6). A *selective radiator* is an object with an emissivity that varies with respect to wavelength. If two objects in the same setting are at the same temperature but have different emissivities, the one having the higher emissivity will radiate more strongly. Because the sensor detects radiant energy (apparent temperature) rather than the kinetic ("true") temperature, precise interpretation of an image requires knowledge of the emissivities of features shown on the image.

Heat

Heat is the internal energy of a substance arising from the motion of its component atoms and molecules. *Temperature* measures the relative warmth or coolness of a substance. It is the kinetic temperature or average thermal energy of molecules within a substance. *Kinetic temperature*, sometimes known as the *true temperature*, is measured using the usual temperature scales, most notably the Fahrenheit, Celsius (centigrade), and Kelvin (absolute) scales. *Radiant* (or *apparent*) *temperature* measures the emitted energy of an object. Photons from the radiant energy are detected by the thermal scanner.

Heat capacity is the ratio of the change in heat energy per unit mass to the corresponding change in temperature (at constant pressure). For example, we can measure the heat capacity of pure water to be 1 calorie (cal) per gram (g), meaning that 1 cal is required for each gram to raise its temperature by 1°C. The specific heat of a substance is the ratio of its heat capacity to that of a reference substance. Because the reference substance typically is pure water, specific heat is often numerically equal to its heat capacity. Because a calorie is defined as the amount of heat required to raise by 1°C the temperature of 1 g of pure water, use of water as the reference means that heat capacity and specific heat will be numerically equivalent. In this context, specific heat can be defined as the amount of heat (measured in calories) required to raise the temperature of 1 g of a substance 1°C.

Thermal conductivity is a measure of the rate at which a substance transfers heat. Conductivity is measured as calories per centimeter per second per degree Celsius, so it measures calories required to transfer a change in temperature over specified intervals

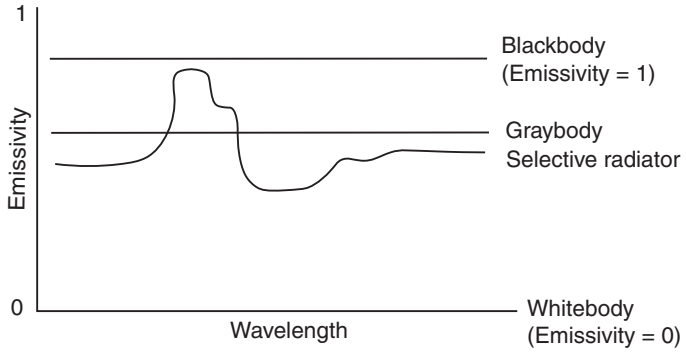


FIGURE 10.6 Thermal emissivity diagrams for a blackbody, graybody, and whitebody. Image by Susmita Sen.

of length and time. Some of these variables can be integrated into a single measure, called *thermal inertia* (P), which measures the tendency of a substance to resist changes in temperature, or, more precisely, the rate of heat transfer at the contact between two substances. P is defined as:

$$P = \sqrt{KC\rho} \quad \text{(EQ. 10.3)}$$

where K is the thermal conductivity (cal · cm^{-1} · sec^{-1} · $°C^{-1}$); C is the heat capacity (cal · g^{-1} · $°C^{-1}$); and ρ is the density (g · cm^{-3}). P is then measured in cal · cm^{-2} · $°C^{-1}$ · $sec^{-1/2}$.

In the context of remote sensing, thermal inertia indicates the ability of a surface to retain heat during the day and reradiate it at night, thereby signaling the nature of the varied thermal properties of the local terrain. The thermal inertia of specific surfaces (perhaps at the depths of several centimeters) will be determined by their physical characteristics, including mineralogy, particle size, compactness and lithification of mineral grains, and presence and depth of unconsolidated surface materials, such as sand, dust, and other loose sediments. Thus, implementation of remote sensing in the context of thermal inertia observes landscapes within the daily cycle of heating and cooling. Although a single thermal image provides only an isolated snapshot of relative temperatures, a *pair* of carefully timed thermal snapshots permits observation of temperature *changes* between the warmest and coolest portions of the day, and therefore offers an opportunity to observe differences in the thermal properties of the materials at the Earth's surface.

Temperatures of materials with low thermal inertia change significantly during the daily heating-cooling cycle, whereas temperatures of materials with high thermal inertias will respond more slowly. Thermal inertia characterizes a material's ability to conduct and store heat, and therefore its ability to retain heat during the day, then reradiate it at night. For remote sensing, thermal inertia represents a complex composite of factors such as particle size, soil cover, moisture, bedrock, and related terrain features. Relative thermal inertia can sometimes be approximated by assessing the amplitude of the diurnal temperature curve (i.e., the difference between daily maximum and minimum surface temperatures). Assessing differences in thermal inertia of the materials at the Earth's surface, in the context of other characteristics, can help to characterize these surface materials and their properties.

10.7 LAND SURFACE TEMPERATURE

Satellite remote sensing has made important contributions, not only to the practice of land remote sensing, but also to meteorology and climatology. Assessment of *land surface temperature (LST)* has provided a strategy for estimating the temperature of the Earth's surface through remotely sensed observations (Gillespie, 2014). (In Chapter 7, discussion of VIIRS introduced the role of such instruments to collect such data.) LST is most appropriate for estimating temperatures of surfaces and for tracking temperature patterns over time and is less suitable for assessing the temperatures of specific objects.

Land surface temperature is defined as the average temperature of a specified area of the Earth's surface (i.e., soil surfaces and vegetation canopies), as calculated from measured radiance observed from above (Jin and Dickinson, 2010). LST estimates are based on application of basic radiation laws (Chapter 2) that define relationships between the

intensity of emitted radiation and observed temperature. Because observed radiation has an intensity proportional to the temperature of the Earth's surface at the observed site, LST can estimate temperatures for specific sites as observed by aerial sensors. Thus, LST itself is not measured directly but through observations collected by airborne or spaceborne sensors that record energy radiated in the TIR region of the spectrum. LST has been estimated using AVHRR, MODIS (Wan and Dozier, 1996), VIIRS, and Landsat data, among others.

Calculating LST

Accurate estimation of LST requires application of relationships between energy emitted by a blackbody and peak wavelength of that energy, as an estimate of temperature. Thus, for example, Planck's Law (Section 2.4) can estimate intensity of radiation emitted by a blackbody in relation to peak wavelength (or frequency) for the radiation. Then, temperature estimates based on the radiation's peak wavelength provide a basis for estimates of the temperature of the surface.

Brightness Temperature

As mentioned above, a satellite sensor observes radiation emitted from a defined area at the Earth's surface. By observing the peak wavelength of that radiation, Planck's Law then forms the basis for estimation of the temperature of that object. One way to estimate LST is based on the inversion of Planck's Law, which defines the temperature (T_b) at which a blackbody (e_1) would emit radiance that is measured by remote sensing. This approach can be applied on a single image band (i.e., the thermal band):

$$T_b = \frac{c_2}{\lambda \ln\left(\frac{c_1}{B(\lambda, T)\pi\lambda^5} + 1\right)} \quad \text{(EQ. 10.4)}$$

where λ is the middle of the band wavelength range, T is temperature, and B is the spectral radiance at λ (Gillespie, 2014).

Model Temperature (Tm)

In some situations, image classification can identify surface composition, from which ε (emissivity) can then be inferred to provide a more refined estimate of LST (van de Griend and Owe, 1993). For example, the normalized difference vegetation index (NDVI) has been shown to be correlated with emissivity for certain vegetation types (van de Griend and Owe, 1993). Also, there are known emissivities for certain materials (e.g., Table 10.1), which could be one or more classes of interest.

Color Temperature (Tc)

Color temperature is defined as the temperature that satisfies Planck's Law measured at two different wavelengths. A simple way to estimate it is to calculate the ratio of two bands, where emissivity is known. This is most frequently done for night images in order to reduce the impact of solar reflection (Gillespie, 2014).

Generalized Split-Window Algorithm

Another important strategy for estimating LST relies on data collected by two separate instruments, using two separate channels with different wavelengths (Wan and Dozier, 1996). Alternative methods (mentioned above) rely on radiances collected in a single spectral channel, which may be effective with a clear atmospheric profile. The *split-window strategy* is based on spectral radiance differences collected by two different spectral channels that permit elimination or reduction of atmospheric effects. We note that this split-window strategy has been applied to the TIRS instrument for Landsat 8 (Caselles et al., 1998; Du et al., 2014). In the case of the Landsat 8 TIRS, there are two thermal channels at slightly different wavelengths, each at a swath width of 190 km, in the longwave infrared region. Thus, both TIRS sensors view the same footprint to record thermal radiation emitted by the Earth and its atmosphere (**Figure 10.7**). The use of these two separate thermal channels to view the same footprint permits separation of atmospheric effects from the ground temperature signal, thereby allowing TIRS to separate the atmospheric signal from the terrestrial signal, thereby improving the accuracy of the estimate of surface temperature. This *split-window algorithm* strategy produces a high-quality thermal signal, relative to alternative approaches.

Difficulties in Estimating LST

Regardless of the algorithmic approach to estimating LST, the variability of most land surfaces can result in factors that confound the ability to estimate and validate it. Some of these factors include the following:

- Varied angles of observation
- Sub-pixel variations in temperature and surfaces
- Variations in spectral emissivity at different wavelengths

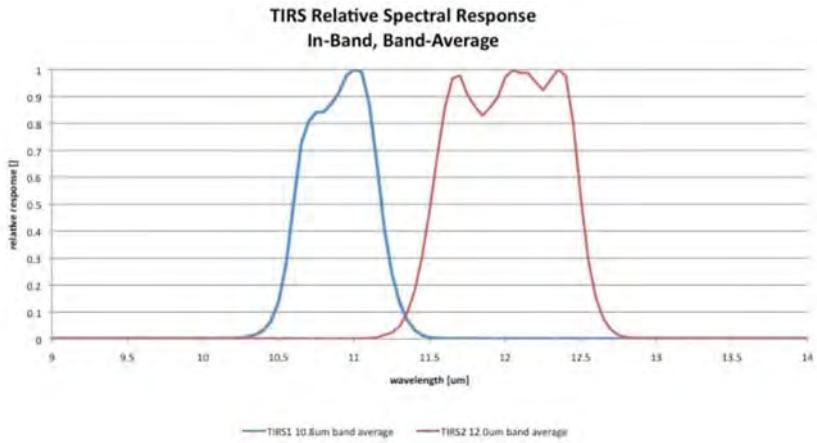


FIGURE 10.7 Landsat TIRS using two separate thermal channels to view a single footprint, permitting separation of atmospheric effects from estimates of ground surface temperatures, thereby improving the accuracy of surface temperature estimates. This strategy, known as the *split-window algorithm* (Caselles et al., 1998), improves the quality of the thermal signal.

- Effects of atmospheric temperature and humidity
- Presence of aerosols and clouds

Even under ideal circumstances, temporal variations (diurnal and seasonal) contribute to LST variations.

LST for Evapotranspiration and Drought

There are many applications for calculation of LST. One widely used application is the examination of LST over time to study evapotranspiration (ET) and drought (**Figure 10.8**). Numerous products have been designed for this purpose for different satellite platforms, including the MODIS evapotranspiration product (MOD16), which provides global estimates of ET at 1 km^2 every 8 days (Mu, Zhao, and Running, 2011). Products are also available for Landsat for some areas, including the Mapping EvapoTranspiration at high Resolution with Internalized Calibration (METRIC) product (Senay et al., 2016; Wang, He, and Hu, 2015). This product was designed to calculate per-pixel ET and has been used to gauge water use for western regions of the United States, where field sizes are large enough to be recorded by the 100m TIRS pixels (Allen et al. 2007; Willardson, 2014). It has been applied using Landsat 7 thermal data and later Landsat 8 TIRS. The basic premise behind these products is that irrigated fields are cooled by evapotranspiration. Therefore, temperatures indicate water use at specific fields or vegetated areas at particular times.

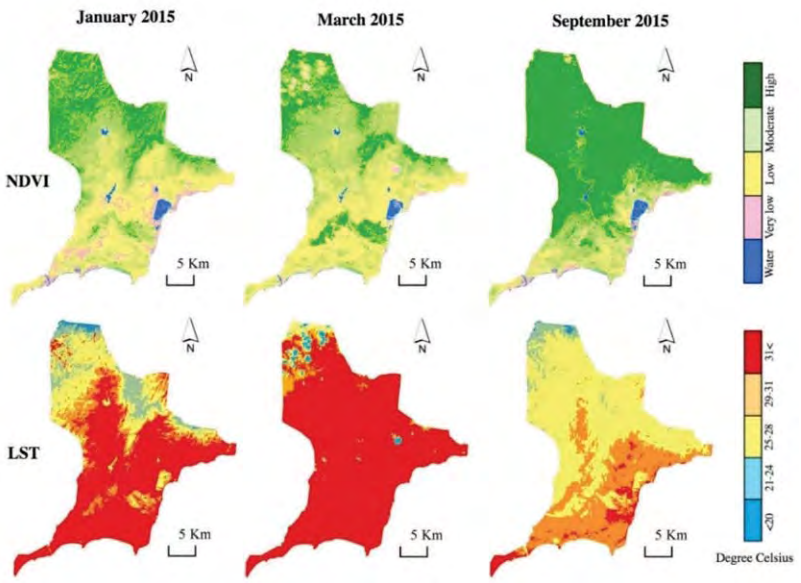


FIGURE 10.8 Classified images of sequential drought, Tuy Phong Province, Vietnam, as observed in January, March, and September 2015, illustrating the effectiveness of Landsat surface temperature estimates and the normalized difference vegetation index as strategies for monitoring drought progression. From Tran et al. (2017). Used by permission of Taylor & Francis.

10.8 GEOMETRY OF THERMAL IMAGES

Thermal scanners, like all remote sensing systems, generate geometric errors as they gather data. These errors mean that representations of positions and shapes of features depicted on thermal imagery do not match to their correct planimetric forms. Therefore, images cannot be directly used as the basis for accurate measurements. Some errors are caused by aircraft or spacecraft instability. As the aircraft rolls and pitches, the scan lines lose their correct positional relationships, and of course, the features they portray are not accurately represented in the image.

Thermal imagery also exhibits relief displacement analogous to that encountered in aerial photography (Figure 10.9). Thermal imagery, however, does not have the single central perspective of an aerial photograph, but rather a separate nadir for each scan line. Thus, the focal point for relief displacement is the nadir for each scan line or, in effect, the trace of the flight path on the ground. Thus, relief displacement is projected from a line that follows the center of the long axis of the image. At the center of the image, the sensor views objects from directly overhead, and planimetric positions are correct. However, as distance from the centerline increases, the sensor tends to view the sides rather than only the tops of features, and relief displacement increases. These effects are visible in Figure 10.9; the tanker and the tanks appear to lean outward from a line that passes through the center of the image. The effect increases toward the edges of the image.

Figure 10.9 also illustrates other geometric qualities of thermal line scan imagery. Although the scanning mirror rotates at a constant speed, the projection of the IFOV

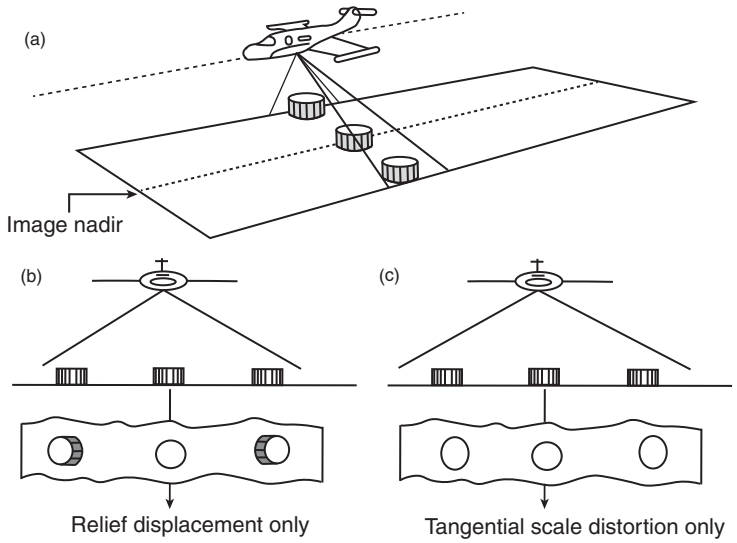


FIGURE 10.9 Relief displacement and tangential scale distortion. (a) Position of the plane relative to three objects. The center object is directly below the plane (i.e., nadir), so there is no relief displacement for this object and only the top is visible. (b) Relief displacement in the two outer objects. The tops of the objects appear to lean away from the sensor. (c) Tangential scale distortion in scanning sensors, caused by the movement of the scanner. The IFOV scans a larger area on the ground at the edge of the scene. Image by Susmita Sen.

onto the ground surface does not move at equal speed (relative to the ground) because of the varied distance from the aircraft to the ground. At nadir, the sensor is closer to the ground than it is at the edge of the image; in a given interval of time, the sensor scans a shorter distance at nadir than it does at the edge of the image. Therefore, the scanner produces a geometric error that tends to compress features along an axis oriented perpendicular to the flight line and parallel to the scan lines. In Figure 10.9, this effect, known as *tangential scale distortion,* is visible in the shapes of the cylindrical storage tanks. The images of those nearest the flight line are more circular, whereas the shapes of those farthest from the flight line (nearest the edge of the image) are compressed along an axis perpendicular to the flight line. Sometimes the worst effects can be removed by corrections applied as the film image is generated, although it is often necessary to avoid use of the extreme edges of the image.

10.9 THE THERMAL IMAGE AND ITS INTERPRETATION

The image generated by a thermal scanner usually appears as a black-and-white strip depicting thermal contrasts in the landscape as variations in gray tones (e.g., Figure 10.10). Brighter tones (whites and light grays) represent warmer features; darker tones (dark grays and blacks) represent cooler features. In some applications, the black-and-white image may be subjected to density slicing (Chapter 5) or other enhancements that assign distinctive hues to specific tones to aid visual interpretation. Often, it is easier for the eye to separate differing shades of color rather than variations in gray on the original image. Such enhancements are simply manipulations of the basic infrared image; they do not represent differences in either the means of acquisition or in the quality of the basic information available for interpretation.

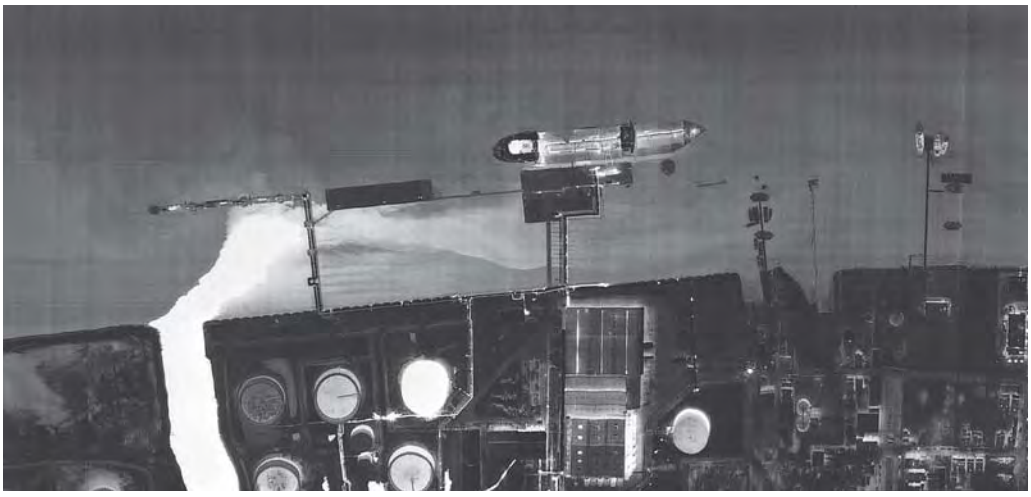


FIGURE 10.10 Thermal image of an oil tanker and petroleum storage facilities near the Delaware River, December 19, 1979. This image, acquired at 11:43 P.M., shows discharge of the warm water from the refinery into the Delaware River and the thermal patterns arising from operation of a large petroleum facility. From Daedalus Enterprises, Inc. Used by permission.

For any TIR image, the interpreter must always determine (1) whether the image at hand is a positive or a negative image, and (2) the time of day that the image was acquired. Sometimes it may not be possible to determine the correct time of day from information within the image itself; misinterpretation can alter the meaning of gray tones on the image and render the resulting interpretation useless.

Thermal scanners are generally uncalibrated, so they show relative radiances rather than absolute measurements of radiances. However, some thermal scanners do include reference sources that are viewed by the scanning system at the beginning and end of each scan. The reference sources can be set at specific temperatures that are related to those expected to be encountered in the scene. Thus, each scan line includes values of known temperature that permit the analyst to estimate temperatures of objects within the image.

As noted above, certain geometric distortions can confound the interpretation of the image. As the sensor views objects near the edge of the image, the distance from the sensor to the ground increases. This relationship means that the IFOV is larger near the edges of the image than it is at the flight line.

In addition, errors caused by the atmosphere and by the system itself prevent precise interpretation of thermal imagery. Typical system errors might include recording noise, variations in reference temperatures, and detector errors. Full correction for atmospheric conditions requires information that is not usually available in detail. Often, then, it is necessary to use approximations, or value-based samples acquired at a few selected times and places, which are then extrapolated to estimate values elsewhere. Also, the *atmospheric path traveled* by radiation reaching the sensor varies with angle of observation, which changes as the instrument scans the ground surface. These variations in angle can lead to errors in observed values within the image.

Even when accurate measures of radiances are available, it is difficult to derive data for kinetic temperatures from the apparent temperature information within the image. Derivation of kinetic temperatures requires knowledge of emissivities of the materials. In some instances, such knowledge may be available, as the survey may be focused on a known area that must be repeatedly imaged to monitor changes over time (e.g., as moisture conditions change). But many other surveys examine areas not previously studied in detail, and information regarding surface materials and their emissivities may not be known.

Emissivity is a measure of the effectiveness of an object in translating temperature into emitted radiation (and in converting absorbed radiation into a change in observed temperature). Because objects differ with respect to emissivity, observed differences in emitted infrared energy do not translate directly into corresponding differences in temperature. As a result, it is necessary to apply knowledge of surface temperature or of emissivity variations to accurately study surface temperature patterns from thermal imagery. Because knowledge of these characteristics assumes a detailed prior knowledge of the landscape, such interpretations should be considered as appropriate for examination of a distribution known already in some detail rather than for reconnaissance of an unknown pattern (e.g., one might already know the patterns of soils and crops at an agricultural experiment station but may wish to use the imagery to monitor temperature patterns). Often, estimated values for emissivity are used, or assumed values are applied to areas of unknown emissivity.

It should also be recognized that the sensor records radiances at surfaces of objects. Because radiances may be determined at the surface of an object by a layer perhaps as thin as 50 μm, a sensor may record conditions that are not characteristic of the overall

subsurface mass, which is often the object of the study. For example, evaporation from a water body or a moist soil surface may cool the thin layer of moisture at the contact point with the atmosphere. Because the sensor detects radiation emitted at this surface layer, the observed temperature may differ considerably from that of the remaining mass of the soil or water body. Leckie (1982) estimates that calibration error and other instrument errors are generally rather small, although they may be important in some instances. Errors in estimating emissivity and in attempts to correct for atmospheric effects are likely to be the most important sources of error in quantitative studies of thermal imagery.

In many instances, a thermal image must be interpreted to yield qualitative rather than quantitative information. Although some applications do require interpretations of quantitative information, there are many others for which qualitative interpretation is completely satisfactory. An interpreter who is well informed about the landscape represented on the image, the imaging system, the thermal behavior of various materials, and the timing of the flight is prepared to derive considerable information from an image, even though it may not be possible to derive precise temperatures.

The Thermal Landscape

The *thermal landscape* is a composite of the familiar elements of surface material, topography, vegetation cover, and moisture. Various rocks, soils, and other surface materials respond differently to solar heating. Thus, in some instances, the differences in thermal properties listed in Table 10.2 can be observed in thermal imagery. However, the thermal behavior of surface materials is also influenced by other factors. For example, slopes that face the Sun will tend to receive more solar radiation than slopes that are shadowed by topography. Such differences are, of course, combined with those arising from different surface materials. Also, the presence and nature of vegetation alter the thermal behavior of the landscape. Vegetation tends to heat rather rapidly but can also shade areas, creating patterns of warm and cool.

Water tends to retain heat, to cool slowly at night, and to warm slowly during the daytime. In contrast, many soils and rocks (if dry) tend to release heat rapidly at night and to absorb heat quickly during the daytime. Even small or modest amounts of moisture can greatly alter the thermal properties of soil and rock. Therefore, thermal sensors can be effective in monitoring the presence and movement of moisture in the environment. In any given image, the influences of surface materials, topography, vegetation, and moisture can combine to cause complex image patterns. However, often it is possible to isolate the effect of some of these variables and therefore to derive useful information concerning, for example, movement of moisture or the patterns of differing surface materials.

Timing the acquisition of thermal imagery is very important. Optimum times vary according to the purpose and subject of the study, so it is not possible to specify universally applicable rules. Because the greatest thermal contrast tends to occur during the daylight hours, sometimes thermal images are acquired in the early afternoon to capture the differences in the thermal properties of landscape features. However, in the 3–6 μm range, the sensor may record reflected as well as emitted thermal radiation, so daytime missions in this region may not be optimum for thermal information. Also, during daytime, the sensor may record thermal patterns caused by topographic or cloud shadowing. Although shadows may sometimes be useful in interpretation, they are more likely to complicate analysis of a thermal image, so it is usually best to avoid acquiring heavily

shadowed images. In a daytime image, water bodies typically appear as cool relative to land, and bare soil, meadow, and wooded areas appear as warm features.

Some of the problems arising from daytime images are avoided by planning missions just before dawn. Shadows are absent, and sunlight, of course, cannot cause reflection (at shorter wavelengths) or shadows. However, thermal contrast is lower, so it may be more difficult to distinguish between broad classes of surfaces based on differences in thermal behavior. On such images, water bodies would appear to be warm relative to land. Forested areas may also appear to be warm. Open meadows and dry, bare, soil patches are likely to appear as cool features.

The thermal images of petroleum storage facilities (Figures 10.10 and 10.11) show thermal contrasts that are especially interesting. A prominent feature in Figure 10.10 is the bright thermal plume discharged by the tributary to the Delaware River. The image clearly shows the sharp contrast in temperature as the warm water flows into the main channel, then disperses and cools as it is carried downstream. Note the contrast between the full and partially full tanks and the warm temperatures of the pipelines that connect the tanker with the storage tanks. Many of the same features are also visible in Figure 10.11, which shows a partially loaded tanker with clear delineation of the separate storage tanks within the ship.

In Figure 10.12, two thermal images depict a portion of the Cornell University campus in Ithaca, New York, acquired in January (left) and again the following November (right). Campus buildings are clearly visible, as are losses of heat through vents in the roofs of buildings and at manholes where steam pipes for the campus heating system join or change direction. The left-hand image shows a substantial leak in a steam pipe as it

FIGURE 10.11 Thermal image of oil tankers and petroleum storage facilities, Delaware River, December 1979. Here we see internal compartments in the larger tanker and a broader view of the petroleum facility, with details of petroleum storage. Note that this image shows the image nadir, oriented left-to-right along the center of the image, with relief displacement revealing the outward-oriented sides of the tankers (at the upper edge) and storage tanks (at the lower edge). From Daedalus, Inc. Used by permission.

FIGURE 10.12 Two images of a portion of the Cornell University campus, Ithaca, New York, as imaged at nighttime, January and November of the same year. These images were acquired to monitor losses of heat in the underground steam transmission system. The large leak at the center right of the January image has been repaired in the November image. From Daedalus Enterprises, Inc. Used by permission.

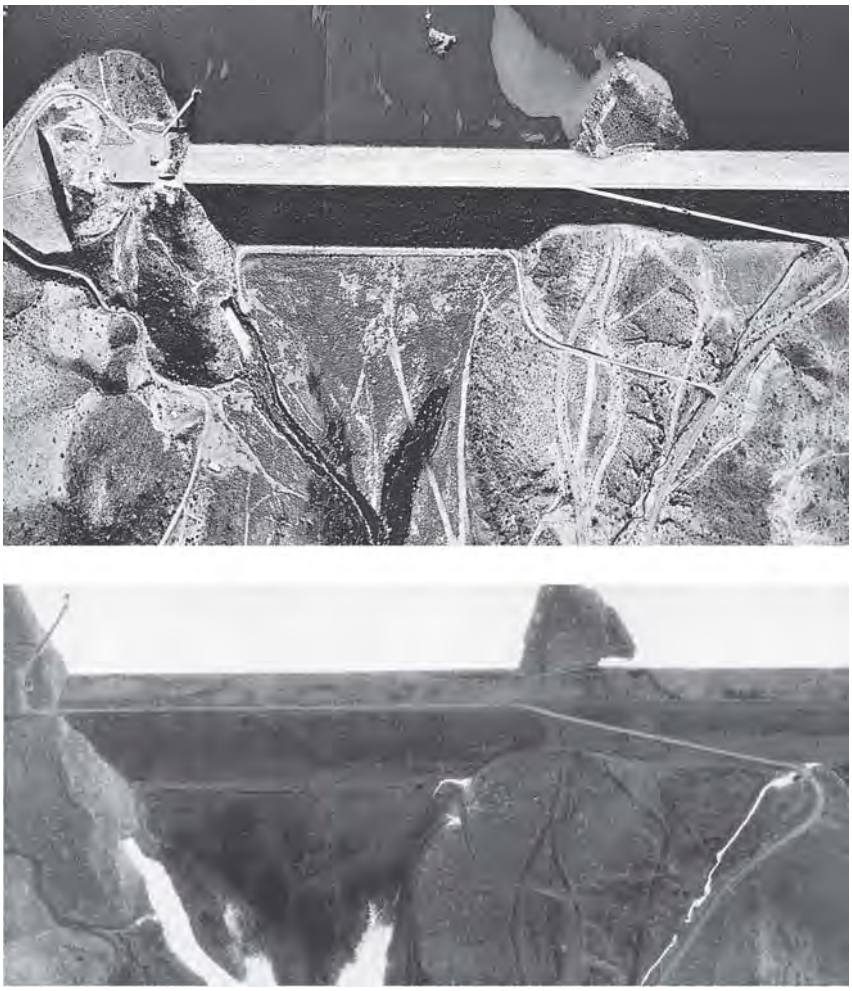


FIGURE 10.13 Painted Rock Dam, Arizona, January 28, 1979. Aerial photograph (top) and thermal image (bottom). From Daedalus Enterprises, Inc. Used by permission.

passes over the bridge in the right center of the image. On the right, a later image of the same region shows clearly the effects of repair of the defective section.

Figure 10.13 shows Painted Rock Dam, Arizona, as depicted by both an aerial photograph and a thermal infrared image. The aerial photograph (top) was taken at about 10:30 A.M., and the thermal image was acquired at about 7:00 A.M. the same day. The prominent linear feature is a large earthen dam, with the spillway visible at the lower left. On the thermal image, the open water upstream from the dam appears as a uniformly white (warm) region, whereas land areas are dark (cool)—a typical situation for the early morning hours, before solar radiation has warmed the Earth. On the downstream side of the dam, the white (warm) regions reveal areas of open water or saturated soil. The open water in the spillway is, of course, expected, but the other white areas indicate places where there may be seepage and potentially weak points in the dam structure.

Thermal Inertia

If imagery or data for two separate times are available, it may be possible to employ knowledge of *thermal inertia* as a means of studying the pattern of different materials at the Earth's surface. Figure 10.14 illustrates the principles involved—it represents two images acquired at times that permit observation of extremes of temperature, perhaps near noontime and again just before dawn. These two sets of data permit estimation of the ranges of temperature variation for each region on the image. Because these variations are determined by the thermal inertias of the substances, they permit interpretation of features represented by the images. Leckie (1982) notes that misregistration can be a source of error in comparisons of day and night images, although such errors are thought to be small relative to other errors.

Landsat Daytime/Nighttime Thermal Examples

Although the timing of Landsat 8–based thermal measurements (and other satellite-based measurements) is constrained by the orbit characteristics, Landsat 8 does occasionally capture data at night in ascending mode. This provides some opportunity to capitalize on the differences between day and night imagery described above.

Figure 10.15 shows the Grand Junction, Colorado, region as imaged by Landsat 8 OLI (daytime) and TIRS (Table 10.1, nighttime) scenes, June 2016. This example provides

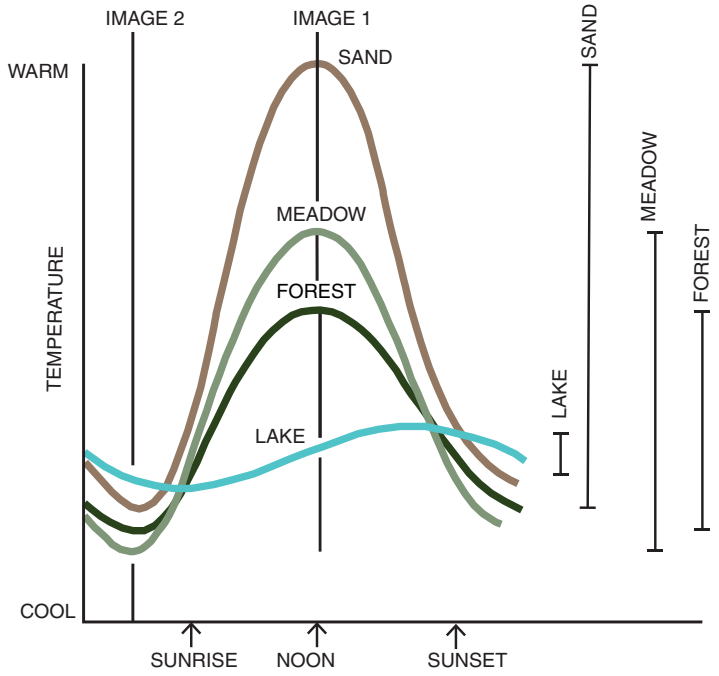


FIGURE 10.14 Schematic illustration of diurnal temperature variation of several broad classes of land cover, illustrating their responses over a 24-hour interval to daily temperature variations. Image by Susmita Sen.

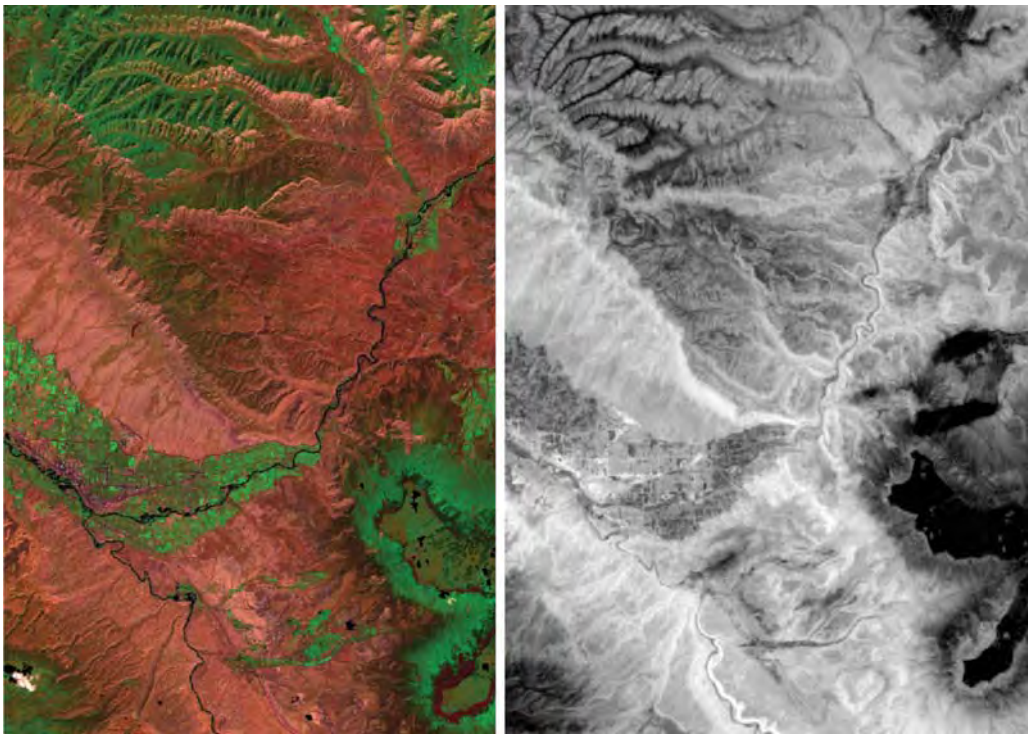


FIGURE 10.15 Grand Junction, Colorado. Left: Daytime Landsat 8 (bands 7-5-8) (at 30-m detail) multispectral imagery. Green represents forested and vegetated areas, often on north-facing slopes, shaded from direct sunlight for much of the day. The reddish tones represent the largely dry, sparsely vegetated, rangelands and unvegetated surfaces open to direct sunlight during daytime. The green strip at the center and left of the image is the settlement of Grand Junction, including urban landscapes and adjacent irrigated lands. Right: Nighttime TIRS thermal imagery (June 2016) (at 100-m detail). In the nighttime image, note the large temperature variations related to terrain elevation and aspect. Imaged prepared by Tammy Parece. Used by permission.

a distinctive situation illustrating differential heating and cooling, and variations in the ability of landscapes to retain heat after sunset. Here, local relief is unusually high in the sense that, for this region of northwestern Colorado, typical valley elevations are about 4,500 feet above sea level, and the highest at about 10,000 feet, so local relief within this image is at least 5,000 feet. The Grand Junction region, at the confluence of the Colorado and Gunnison rivers, is depicted as a patchwork of light and dark features—paved surfaces, parks, local irrigation, and vegetation, characteristic of urban landscapes. At the center right edge of the image, the jet-black shapes show the cooler surfaces of Grand Mesa (about 5,000 ft higher than Grand Junction). In low-lying terrain at the center of the image, irrigated landscapes bordering rivers at this time of year provide enough moisture to retain, then slowly release, heat during the nighttime hours. In higher canyons, steep slopes that face the Sun accumulate heat during the day, to slowly reradiate it at night, forming the light-toned surfaces in the image. Narrow, steep-walled canyons, shadowed during the day, are cool at night as they quickly lose the radiation accumulated

during the day. On average, in June, the mesa surface is about 18° cooler than the valley terrain, so it shows as a jet-black surface on this nighttime thermal image.

The significant amount of impervious surface and human activities in urban areas causes notable thermal inertial patterns that are distinctive in thermal energy. The *urban heat island* effect, where urban areas are warmer than the surrounding landscape, can often be seen, especially on calm nights. Figure 10.16 presents two Landsat images of the Roanoke-Salem, Virginia, region, in April 2010 (Landsat 7, day), and June 2016 (Landsat 8, night). The images were acquired for clear, calm conditions after several consecutive days of solar heating, creating a distinctive meteorological situation that coincided with cloud-free conditions at the time of these Landsat overpasses. In the daytime image (left), the red tones (chiefly the TIR channel) signify peak emission in thermal region (roughly 10 μm)—mainly pavement, parking lots, central business district, arterial highways, and impervious surfaces, bordered by structures and parking lots. Such surfaces, if exposed to the solar beam, will heat up, emitting preferentially in the TIR. The yellowish colors (green + red) depict hot surfaces emitting in or near the SWIR and far-infrared regions (see Table 10.1 for wavelengths). Blue-tinted white tones (bright in all three bands) signify surfaces such as the roofs of large buildings, often warehouses or industrial buildings, reflecting or emitting radiation in all three channels. In the nighttime image (right), the light-toned network is formed mainly by warmer roadways and parking lots that are emitting thermal radiation from impervious surfaces. The light-toned gray features between roadways are largely vegetated or forested regions. Dark features in the daytime image are mainly cooler shadowed terrain that did not retain much heat during the day; most of the black, dot-shaped features are mainly rooftops of large refrigerated warehouses.

For the same daytime image shown in Figure 10.16, Figure 10.17 shows the urban heat island during the day for nearby Blacksburg, Virginia. This example illustrates the

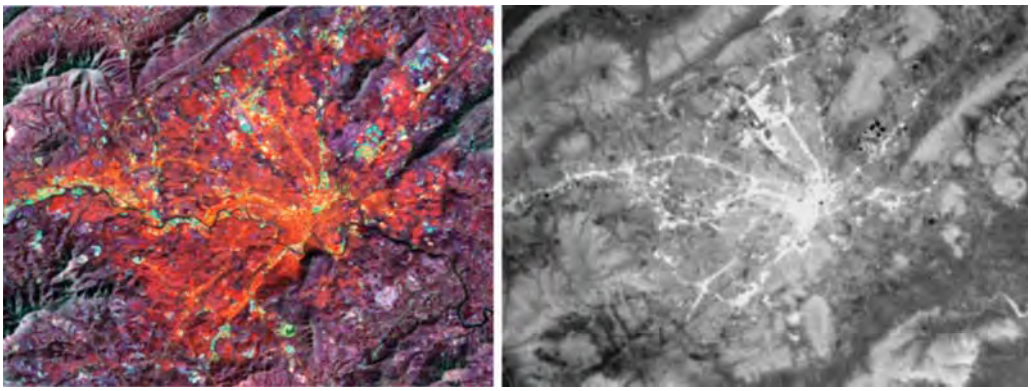


FIGURE 10.16 Shortwave infrared region (SWIR) and thermal images for Roanoke, Virginia, illustrating differences between daytime and nighttime thermal patterns. Left: Landsat 7 daytime multispectral image (bands 675) (approximately 10:30 A.M., September 2010). This daytime image shows both thermal emissions from the landscape and reflected solar radiation. Right: Nighttime Landsat 8 TIRS thermal imagery (approximately 11 P.M., June 2016). The nighttime image shows thermal radiation emitted from impervious surfaces heated by solar radiation throughout daytime hours. To a lesser extent, the nighttime image may also record emitted radiation from industrial activities.

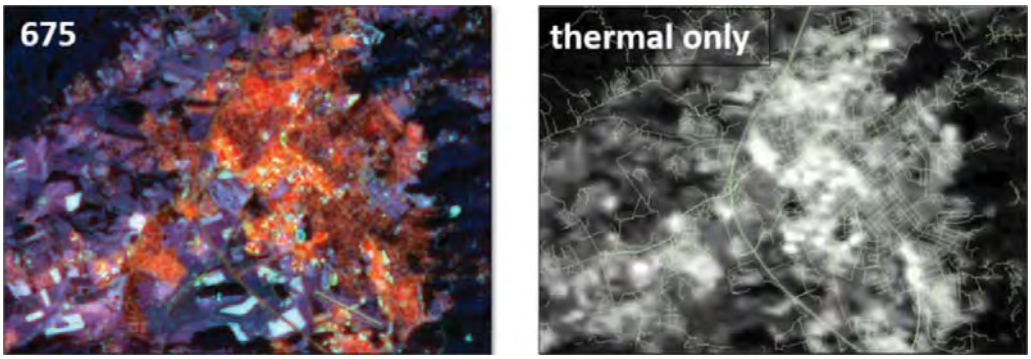


FIGURE 10.17 Urban heat island effect from Landsat 7 imagery, Blacksburg, Virginia (September 2010). Left: Thematic Mapper (TM) bands 675 (see **Table 10.1** for wavelengths). Right: TM band 6 (thermal band, at 60-m resolution), with a shapefile overlay depicting the street pattern.

warming of calm air in proximity to sizeable areas of impervious surface. The left-hand image (TM bands 675) displays warm air visible as the red shapes. The right-hand image displays only the daytime thermal band (TM band 6). Here, pockets of warm air are visible as the cloud-like shapes positioned at impervious surfaces, such as parking lots, large buildings, and dense street patterns.

Landsat Observations of Fire

During the summer of 2015, there was a very serious outbreak of fire in California known as the Shasta-Trinity fires (named for two adjacent northern California counties). Officially, the Shasta-Trinity fires started on July 30, initiated from widespread lightning strikes, extending, despite efforts to suppress the spread of fire, until October 29–November 3, 2015, together burning about 113,584 km^2. Irregular terrain, dense forests, sparse road networks, and dry summers in the Mediterranean climate of the California coastline are several of the many factors that create serious fire hazards in this region. The fires and their smoke plumes can be clearly seen in Landsat 8 imagery (Figure 10.18). The VNIR reflectance data (left) highlights the smoke plumes but does not depict the fires very clearly because it does not capture their emitted energy. In contrast, the SWIR data (right) minimizes smoke and haze, while depicting burned areas.

To depict burned areas more clearly, two normalized burn ratios have been developed and commonly applied, and can be downloaded as products from Landsat imagery. The normalized burn ratio (NBR) uses the standard normalized form with NIR and SWIR data, as follows (Lopez Garcia and Caselles, 1991; Key and Benson, 2006):

$$\mathrm{NBR} = \left(\frac{\mathrm{NIR} - \mathrm{SWIR}}{\mathrm{NIR} + \mathrm{SWIR}} \right) \qquad \text{(EQ. 10.5)}$$

The SWIR2 band is most commonly used with NBR (see Table 10.1 for wavelengths) and provides great contrast with the NIR band, minimizing clouds and highlighting burned areas (Figure 10.19, left).

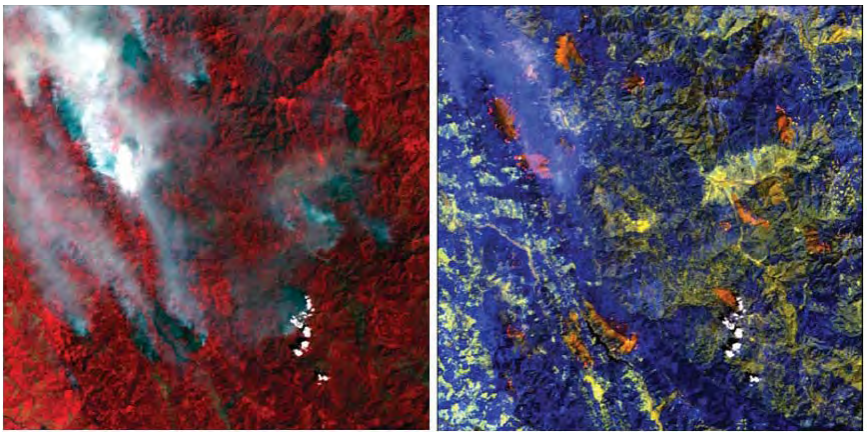


FIGURE 10.18 Forest fires, Shasta-Trinity Forest, California, August 2015, as recorded by Landsat 8 OLI. Left: Standard false color near-infrared-radiation band combination—fires and burned areas are poorly delineated and largely obscured by smoke. Right: Two-band combination of SWIR 1 and SWIR 2 (red). Note the effectiveness of the SWIR ratio in minimizing effects of smoke and haze, permitting observation of the extent of burned areas (orange patches). Images prepared by Hoa Tran. Used by permission.

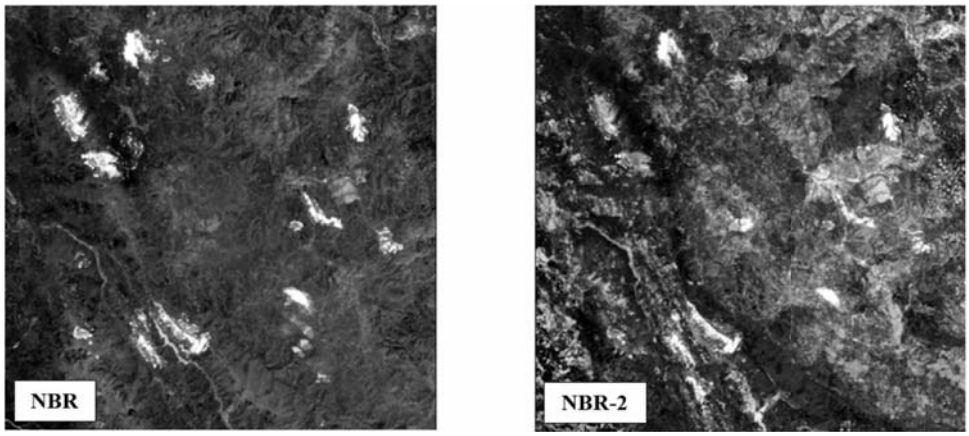


FIGURE 10.19 Forest fires in the Shasta-Trinity Forest, August 2015, as observed using alternative burn ratios. Left: NBR ratio. Right: NBR-2 ratio. The alternative burn ratios provide differing perspectives on this landscape. Images prepared by Hoa Tran. Used by permission.

In contrast, the second normalized burn ratio (NBR-2) (Figure 10.19, right) substitutes the SWIR1 band for the NIR (*www.usgs.gov/land-resources/nli/landsat/landsat-normalized-burn-ratio-2*). It is designed to highlight burned areas, but also water sensitivity in vegetation, which may be important for recovery. Because SWIR radiation can penetrate atmospheric smoke and haze, it is effective in monitoring forest and range fires, defining regions of active burning, detecting hot spots, and estimating the severity of burned regions. The NBR-2 image in Figure 10.19 (right) shows recently burned areas, but also depicts variations in relief, vegetation patterns, and local terrain. It is calculated as follows:

$$\text{NBR} = \left(\frac{\text{SWIR1} - \text{SWIR2}}{\text{SWIR1} + \text{SWIR2}} \right) \quad \text{(EQ. 10.6)}$$

To further explore the impacts of fire on the landscape, it is also possible to calculate the difference in burn ratios for images taken pre- and postfire (Key and Benson, 2006; Escuin, Navarro, and Fernandez, 2008). This is referred to as a differenced NBR.

10.10 SUMMARY

Thermal imagery is a valuable asset for remote sensing because it conveys information that is not easily derived from other forms of imagery. The thermal behavior of different soils, rocks, and construction materials can permit derivation of information not present in other images. The thermal properties of water contrast with those of many other landscape materials, so that thermal images can be sensitive to the presence of moisture in the environment. Furthermore, the presence of moisture is itself often a clue to differences between different classes of soil and rock.

Of course, use of data from the far-infrared region can present its own problems. Like all images, thermal imagery has geometric errors. Moreover, the analyst cannot derive detailed quantitative interpretations of temperatures unless detailed knowledge of emissivity is at hand. Timing of image acquisition can be critical. Atmospheric effects can pose serious problems, especially from satellite altitudes. Because the thermal landscape differs so greatly from the visible landscape, it may often be necessary to use aerial photography to locate familiar landmarks while interpreting thermal images. Existing archives of thermal imagery are not comparable in scope to those for aerial photography or satellite data (such as those of Landsat or SPOT), so it may be difficult to acquire suitable thermal data unless it is feasible to purchase custom-flown imagery.

REVIEW QUESTIONS

1. Explain why choice of time of day is so important in planning acquisition or analysis of thermal imagery.
2. Why do you expect season of the year to be important for analysis?
3. In your new job as an analyst for an institution that studies environmental problems in coastal areas, it is necessary for you to prepare a plan to acquire thermal imagery of a tidal marsh. List the important factors you must consider as you plan the mission.

4. In what ways would thermal imagery be important for agricultural research?
5. How would thermal imagery be useful for studies of the urban landscape?
6. Examine **Figure 10.10**. Is the tanker empty, partially full, or full? Compare this to **Figure 10.11**; is the tanker empty, partially full, or full? From your inspection of the imagery, what can you determine about the construction of tankers and the procedures used to empty or fill tankers?
7. Why is thermal imagery considered so useful for scientists even though it does not usually provide measurements of actual temperature? Can you identify situations where it might be important to determine actual temperatures from imagery?

REFERENCES

Allen, R. G., M. Tasumi, A. Morse, R. Trezza, J. L. Wright, . . . and C. W. Robison. 2007. Satellite-Based Energy Balance for Mapping EvapoTranspiration at High Resolution with Internalized Calibration (METRIC)—Applications. *Journal of Irrigation and Drainage Engineering*, Vol. 133, No. 4, pp. 395–406.

Barsi, J. A., K. Lee, G. Kvaran, B. L. Markham, & J. A. Pedelty. 2014. The Spectral Response of the Landsat-8 Operational Land Imager. *Remote Sensing*, Vol. 6, No. 10, pp. 10232–10251.

Caselles, V., M. M. Artigao, E. Hurtado, C. Col, and A. Brasa. 1998. Mapping Actual Evapotranspiration by Combining Landsat TM and NOAA-AVHRR Images: Application to the Barrax Area, Albacete, Spain. *Remote Sensing of Environment,* Vol. 63, pp. 1–10.

Colcord, J. E. 1981. Thermal Imagery Energy Surveys. *Photogrammetric Engineering and Remote Sensing*, Vol. 47, pp. 237–240.

Dennison, P. E., K. Charoensir, D. A. Roberts, S. H. Petersen, and R. O. Green. 2006. Wildfire Temperature and Land Cover Modeling Using Hyperspectral Data. *Remote Sensing of Environment,* Vol. 100, pp. 212–222.

Du, C., H. Ren, Q. Qin, J. Meng and J. Li. 2014. Split-Window Algorithm for Estimating Land Surface Temperature from Landsat 8 TIRS Data. 2014 IEEE Geoscience and Remote Sensing Symposium, Quebec City, QC, pp. 3578–3581.

Escuin, S., R. Navarro, and P. Fernandez. 2008. Fire Severity Assessment by Using NBR (Normalized Burn Ratio) and NDVI (Normalized Difference Vegetation Index) Derived from LANDSAT TM/ETM Images. *International Journal of Remote Sensing,* Vol. 29, No. 4, pp. 1053–1073.

Explorer1001. 2013, 14 January. Termowizja [Thermovision]. In Wikipedia. *https://commons.wikimedia.org/wiki/File:Termowizja.jpg*

Fagerlund, E., B. Kleman, L. Sellin, and H. Svenson. 1970. Physical Studies of Nature by Thermal Mapping. *Earth Science Reviews,* Vol. 6, pp. 169–180.

Fraser, R., and J. J. Kay. 2004. Energy Analysis of Ecosystems: Establishing a Role for Thermal Remote Sensing. Chapter 8 in *Thermal Remote Sensing in Land Surface Processes* (D. A. Quattrochi and J. C. Luvall, eds.). Boca Raton, FL: CRC Press, pp. 283–360.

Gillespie, A. R. 2014. Land Surface Emissivity. In *Encyclopedia of Remote Sensing* (E. G. Njoku, ed.), New York: Springer, pp. 303–310.

Gillespie, A. R., and A. B. Kahle. 1978. Construction and Interpretation of a Digital Thermal Inertia Image. *Photogrammetric Engineering and Remote Sensing,* Vol. 43, pp. 983–1000.

Gluch, R., D. A. Quattrochi, and J. C. Luvall. 2006. A Multiscale Approach to Urban Thermal Analysis. *Remote Sensing of Environment,* Vol. 104, pp. 123–131.

Goddard Space Flight Center. 1978. *Data Users Handbook, Heat Capacity Mapping Mission (HCMM), for Applications Explorer Mission-A (AEM)).* Greenbelt, MD: NASA.

Hudson, R. D. 1969. *Infrared Systems Engineering.* New York: John Wiley, 642 pp.
Jakosky, B. M., M. T. Mellon, H. H. Kieffer, and P. R. Christensen. 2000. The Thermal Inertia of Mars from the Mars Global Surveyor Thermal Emission Spectrometer. *Journal of Geophysical Research,* Vol. 105, pp. 9643–9652.
Jeevalakshimia, D., S. S. Nareayana Reddy, and B. Manikiam. 2017. Land Surface Temperature Retrieval from Landsat Using Emissivity Estimation. *International Journal of Applied Engineering,* Vol. 12, No. 20, pp. 9679–9687.
Jin, M., and R. E. Dickinson. 2010. Land Surface Skin Temperature Climatology: Benefitting from the Strengths of Satellite Observations. *Environmental Research Letters*, Vol. 5, No. 4, p. 044004.
Kahle, A. B. 1977. A Simple Model of the Earth's Surface for Geologic Mapping by Remote Sensing. *Journal of Remote Sensing,* Vol. 82, pp. 1673–1680.
Key, C., and N. Benson, 2006. Landscape Assessment: Remote Sensing of Severity, the Normalized Burn Ratio; and Ground Measure of Severity, the Composite Burn Index. In *FIREMON: Fire Effects Monitoring and Inventory System,* RMRS-GTR. Ogden, UT: USDA Forest Service, Rocky Mountain Research Station.
Leckie, D. G. 1982. An Error Analysis of Thermal Infrared Line-Scan Data for Quantitative Studies. *Photogrammetric Engineering and Remote Sensing,* Vol. 48, pp. 945–954.
Lopez Garcia, M. J., and V. Caselles. 1991. Mapping Burns and Natural Reforestation Using Thematic Mapper Data. *Geocarto International,* Vol. 6, pp. 31–37.
McVicar, T. R., and D. L. B. Jupp. 1999. Estimating One-Time-of-Day Meteorological Data from Standard Daily Data as Inputs to Thermal Remote Sensing-Based Energy Balance Models. *Agricultural and Forest Meteorology,* Vol. 96, pp. 219–238.
Mellon, M. T., B. M. Jakosky, H. H. Kieffer, and P. R. Christensen. 2000. High-Resolution Thermal-Inertia Mapping from the Mars Global Surveyor Thermal Emission Spectrometer. *Icarus,* Vol. 148, pp. 437–455.
Moore, R. K. (ed.). 1975. Microwave Remote Sensors. Chapter 9 in *Manual of Remote Sensing* (R. G. Reeves, ed.). Falls Church, VA: American Society of Photogrammetry, pp. 399–538.
Mu, Q., M. Zhao, and S. W. Running. 2011. Improvements to a MODIS Global Terrestrial Evapotranspiration Algorithm. *Remote Sensing of Environment,* Vol. 115, pp. 1781–1800.
Norman, J. R., and F. Becker. 1995. Terminology in Thermal Infrared Remote Sensing of Natural Surfaces. *Remote Sensing Reviews,* Vol. 12, pp. 159–173.
Prata, A. J., V. Caselles, C. Coll, J. A. Sobrino, and C. Ottlé. 1995. Thermal Remote Sensing of Land Surface Temperature from Satellites: Current Status and Future Prospects. *Remote Sensing Reviews,* Vol. 12, pp. 175–224.
Pratt, D. A., and C. D. Ellyett. 1979. The Thermal Inertia Approach to Mapping of Soil Moisture and Geology. *Remote Sensing of Environment,* Vol. 8, pp. 151–168.
Price, J. C. 1978. Thermal Inertia Mapping: A New View of the Earth. *Journal of Geophysical Research,* Vol. 82, pp. 2582–2590.
Price, J. C. 1981. The Contribution of Thermal Data in Landsat Multispectral Classification. *Photogrammetric Engineering and Remote Sensing*, Vol. 47, pp. 229–236.
Price, J. C. 1985. On the Analysis of Thermal Infrared Imagery: The Limited Utility of Apparent Thermal Inertia. *Remote Sensing of Environment,* Vol. 18, pp. 59–73.
Quattrochi, D. A., and J. C. Luvall. 1999. Thermal Infrared Remote Sensing for Analysis of Landscape Ecological Processes: Methods and Applications. *Landscape Ecology,* Vol. 14, pp. 577–598.
Quattrochi, D. A., and J. C. Luvall. 2004. *Thermal Remote Sensing in Land Surface Processing.* Boca Raton, FL: CRC Press, 142 pp.
Quattrochi, D. A., and J. C. Luvall. 2009. Thermal Remote Sensing in Earth Science Research. Chapter 5 in *The Sage Handbook of Remote Sensing* (T. A. Warner, M. D. Nellis, and G. M. Foody, eds.). London: SAGE, pp. 64–78.
Ramsey, M. S., and J. H. Fink. 1999. Estimating Silicic Lava Vesicularity with Thermal Remote

Sensing: A New Technique for Volcanic Mapping and Monitoring. *Bulletin of Volcanology,* Vol. 61, pp. 32–39.

Sabins, F. 1969. Thermal Infrared Imaging and Its Application to Structural Mapping, Southern California. *Geological Society of America Bulletin,* Vol. 80, pp. 397–404.

Schott, J. R., and W. J. Volchok. 1985. Thematic Mapper Infrared Calibration. *Photogrammetric Engineering and Remote Sensing,* Vol. 51, pp. 1351–1358.

Senay, G. B., M. Friedrichs, R. K. Singh, N. Manohar, and N. M. Velpuri. 2016. Evaluating Landsat 8 Evapotranspiration for Water Use Mapping in the Colorado River Basin. *Remote Sensing of Environment,* Vol. 185, pp. 171–185.

Short, N. M., and L. M. Stuart. 1982. *The Heat Capacity Mapping Mission (HCMM) Anthology* (NASA Special Publication 465). Washington, DC: U.S. Government Printing Office, 264 pp.

Toll, D. L. 1985. Landsat-4 Thematic Mapper Scene Characteristics of a Suburban and Rural Area. *Photogrammetric Engineering and Remote Sensing,* Vol. 51, pp. 1471–1482.

Tran, H. T., J. B. Campbell, T. D. Tran, and H. T. Tran. 2017. Monitoring Drought Vulnerability Using Multispectral Indices Observed from Sequential Remote Sensing (Case Study: Tuy Phong, Binh Thuan, Vietnam). *GIScience and Remote Sensing*, Vol. 54, No. 2, pp. 167–184.

Van de Griend, A. A., and M. Owe. 1993. On the Relationship between Thermal Emissivity and the Normalized Difference Vegetation Index for Natural Surfaces. *International Journal of Remote Sensing*, Vol. 14, No. 6, pp. 1119–1131.

Wan, Z., and J. Dozier. 1996. A generalized split-window algorithm for retrieving land-surface temperature from space. *IEEE Transections on Geoscience and Remote Sensing,* Vol. 34(4), pp. 892–905.

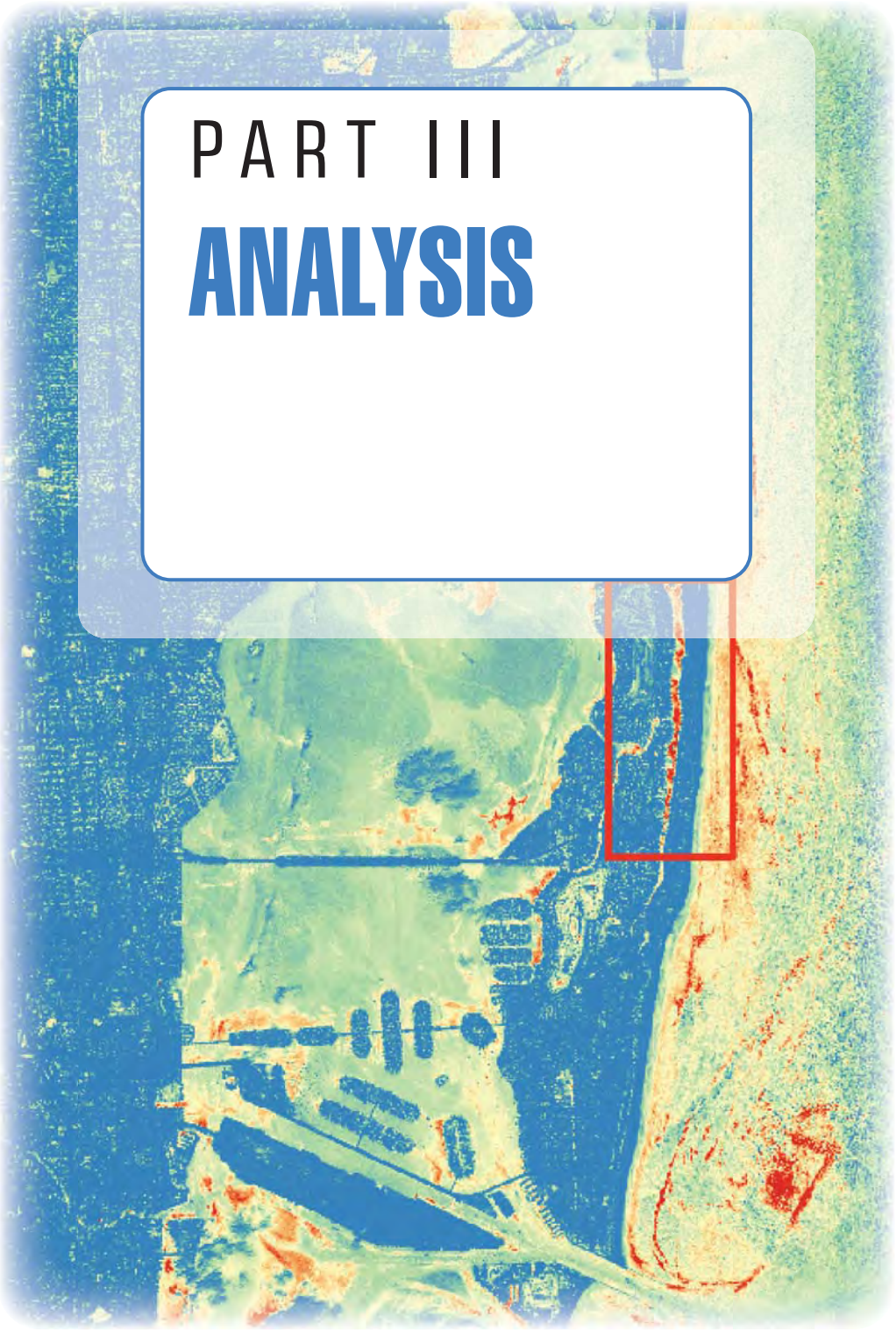
Wang, S., L. He, and W. Hu. 2015. A Temperature and Emissivity Separation Algorithm for Landsat-8 Thermal Infrared Sensor Data. *Remote Sensing,* Vol. 7, No. 8, pp. 9904–9927.

Watson, K., S. Hummer-Miller, and D. L. Sawatzky. 1982. Registration of Heat Capacity Mapping Mission Day and Night Images. *Photogrammetric Engineering and Remote Sensing,* Vol. 48, pp. 263–268.

Weast, R. C. (ed.). 1986. *CRC Handbook of Chemistry and Physics* (68th ed.). Boca Raton, FL: CRC Press, 2440 pp.

Willardson, A. G. 2014. Landsat Thermal Infrared Imagery and Western Water Management. *Journal of Contemporary Water Research and Education,* Vol. 153, pp. 42–48.

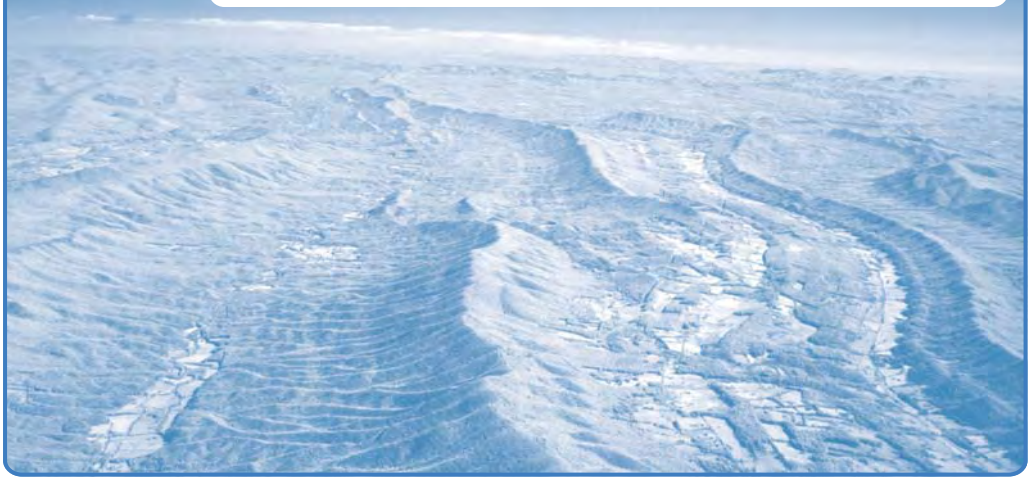
Xue, Y., and A. P. Cracknell. 1995. Advanced Thermal Inertia Modeling. *International Journal of Remote Sensing,* Vol. 16, pp. 431–446.



PART III

ANALYSIS

11 Statistics and Preprocessing



MAJOR TOPICS TO UNDERSTAND

- Image Statistics
- Feature Extraction
- Radiometric Preprocessing
- Geometric Preprocessing
- Image Data Processing Standards

11.1 INTRODUCTION

In the context of digital analysis of remotely sensed data, preprocessing refers to those operations that are preliminary to the principal analysis. Typical preprocessing operations could include (1) radiometric preprocessing to adjust digital values for the effect of a hazy atmosphere, and/or (2) geometric preprocessing to bring an image into registration with a map or another image. Once corrections have been made, the data can then be subjected to the primary analyses described in subsequent chapters. Thus, preprocessing forms a preparatory phase that, in principle, improves image quality as the basis for later analyses that will extract information from the image.

It should be emphasized that, although certain preprocessing procedures are frequently used, there can be no definitive list of "standard" preprocessing steps because each data type or project requires individual attention and some preprocessing decisions may be a matter of personal preference or organizational protocols. Furthermore, the quality of remotely sensed data varies greatly, so some data may not require the preprocessing that would be necessary in other instances. Also, preprocessing changes data. We may assume that such changes are beneficial, but the analyst should remember that preprocessing may create artifacts that are not immediately obvious. As a result, the analyst should tailor preprocessing to the data at hand and the needs of specific projects, using only those preprocessing operations essential to obtain a specific result.

Preprocessing includes a wide range of operations, from the very simple to extremes of abstractness and complexity. Most (at least in the context of spectral data) can be categorized into one of three groups: (1) feature extraction, (2) radiometric corrections, and (3) geometric corrections. Although there are far too many preprocessing methods to discuss in detail here, we will illustrate some of the principles important for each group.

This chapter focuses on the preprocessing of spectral data from passive sensors. Data from lidar sensors, radar sensors, thermal imagers, and mapping cameras require specialized preprocessing that is nonetheless predicated on the same basic principles. It should also be noted that the concept of "analysis-ready data" is increasingly a goal of most public agencies and commercial entities that acquire and distribute remotely sensed data. Thus, radiometric and geometric preprocessing are increasingly incorporated into higher level data products from myriad sensors and programs.

11.2 IMAGE STATISTICS

There are common statistical measures by which remotely sensed data are commonly summarized. These include, for each band, the mean, minimum, maximum, variance, and standard deviation. Among bands, covariance and correlation matrices provide important summary descriptors.

Figure 11.1 shows 3 × 3 images of digital numbers (DNs) varying from 0 to 15 (4-bit radiometric resolution). The summary statistics for each subfigure are as follows: Figure 11.1a, m = 4.0 DN and s^2 = 17.5 DN2; Figure 11.1b, m = 9.0 and s^2 = 7.8 DN2; Figure 11.1c, m = 9.1 DN and s^2 = 11.6 DN2. Note that the lighter blue (higher overall values) in Figures 11.1b and 11.1c are reflected in their means being much higher than the mean of Figure 11.1a. The greater spread of values in Figure 11.1a versus the other two subfigures is reflected in its higher variance.

The mean, m, is defined as follows (using the image in Figure 11.1a as an example):

$$\begin{aligned} m = \bar{x} &= \frac{\sum_{i}^{n} x_i}{n} \\ &= \frac{[12+1+10+3+2+0+1+4+3]}{9} \\ &= \frac{36}{9} \\ &= 4 \end{aligned} \qquad \text{(EQ. 11.1)}$$

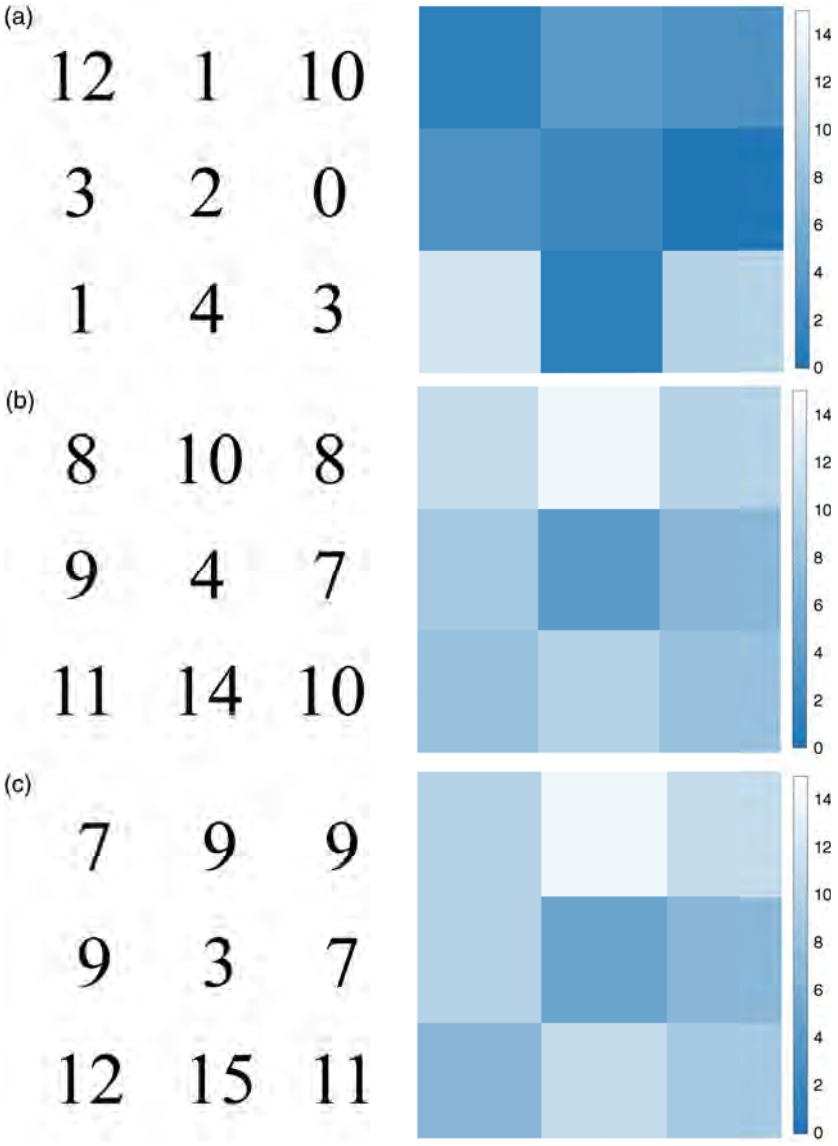


FIGURE 11.1 3 × 3 images of digital numbers (DNs) varying from 0 to 15 (4-bit radiometric resolution) with their corresponding heat maps.

where i is the cell, xi is the value of the cell, and n is the number of cells. (Note that i increments left to right, top to bottom.)

The sample variance, s^2, tells us how spread out our values are, defined as the average of the squared differences from the mean, m, as follows (using the image in Figure 11.1a as an example):

$$
\begin{aligned}
s^2 &= \frac{\sum_{i}^{n}\left(x_i - m\right)^2}{n-1} \\
&= \frac{\left[(12-4)^2 + (1-4)^2 + (10-4)^2 + (3-4)^2 + (2-4)^2 + (0-4)^2 + (1-4)^2 + (4-4)^2 + (3-4)^2\right]}{8} \\
&= \frac{\left[8^2 + (-3)^2 + 6^2 + (-1)^2 + (-2)^2 + (-4)^2 + (-3)^2 + 0^2 + (-1)^2\right]}{8} \\
&\quad \frac{\left[64 + 9 + 36 + 1 + 4 + 16 + 9 + 0 + 1\right]}{8} \\
&= \frac{140}{8} \\
&= 17.5
\end{aligned}
\qquad \text{(EQ. 11.2)}
$$

The square root of the sample variance is the sample standard deviation, s, which has the same units as our original values (or their mean).

These summary statistics are ways of quantitatively describing the distribution of data in a spectral band, but visual representations are often as helpful, or more. One of the most common visual representations of data in a single band is an image histogram, which is a vertical bar chart in which the range of values in a given spectral band is first divided into contiguous bins, usually (but not necessarily) of equal width. The height of each bar is then proportional to the frequency of values (e.g., reflectances) present in each bin. Figure 11.2 is an example of a histogram with just three equal-width bins derived from Figure 11.1a.

One of the most common sets of descriptive statistics, as useful in remote sensing as in any other branch of data analytics, is the *five-number summary*. This consists of the minimum, maximum, median, first quartile, and third quartile. As an example, let's derive each of these from Figure 11.1a. The minimum is the smallest value, 0. The maximum is the largest value, 12. The median is the middle value (when the data are represented as an ordered list), 3 (Equation 11.3), or the mean of the two middle values if n, the number of values, is even. The median is also known as the second quartile, Q2.

$$
\begin{aligned}
\text{median} &= \text{Q2} \\
&= \text{center value in ordered list } \{0, 1, 1, 2, \mathit{3}, 3, 4, 10, 12\} \\
&= 3
\end{aligned}
\qquad \text{(EQ. 11.3)}
$$

There are numerous ways to compute quartiles, and as such, any method we present here will have its proponents and detractors among the statistical cognoscenti. For most remote sensing analysts, the specific method is less important than understanding that the *first quartile* is the value between the first quarter of the data and the remaining three quarters. By extension, the *third quartile* is the value between the first three quarters of

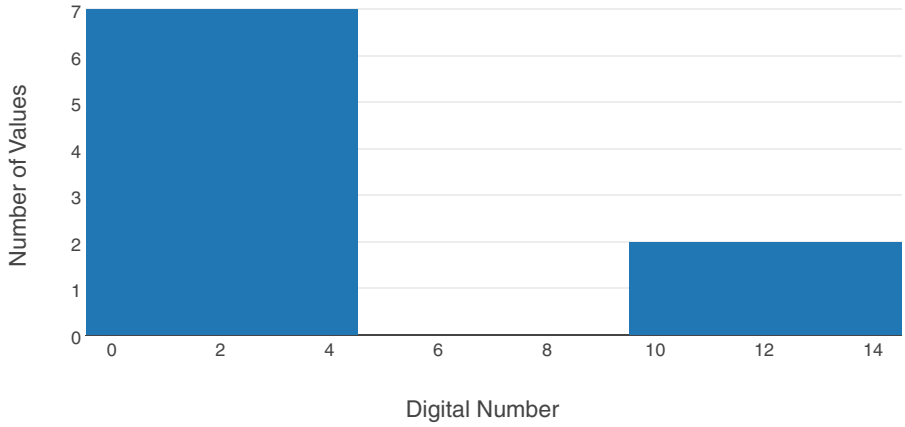


FIGURE 11.2 Histogram of image shown in **Figure 11.1a**.

the data and the last quarter. As such, the following paragraph is only for readers interested in the details of constructing a quartile using one of many common methods, but it is not necessary for a gestalt understanding.

However, with the cognizance that there are multiple ways in which quartiles can be calculated, let's look at a common method using the nine values in the image shown as Figure 11.1a. We first order the values, as shown in Equation 11.3. Since there are an odd number of points, we remove the median and split the data into two halves. (If there were an even number of points we would leave the median and split the dataset into two halves.) The first quartile, Q_1, is the median of the lower half. The third quartile, Q_3, is the median of the upper half. Equations 11.4 and 11.5 show these calculations.

$$\text{first quartile} = Q_1 = \text{median}\,\{0,1,1,2\} = \frac{1+1}{2} = 1 \qquad \text{(EQ. 11.4)}$$

$$\text{third quartile} = Q_3 = \text{median}\,\{3,4,10,12\} = \frac{4+10}{2} = 7 \qquad \text{(EQ. 11.5)}$$

The resulting five-number summary for the image in Figure 11.1a is shown in Table 11.1.

TABLE 11.1 Five-Number Summary from Image Shown in **Figure 11.1a**

	Value
min (Q_0)	0
Q_1	1
Q_2	3
Q_3	7
max (Q_4)	12

We are often interested in how much two variables (e.g., spectral bands, or the same band acquired on different dates) change together. This is called the sample covariance, C, defined as follows with the calculation example using Figures 11.1a and 11.1b:

$$
\begin{aligned}
C &= \frac{\sum_{i}^{n}(x_i - \bar{x})(y_i - \bar{y})}{n-1} \\
&= \frac{\left[\begin{array}{l}(12-4)(8-9)+(1-4)(10-9)+(10-4)(8-9)+\\(3-4)(9-9)+(2-4)(4-9)+(0-4)(7-9)+\\(1-4)(11-9)+(4-4)(14-9)+(3-4)(10-9)\end{array}\right]}{8} \\
&= \frac{\left[\begin{array}{l}8(-1)+(-3)(1)+(6)(-1)+(-1)(0)+\\(-2)(-5)+(-4)(-2)+(-3)(2)+0(5)+(-1)(1)\end{array}\right]}{8} \\
&= \frac{[-8-3-6+0+10+8-6+0-1]}{8} \\
&= \frac{-6}{8} \\
&= -0.75
\end{aligned}
\qquad \text{(EQ. 11.6)}
$$

Since this value is so low compared to the variance of a given image (e.g., 17.5 DN2 for the first image and 7.75 DN2 for the second image), change in one does not appear related to change in the other, reflected in the very low covariance of –0.75 DN2. In contrast, however, the values in Figures 11.1b and 11.1c do appear to vary together, and this is reflected in a higher covariance, 9.3 DN2.

It is possible, if bands are not in the same scale (e.g., for DNs with different gains and biases), to calculate interband correlation, r (which is just the sample covariance divided by the product of the two sample standard deviations), affording standardization. However, as pointed out by Mather and Koch (2011), such standardization is inherently "undesirable" when the units and measurement scale are the same (such as with radiance or reflectance, as discussed below), as it "removes the effects of changes in the degree of variability between the bands." However, many analysts nonetheless find correlation easier to interpret than covariance, as values always range from –1 to 1. This indicates both the strength and direction of the linear relationship between two spectral bands, with absolute values close to zero having very little (to no) relationship and absolute values close to one having a very strong relationship. With respect to direction, correlation is positive when the band values (DN, radiance, reflectance, etc.) increase together, and negative when one value decreases as the other increases. The following two tables show the covariance matrix (Table 11.2) and correlation matrix (Table 11.3) for the 3 × 3 images shown in Figure 11.1.

A scatter plot is a plot of the values of one spectral band (or other variable) against another spectral band (or other variable). The scatter plots associated with the images shown in Figure 11.1 are shown in Figure 11.3. Note the differences in appearance between high and low interband correlation.

The First Law of Geography states that "everything is related to everything else, but near things are more related than distant things" (Tobler, 1970). While "near" is typi-

TABLE 11.2 Covariance Matrix from Images Shown in **Figure 11.1**

	11.1a	11.1b	11.1c
11.1a	17.5	−0.75	−0.88
11.1b	−0.75	7.75	9.25
11.1c	−0.88	9.25	11.61

TABLE 11.3 Correlation Matrix from Images Shown in **Figure 11.1**

	11.1a	11.1b	11.1c
11.1a	1.0	−0.06	−0.06
11.1b	−0.06	1.0	0.98
11.1c	−0.06	0.98	1.0

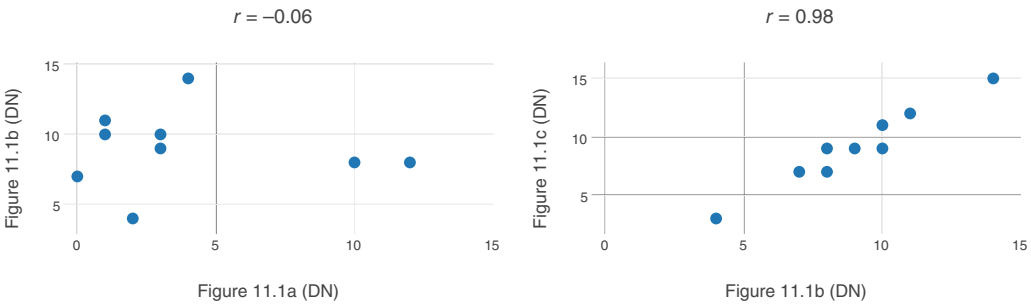


FIGURE 11.3 Scatter plots associated with the images shown in **Figure 11.1**.

cally construed as spatial proximity, temporal or spectral proximity also typically has the same effect (autocorrelation). To illustrate this from a spectral standpoint, let's take a look at a practical illustration of the concepts in this section using descriptive statistics from a spatial and spectral subset of a Landsat 8 Operational Land Imager (OLI) scene (WRS-2 row 17 path 34) acquired May 11, 2015 (Figure 11.4).

As we can see from Tables 11.4 and 11.5, spectrally proximate wavelength bands (e.g., blue and green with $r = 0.96$) are more correlated than those that are less spectrally proximate (e.g., blue and near infrared [NIR] with $r = -0.30$). As spectral resolution increases (e.g., imaging spectroscopy data), this general trend is exacerbated.

11.3 FEATURE EXTRACTION

In the context of image processing, the term *feature extraction* (or *feature selection*) has specialized meaning. "Features" are not geographical features, visible on an image; rather, they are "statistical" characteristics of image data—individual bands or combinations of band values that carry information concerning systematic variation within the scene.

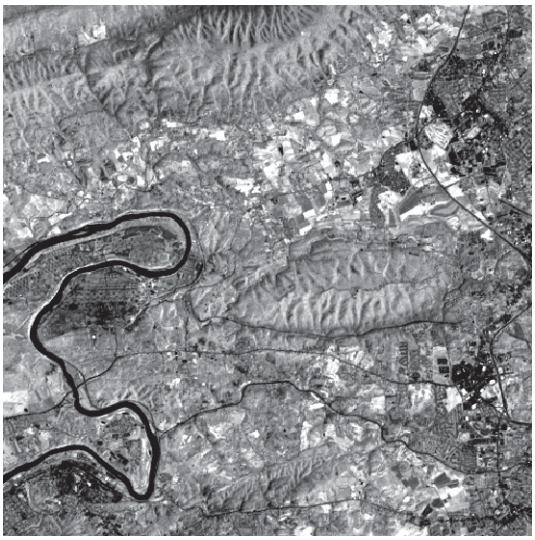


FIGURE 11.4 Near-infrared (NIR) band (0.845–0.885 μm, surface reflectance) from a subset of a Landsat 8 OLI scene (WRS-2 row 17 path 34) acquired May 11, 2015 (Blacksburg, Virginia, and vicinity).

TABLE 11.4 Covariance Matrix from a Six-Band Subset of a Landsat OLI Image Acquired May 11, 2015

	Blue	Green	Red	NIR	SWIR1	SWIR2
Blue	707704	724592	1105252	–817543	949071	1106350
Green	724592	802040	1179270	–424747	1212915	1232224
Red	1105252	1179270	1867404	–1247069	1855659	1998525
NIR	–817543	–424747	–1247069	10257846	1673930	–652673
SWIR1	949071	1212915	1855659	1673930	3810167	2851772
SWIR2	1106350	1232224	1998525	–652673	2851772	2751335

TABLE 11.5 Correlation Matrix from a Six-Band Subset of a Landsat OLI Image Acquired May 11, 2015

	Blue	Green	Red	NIR	SWIR1	SWIR2
Blue	1	0.96	0.96	–0.30	0.58	0.79
Green	0.96	1	0.96	–0.15	0.69	0.83
Red	0.96	0.96	1	–0.28	0.7	0.88
NIR	–0.30	–0.15	–0.28	1	0.27	–0.12
SWIR1	0.58	0.69	0.7	0.27	1	0.88
SWIR2	0.79	0.83	0.88	–0.12	0.88	1

Thus, feature extraction could also be known as "information extraction," isolation of components within spectral data that are most useful in portraying the essential elements of an image, its "hidden structure" (Schlens, 2014). In theory, discarded data contain noise and errors present in original data. Thus, feature extraction may increase accuracy. In addition, feature extraction reduces the number of spectral channels, or bands, that must be analyzed, thereby reducing computational demands. After feature selection is complete, the analyst works with fewer but more potent channels. The reduced data set may convey almost as much information as does the complete data set. Feature selection may increase speed and reduce the costs of analysis.

Multispectral data, by their nature, consist of multiple channels of data. Although some images may have as few as 3 or 4 channels (Chapter 6), other image data may have many more, possibly 200 or more channels (imaging spectroscopy, Chapter 14). Data volume is further increased with the routine analysis of multitemporal data. With so much data, processing of even modest-sized images requires considerable time. In this context, feature selection assumes considerable practical significance, as image analysts wish to reduce the amount of data while retaining effectiveness and/or accuracy.

High correlation between pairs of bands (see parts **b** and **c** of **Figure 11.1**; **Figure 11.3**, right) means that the values in the two channels are closely related. Feature selection attempts to identify, then remove, such duplication so that the data set can include maximum information using the minimum number of channels.

For example, for data represented by **Table 11.5,** the blue, NIR, and SWIR1 bands might include almost as much information as the entire set of six channels because the blue band is closely related to the red and green, the two SWIR bands are highly correlated ($r = 0.88$), and the NIR band carries information largely unrelated to any others. Therefore, each of the discarded channels (green, red, and SWIR2) resembles one of the channels that has been retained. So a simple approach to feature selection discards unneeded bands, thereby reducing the number of channels. Although this kind of selection can be used as a kind of rudimentary feature extraction (particularly when coupled with analysis-specific variable selection), feature selection is typically a more complex process based on statistical interrelationships among channels.

Principal Components Analysis

A powerful approach to feature selection applies a method of data analysis called principal components analysis (PCA). This presentation offers only a superficial description of PCA. More complete explanation requires the level of detail provided by Davis (2002), Richards (2013), and Schlens (2014). In essence, the axes (variables) in the original multidimensional feature space (the multidimensional scatter plot) are transformed such that (1) there is no remaining covariance among the variables (i.e., spectral bands) and (2) each principal component is oriented to capture as much of the variability that remains in a serial (and successively orthogonal) fashion. These new axes form the new *basis* for the data contained in the original reflectance vectors from each pixel.

In matrix notation (and this paragraph is not essential for a conceptual understanding of PCA), the new basis is a linear transformation of the original basis (Schlens, 2003) of the form

$$\mathbf{Y} = \mathbf{PX} \qquad \text{(EQ. 11.7)}$$

where **Y** is the transformed (in this case, decorrelated) image data set with k rows × b columns, where k is the number of pixels in the image and b is the number of bands in the image), **X** is the original data set (also $k \times b$), and **P** is the linear transformation matrix (TM; in this case, the *eigen matrix*) that transforms **X** into **Y** via a rotation and stretch ($b \times b$). The rows of **P** are a set of new orthogonal basis vectors (in this case, the *eigen vectors*) for expressing the columns of **X**. When $k = 1$, a new PC vector is calculated from the original reflectance vector using the eigen matrix. Since the eigen matrix is the linear transformation that eliminates all covariance, it varies from image to image.

The effect is most easily seen using only two bands. Figure 11.5 is a scatter plot of the blue and green bands from the image shown in Figure 11.4. The high correlation between the bands (0.96) is evident. Figure 11.6 shows the same data subsequent to the PCA, with the covariance matrices before and after the transformation shown as Table

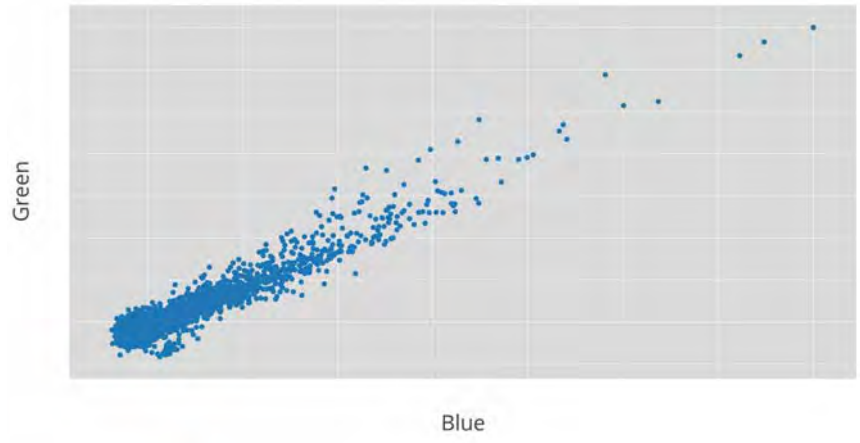


FIGURE 11.5 Scatter plot of the blue and green bands.

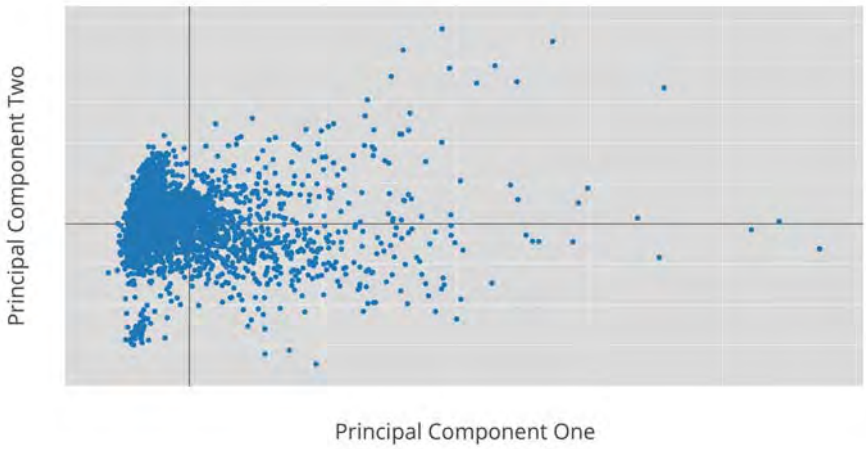


FIGURE 11.6 Scatter plot of the first two principal component bands.

11.6 and Table 11.7. Note that the nondiagonal elements of the covariance matrix are now zero. Tables 11.4 and 11.8 show the effects of the transformation on covariance for a six-band subset of the same date.

The total variance is the sum of the individual PC band variances (also known as the eigen values). For the two-band example (Table 11.7), this is the sum of the diagonal elements (the matrix *trace*): 1,480,998 + 28,746 = 1,509,744. The percentage variance explained by each principal component can be found by dividing each variance by the total. So, the percentage variance explained by principal component 1 is 1,480,998 / 1,509,744 = 98%. We are also commonly interested in the percent variance explained by the first *n* principal components. This is easily found by first finding the cumulative variance and then dividing by the sum of the eigen values (the trace of the principal component image covariance matrix). So, the percent variance explained by the first three principal components shown in Table 11.8 is as follows:

$$\frac{10{,}843{,}409 + 8{,}532{,}934 + 667{,}807}{10{,}843{,}409 + 8{,}532{,}934 + 667{,}807 + 109{,}525 + 27{,}080 + 15{,}741} = 99.2\% \quad \text{(EQ. 11.8)}$$

TABLE 11.6 Covariance Matrix from a Two-Band Subset of a Landsat OLI Image Acquired May 11, 2015

	Blue	Green
Blue	707704	724592
Green	724592	802040

TABLE 11.7 Covariance Matrix from the Principal Components Computed Using a Two-Band Subset of a Landsat OLI Image Acquired May 11, 2015

	PC1	PC2
PC1	1480998	0
PC2	0	28746

TABLE 11.8 Covariance Matrix from the Principal Components Computed Using a Six-Band Subset of a Landsat OLI Image Acquired May 11, 2015

	PC1	PC2	PC3	PC4	PC5	PC6
PC1	10843409	0	0	0	0	0
PC2	0	8532934	0	0	0	0
PC3	0	0	667807	0	0	0
PC4	0	0	0	109525	0	0
PC5	0	0	0	0	27080	0
PC6	0	0	0	0	0	15741

For data with similar characteristics (e.g., signal-to-noise ratio), the number of principal component bands needed to explain a given amount of variance (often 99%) is a way to understand the *inherent dimensionality* (a way to express the information content) of an image.

Let's use Figure 11.7 to illustrate this concept with multitemporal imagery of the same area. Using 99% as the cumulative variance threshold, the inherent dimensionality of one OLI image is 3 (bands), and that of a six-date stack is 12. As such, while we went from 6 multispectral bands from one date to 36 spectral bands from six dates, a sixfold increase, we only increased the inherent dimensionality four times (3 vs. 12). Furthermore, of course, since the first 12 PC bands represent almost all the variance, we can use the 12 PC bands instead of the 36 original spectral bands for many applications, reducing data volume and enhancing the speed (and often quality) of subsequent analyses.

Figure 11.8 shows transformed data for a subset of a Landsat TM scene. Images PC I and PC II are the most potent; PC III, PC IV, PC V, and PC VI show the decline in information content, such that the images for the higher PCs (such as PCs V and VI) record artifacts of system noise, atmospheric scatter, topographic shadowing, and other undesirable contributions to image brightness. If such components are excluded from subsequent analysis, it is likely that accuracy can be retained (relative to the entire set of six channels) while also reducing time and cost devoted to the analysis. A color presentation of the first three components, assigning each to one of the additive primaries (red, green, and blue), is usually effective in presenting a concise, potent portrayal of the information conveyed by a multichannel multispectral image. Note, however, that because each band is a linear combination of the original channels, the analyst must be prepared to interpret the meaning of the new channels. In some instances, this task is relatively straightforward; in other instances, it can be difficult to unravel the meaning of a PCA image. Furthermore, the PCA transformation applies only to the specific image at hand, and each new image

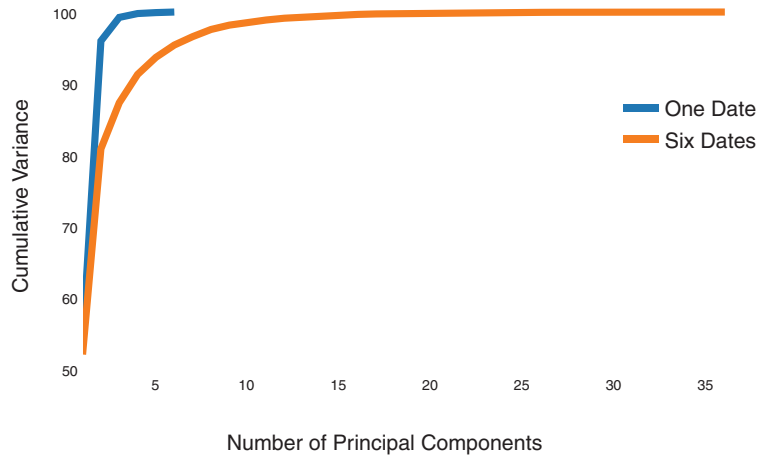


FIGURE 11.7 Cumulative percent variance from principal components derived from one Landsat OLI scene versus a stack of six Landsat OLI scenes (WRS-2 row 17 path 34) of the same area over the course of just over a year. The blue, green, red, NIR, SWIR1, and SWIR2 bands are used for each date. The dates are as follows: March 21, 2014; May 8, 2014; May 24, 2014; January 19, 2015; February 4, 2015; May 11, 2015 (last date used for the one scene cumulative variance calculation).

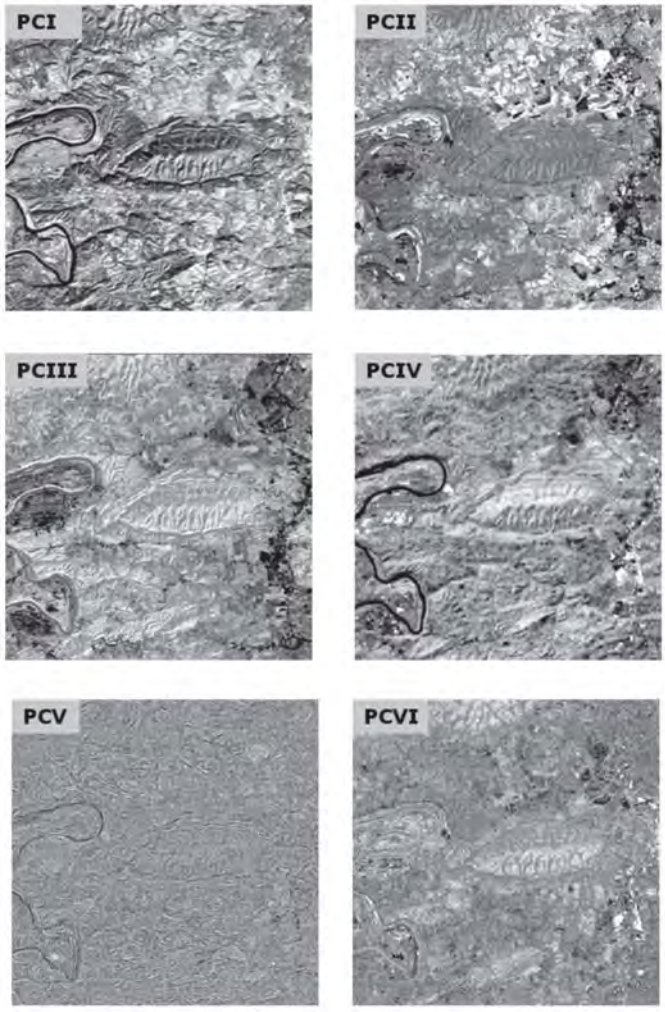


FIGURE 11.8 Feature selection by PCA. The first principal component image (PC I), formed from a linear combination of data from all seven original bands, accounts for over 80% of the total variation of the image data. PC II and PC III present about 10% and 5% of the total variation, respectively. The higher components (e.g., PC V and PC VI) account for very low proportions of the total variation and convey mainly noise and error, as is clear by the image patterns they show.

requires a new recalculation of the PCA. For some applications, this constraint limits the effectiveness of the technique.

Tasseled Cap Transformation

The "tasseled cap" (TC) transformation (Kauth and Thomas, 1976) is a linear transformation that projects soil and vegetation information into a single plane in multispectral data space—a plane in which the major spectral components of an agricultural scene are

displayed in two dimensions (the *plane of vegetation,* Cohen and Goward, 2004; see also Figure 11.9). Algebraically, it, like PCA, is just a change of basis (Equation 11.7), but with a tranformation matrix (**P**) that is constant for a given sensor (e.g., Tables 11.9 and 11.10). Though defined initially for Multispectral Scanner System (MSS) data, subsequent research (Crist and Cicone, 1984) has extended the concept to the six nonthermal bands of the TM and later, to the OLI (Baig et al., 2014; Zhai et al., 2022). The transformation can be visualized as a rotation of a solid multidimensional figure (the feature space representing all spectral bands) in a manner that permits the analyst to view the major spectral components of an agricultural scene as three two-dimensional figures (Figures 11.9, 11.10, and 11.11).

The transformation consists of a linear combination of the original spectral channels to produce a set of new variables (Equation 11.7), each describing a specific dimension of the agricultural scene. Multiplication of the sensor-specific transformation matrix by a reflectance vector from a given OLI pixel results in the TC transformation values for that pixel, as shown in Equation 11.9.

$$\begin{bmatrix} \text{brightness} \\ \text{greenness} \\ \text{wetness} \end{bmatrix} = \begin{bmatrix} 0.4596 & 0.5046 & 0.5458 & 0.4114 & 0.2589 \\ -0.3374 & -0.4901 & 0.7909 & 0.0177 & -0.1416 \\ 0.2254 & 0.3681 & 0.2250 & -0.6053 & -0.6298 \end{bmatrix} \begin{bmatrix} \rho_{\text{green}} \\ \rho_{\text{red}} \\ \rho_{\text{NIR}} \\ \rho_{\text{SWIR1}} \\ \rho_{\text{SWIR2}} \end{bmatrix} \quad \text{(EQ. 11.9)}$$

The coefficients are calculated by means of an iterative procedure (Jackson, 1983). The transformation is based on the calibration information for each specific sensor and so requires a dedicated effort to derive the values for a specific sensor. Once the values are available, they can be applied to imagery from that sensor.

TABLE 11.9 Tasseled Cap Transformation Matrix for OLI Surface Reflectance

	Component	Green	Red	NIR	SWIR1	SWIR2
5-band	Brightness	0.4596	0.5046	0.5458	0.4114	0.2589
	Greenness	−0.3374	−0.4901	0.7909	0.0177	−0.1416
	Wetness	0.2254	0.3681	0.2250	−0.6053	−0.6298

Note: From Zhai et al. (2022). Used by permission.

TABLE 11.10 Tasseled Cap Transformation Matrix for OLI Top-of-Atmosphere Reflectance

	Component	Green	Red	NIR	SWIR1	SWIR2
5-band	Brightness	0.4321	0.4971	0.5695	0.4192	0.2569
	Greenness	−0.3318	−0.4844	0.7856	−0.0331	−0.1923
	Wetness	0.2633	0.3945	0.1801	−0.6121	−0.6066

Note: From Zhai et al. (2022). Used by permission.

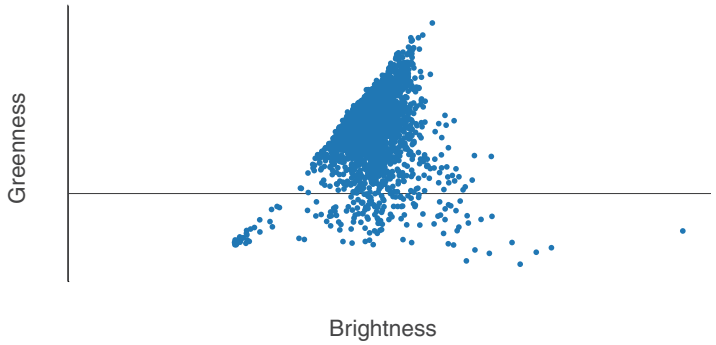


FIGURE 11.9 Tasseled cap plane of vegetation calculated using a subset of an OLI scene (WRS-2 row 17 path 34) acquired May 11, 2015.

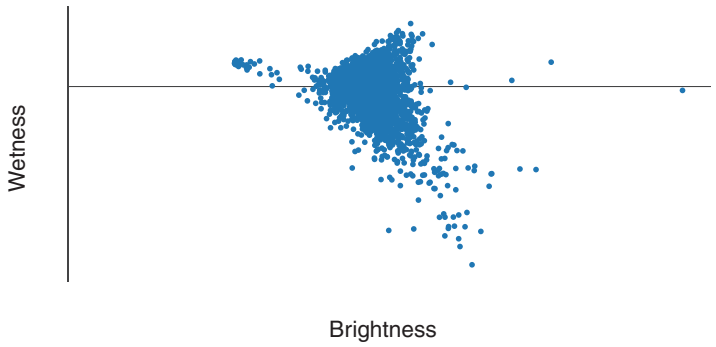


FIGURE 11.10 Tasseled cap plane of soils calculated using a subset of an OLI scene (WRS-2 row 17 path 34) acquired May 11, 2015.

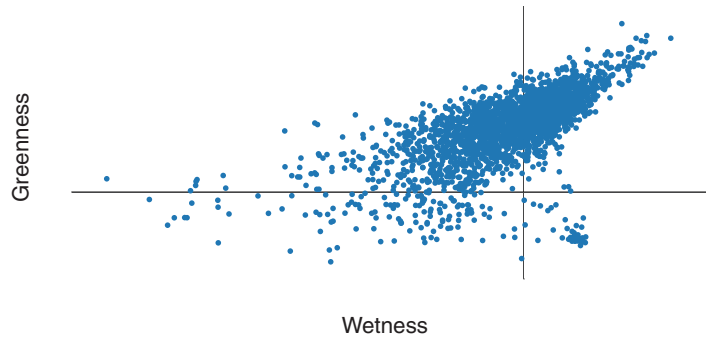


FIGURE 11.11 Tasseled cap transition zone calculated using a subset of an OLI scene (WRS-2 row 17 path 34) acquired May 11, 2015.

Although the TC bands no longer match directly to observable spectral bands, they do carry specific information concerning agricultural scenes. Crist and Cicone (1984) interpret TC1 as brightness, a weighted sum of all five or six input bands. TC2 is designated as greenness, a band that conveys information concerning the abundance and vigor of living vegetation. TC3, wetness (only available in the post-MSS era), contrasts "the SWIR bands against the visible and NIR bands in an effort to express the water content of soils" (Cohen and Goward, 2004) and moisture in plant tissues.

The first two bands (brightness and greenness) usually convey almost all the information in an agricultural scene—often 95% or more. Therefore, the essential components of an agricultural landscape are conveyed by a two-dimensional diagram, the plane of vegetation, using brightness and greenness (Figures 11.9 and 11.12).

Over the interval of an entire growing season, brightness and greenness values for a specific field follow a stereotyped trajectory (Figure 11.12). Initially, the spectral response of a field is dominated by soil, as the field is plowed, disked, and planted (point a in Figure 11.12). The field has a position near the soil brightness line. As the crop emerges (b) and grows, it simultaneously increases in greenness and decreases in soil brightness as the green canopy covers more and more of the soil surface (b to c). Then, as senescence, maturity, and harvest occur, the field decreases in greenness and increases in soil brightness to return the field to near its original position on the diagram (c to d).

Figure 11.13 illustrates the application of the Landsat 9 Operational Land Imager–2 (OLI-2) tasseled cap model to OLI-2 imagery of analysis-ready surface reflectance data acquired in February 2022 over Jefferson County, Alabama, showing the original image, brightness, greenness, and wetness.

As noted earlier, the principal components transformation must be calculated (and re-interpreted) individually for each new image. In contrast, the TC coefficients, once calculated for a specific sensor, apply to all images acquired by that sensor; the meanings of the resulting TC components are thus consistent across images. As a result, the TC transformation has been used for understanding not only images of agricultural scenes, but also a variety of other ecological settings.

11.4 RADIOMETRIC PREPROCESSING

Many preprocessing operations fall into the category of image restoration (Estes et al., 1983), the effort to remove the undesirable influence of atmospheric interference, system noise, and sensor motion. By applying knowledge of the nature of these effects, it is possible to estimate their magnitude, then to remove or minimize their influence upon the data used in later steps of the analysis. After removing these effects, the data are said

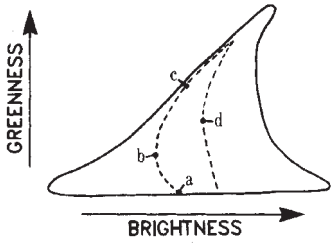


FIGURE 11.12 Seasonal variation of a field in data space defined by the greenness and brightness axes. Point descriptions: (a) bare soil, (b) greening up, (c) full canopy closure, (d) senescence. From Crist and Cicone (1984). Copyright © 1984 IEEE. Used by permission.

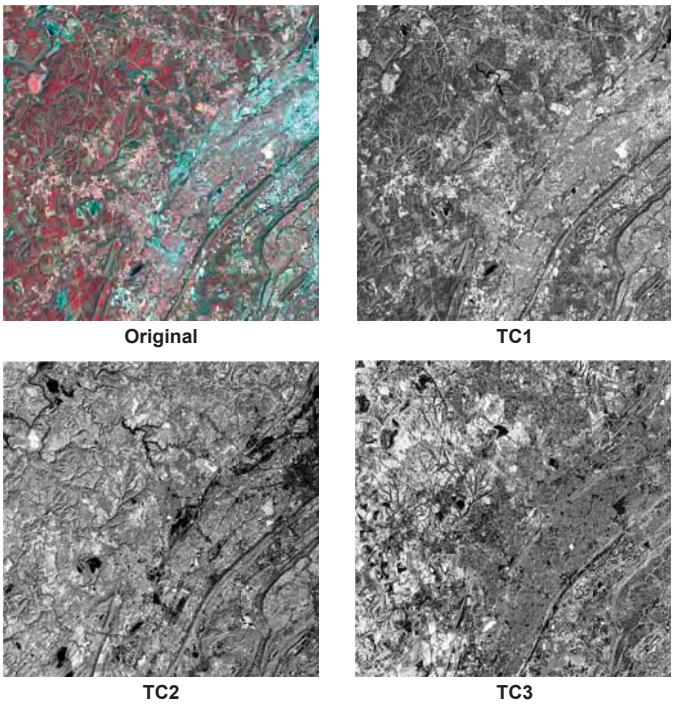


FIGURE 11.13 The tasseled cap transformation applied to a Landsat 9 Operational Land Imager–2 image of Jefferson County, Alabama. Top left shows the standard false color composite; then, in sequence, tasseled cap band (TC) 1, brightness; TC2, greenness; and TC3, wetness.

to be "restored" to their (hypothetical) correct condition, although we can, of course, never know what the correct values might be; we must always remember that attempts to correct data may themselves introduce errors. So the analyst must decide if the errors removed are likely to be greater than those that might be introduced. Typically, image restoration includes efforts to correct for both radiometric and geometric errors.

Recollect from Chapter 5 that the discrete digital values recorded by the sensor are known as digital numbers (DNs) or brightness values (BVs). Most analysts, however, prefer to work with reflectance, which is the ratio of reflected to incident energy (Chapter 2). Two forms of reflectance data are commonly used. *Top-of-atmosphere (TOA) reflectance* corrects for differential (1) sensor- or band-specific sensitivity and (2) solar illumination (for passive optical sensors), but assumes there is no need for atmospheric correction. *Bottom-of-atmosphere (BOA)* or, equivalently, *surface reflectance* (SR), has (at least nominally) removed all atmospheric effects as well.

As a simple example of why atmospheric corrections are necessary, remember that any sensor that observes the Earth's surface using visible or near-visible radiation will record a mixture of two kinds of brightnesses. One brightness is due to the reflectance from the Earth's surface—the brightnesses that are of interest for remote sensing. But the sensor also observes the brightness of the atmosphere itself—the effects of scattering (path radiance; see Chapter 2). Thus, an observed digital brightness value (e.g., "56") might be in part the result of surface reflectance (e.g., "45") and in part the result of

atmospheric scattering (e.g., "11"). Of course we cannot immediately distinguish the two brightnesses, so one objective of atmospheric correction is to identify and separate these two components so that the main analysis can focus on examination of correct surface brightness (the "45" in this example). Ideally, atmospheric correction should find a separate correction for each pixel in the scene. In practice, we may apply the same correction to an entire band, or apply a single factor to a local region within the image.

Preprocessing operations to correct for atmospheric degradation fall into two rather broad categories. First are those procedures based on efforts to model the physical behavior of the radiation as it passes through the atmosphere. Application of such models permits digital numbers to be adjusted to approximate true values that might be observed under a clear atmosphere, thereby improving image quality and analysis accuracy. Physical models (i.e., models that attempt to model the physical process of scattering at the level of individual particles and molecules) have important advantages with respect to rigor, accuracy, and applicability to a wide variety of circumstances. But they also have significant disadvantages. Often they are very complex, usually requiring detailed data and intricate computer programs. An important limitation is the requirement for detailed meteorological information pertaining to atmospheric humidity and the concentrations of atmospheric particles. Such data may be difficult to obtain in the necessary detail and may apply to only a few points within a scene. Also, atmospheric conditions vary with altitude and over space. Although meteorological satellites, as well as a growing number of remote sensing systems, collect atmospheric data that can contribute to atmospheric corrections of imagery, procedures for local implementation of such methods require a substantial investment on the part of the analyst and (if not standardized) preclude data normalization across space and time.

Fortunately, however, image providers are now cognizant of the value of analysis-ready imagery, so they increasingly provide it as a higher-order data product. For example, TOA reflectances are now provided as the default by the European Space Agency for Sentinel-2 MultiSpectral Instrument (MSI) (levels 1B and 1C) as well as by the U.S. Geological Survey (USGS) for Landsat (level 1 products with user-applied rescaling; see **Figures 11.14 and 11.15**). The two programs differ in their approach to surface reflectance. To date, the Landsat program calculates surface reflectance using the radiative transfer model 6S (Vermote et al., 1997), a product for direct download, whereas the Sentinel program requires users to calculate surface reflectance themselves using the Sentinel-2 Toolbox. The Sentinel-2 MSI algorithm also uses a different radiative transfer model (ATCOR/LIBRADTRAN; Mayer and Kylling, 2005; Richter and Schlaepfer, 2011).

A second approach to atmospheric correction of remotely sensed imagery is based on examination of reflectances from objects of known or assumed brightness recorded by multispectral imagery. From basic principles of atmospheric scattering, we know that scattering is related to wavelength, sizes of atmospheric particles, and their abundance. If a known target is observed using a set of multispectral measurements, the relationships between values in the separate bands can help assess atmospheric effects.

Ideally, the target consists of a natural or human-made feature that can be observed with airborne or ground-based instruments at the time of image acquisition, so the analyst could learn from measurements independent of the true brightness of the object when the image was acquired. However, in practice we seldom have such measurements; therefore, we must look for features of known brightness that commonly, or fortuitously, appear within an image. In its simplest form, this strategy can be implemented by identifying a very dark object or feature within the scene. Such an object might be a large

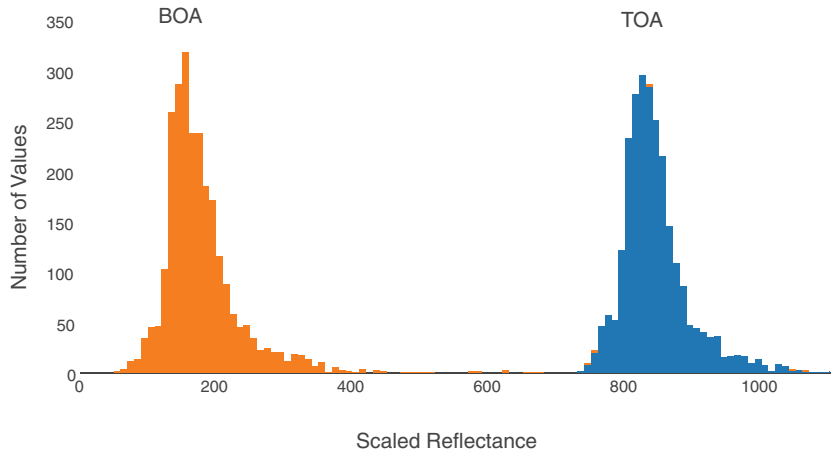


FIGURE 11.14 Histogram from the blue band of a subset of a Landsat 8 scene acquired September 29, 2015, WRS-2 path 28, row 27, showing the difference between top-of-atmosphere and bottom-of-atmosphere reflectance.

water body or possibly shadows cast by clouds or by large topographic features. In the infrared portion of the spectrum, both water bodies and shadows should have brightness at or very near zero because clear water absorbs strongly in the near-infrared spectrum and because very little infrared energy is scattered to the sensor from shadowed pixels. Analysts who examine such areas, or the histograms of the digital values for a scene, can observe that the lowest values (for dark areas, such as clear water bodies) are not zero but some larger value. Typically, this value will differ from one band to the next. For

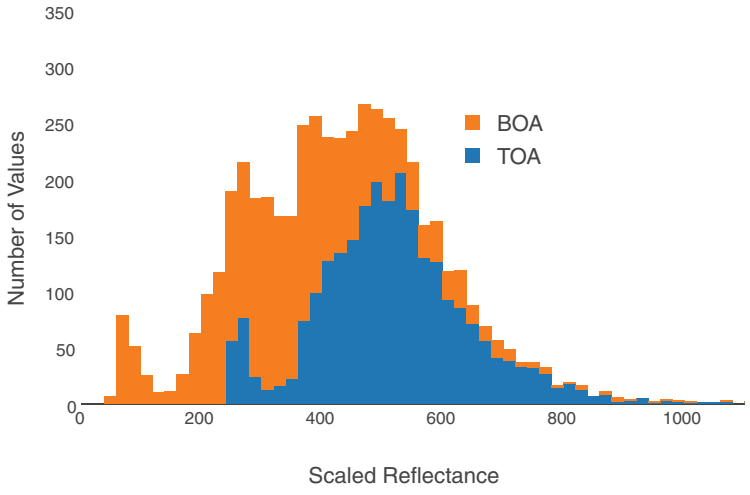


FIGURE 11.15 Histogram from the red band from the same scene. Note that reduced scattering in the red leads to less need for (and results from) atmospheric correction.

example, for Landsat, the values might be 12 for band 1, 7 for band 2, 2 for band 3, and 2 for band 4. These values, assumed to represent the value contributed by atmospheric scattering for each band, are then subtracted from all digital values for that scene and that band. Thus, the lowest value in each band is set to zero, the dark black color assumed to be the correct tone for a dark object in the absence of atmospheric scattering. This procedure forms one of the simplest, most direct methods for adjusting digital values for atmospheric degradation (Chavez, 1975), known sometimes as the histogram minimum method, or the dark object subtraction technique.

This procedure has the advantages of simplicity, directness, and almost universal applicability, as it exploits information present within the image itself. Yet it must be considered as an approximation; atmospheric effects change not only the position of the histogram on the axis, but also its shape (i.e., not all brightnesses are affected equally). (Chapter 2 explained that the atmosphere can cause dark pixels to become brighter and bright pixels to become darker, so application of a single correction to all pixels will provide only a rough adjustment for atmospheric effects.) In addition, in arid regions observed at high Sun angles, shadows, clouds, and open water may be so rare or of such small areal extent that the method cannot be applied.

There are numerous other approaches to atmospheric correction, but as we noted earlier, their need is increasingly obviated by the use of ever more sophisticated physically based models by image providers. Newer sensor designs incorporate spectral bands that provide pixel-specific information on radiative transfer model parameters.

How can the analyst decide if additional atmospheric corrections are necessary beyond those implemented by the image provider? This may be a difficult decision, as the effects of atmospheric degradation are not always immediately obvious from casual inspection. The analyst should always examine summary statistics for each scene (see Section 11.2), inspecting means, variances, covariances, and frequency histograms for suggestions of poor image quality and the absence of dark values, especially if the image is known to show large water bodies.

Of course, inspection of the image may reveal evidence suggesting a requirement for correction. Loss of resolution and low contrast may indicate poor atmospheric conditions. Sometimes the image date may itself suggest the nature of atmospheric quality. In the central United States, summer dates often imply high humidity, haze, and poor visibility, whereas autumn, winter, and spring dates are often characterized by clearer atmospheric conditions. Thus, the image date may provoke further investigation by the analyst to determine if corrections are necessary. Finally, the analyst should examine summary statistics for the scene, especially the frequency histograms for each band.

Fundamentally, however, an analyst will rarely use an image at only one point in time. More commonly, multiple images through time are used. It is the very ubiquity of multitemporal analysis that has led, in part, to organizational mandates to provide calibrated, analysis-ready data with the highest radiometric fidelity.

11.5 GEOMETRIC PREPROCESSING

A critical consideration in the application of remote sensing is preparation of planimetrically correct versions of aerial and satellite images so that they will match to other imagery and to maps and will provide the basis for accurate measurements of distance and

area. Previously, in Chapter 4, we learned that aerial imagery has inherent positional errors that prevent use of the raw imagery for positional measurements. Chapter 4 also introduced the orthophoto, which is a planimetrically correct version of an aerial (or satellite) image created by analysis of stereo imagery through removal of positional effects of topographic relief to produce a positionally accurate image. Similar products can be prepared by applying precise knowledge of the internal geometry of the instrument and of the sensor's position in space relative to the terrain to derive planimetrically correct versions of an aerial or satellite image. Because such images are prepared by applying basic optical principles and details of the instrument calibration, they form the preferred standard for positionally correct images. As such, they are increasingly implemented by image providers by default.

A second approach to image registration, known as *image resampling*, approaches the problem in a completely different manner. No effort is made to apply our knowledge of system geometry. Instead, the images are treated simply as an array of values that must be manipulated to create another array with the desired geometry. Resampling scales, rotates, translates, and performs related manipulations as necessary to bring the geometry of an image to match a particular reference image of desired properties. Such operations can be seen essentially as an interpolation problem similar to those routinely considered in cartography and related disciplines. Such processes constitute the practice of image resampling—the application of interpolation to bring an image into registration with another image or a planimetrically correct map. Image resampling forms a convenient alternative to the analytical approach, as it does not require the detailed data describing the instrument and its operation (which may not be at hand). Although resampling may provide useful representations of images, users should recognize that resampled images are not equivalent to orthographic representations but are produced by arbitrary transformations that bring a given image into registration with another map or image. Although resampling can apply to vector data, our discussion here refers to raster images.

Resampling is related to but distinct from *georeferencing*. Georeferencing indicates that resampling of an image requires matching not only to a reference image but also to reference points that correspond to specific known locations on the ground. Georeferenced images are presented in a specific geographic projection so that the image is presented in a defined projection and coordinate system. Georeferencing is, of course, important for images that are to be used as locational references or to be matched to other maps and images. Although our discussion here focuses on the resampling process, in most image processing systems resampling and georeferencing are usually part of a single process.

In Figure 11.16, the input image is represented as an array of open dots, each representing the center of a pixel in the uncorrected image. Superimposed over this image is a second array, symbolized by solid dots, which shows the centers of pixels in the image transformed (as described below) to have the desired geometric properties (the "output" image).

The locations of the output pixels are derived from locational information provided by *ground control points* (GCPs), places on the input image that can be located with precision on the ground and on planimetrically correct maps. (If two images are to be registered, GCPs must be easily recognized on both images.) The locations of these points establish the geometry of the output image and its relationship to the input image. Thus,

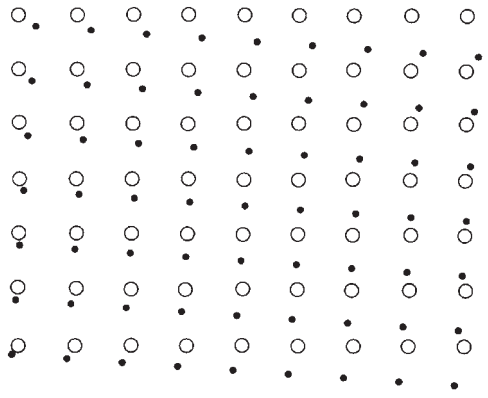


FIGURE 11.16 Resampling. Open circles (O) represent the reference grid of known values in the input image. Black dots (•) represent the regular grid of points to be estimated to form the output image. Each resampling method employs a different strategy to estimate values at the output grid, given known values for the input grid.

this first step establishes the framework of pixel positions for the output image using the GCPs.

The next step is to decide how best to estimate the values of pixels in the corrected image, based on information in the uncorrected image. The simplest strategy from a computational perspective is simply to assign each "corrected" pixel the value from the nearest "uncorrected" pixel. This is the nearest-neighbor approach to resampling (**Figure 11.17**). It has the advantages of simplicity and the ability to preserve the original values of the unaltered scene—an advantage that may be critical in some applications. The nearest-neighbor method is considered the most computationally efficient of the methods usually applied for resampling. On the other hand, it may create noticeable positional errors, which may be severe in linear features where the realignment of pixels may be noticeable and may exhibit similar artifacts when applied to imagery of uneven terrain acquired by a pushbroom scanner. Its principal advantage is that it makes few, if any, alterations to pixel values. This is an advantage in applications where even minor changes may be considered significant.

A second, more complex, approach to resampling is bilinear interpolation (**Figure 11.18**). This method calculates a value for each output pixel based on a weighted average

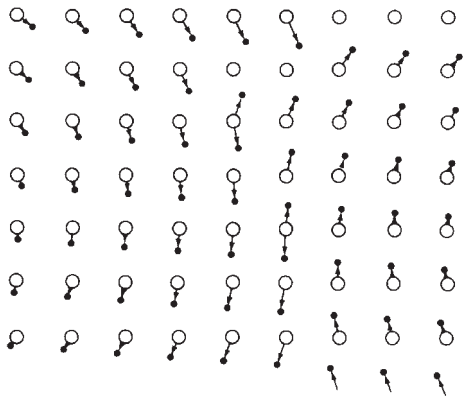


FIGURE 11.17 Nearest-neighbor resampling. Each estimated value (•) receives its value from the nearest point on the reference grid (O).

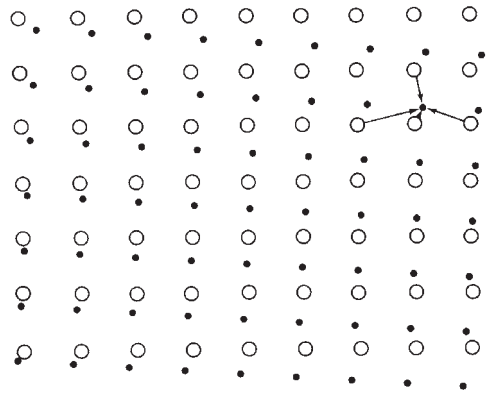


FIGURE 11.18 Bilinear interpolation. Each estimated value (•) in the output image is formed by calculating a weighted average of the values of the four nearest neighbors in the input image (O). Each estimated value is weighted according to its distance from the known values in the input image.

of the four nearest input pixels. In this context, "weighted" means that nearer pixel values are given greater influence in calculating output values than are more distant pixels. Because each output value is based on several input values, the output image will not have the unnaturally blocky appearance of some nearest-neighbor images. The image therefore has a more "natural" look. Yet there are important changes. First, because bilinear interpolation creates new pixel values, the brightness values in the input image are lost. The analyst may find that the range of brightness values in the output image differs from that of the input image. Such changes to digital brightness values may be significant in later processing steps. Second, because the resampling is conducted by averaging over areas (i.e., blocks of pixels), it decreases spatial resolution by a kind of "smearing" caused by averaging small features with adjacent background pixels.

Finally, the most sophisticated, most complex, and most widely used resampling method is cubic convolution (Figure 11.19). Cubic convolution uses a weighted average of values within a neighborhood that extends about two pixels in each direction, usually encompassing 16 adjacent pixels. Typically, the images produced by cubic convolution resampling are much more attractive than those of other procedures, but the data are

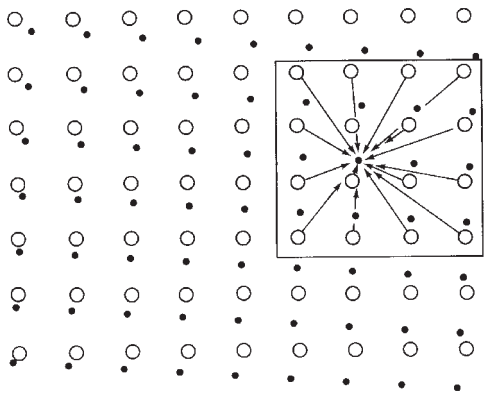


FIGURE 11.19 Cubic convolution. Each estimated value in the output matrix (•) is found by assessing values within a neighborhood of 16 pixels in the input image (O).

altered more than are those of nearest-neighbor or bilinear interpolation, the computations are more intensive, and the minimum number of GCPs is larger. Operational programs commonly do not resample using nearest-neighbor. At the time of this writing, for example, Landsat 8 OLI data are resampled using cubic convolution, and Sentinel-2 MSI data are resampled using bilinear interpolation.

Identification of GCPs

A practical problem in applying image registration procedures is the selection of control points (Figure 11.20). GCPs are features that can be located with precision and accuracy on accurate maps, yet are also easily located on digital images. Ideally, GCPs could be as small as a single pixel, if one could be easily identified against its background. In practice, most GCPs are likely to be spectrally distinct areas as small as a few pixels. Examples might include intersections of major highways, distinctive water bodies, edges of land cover parcels, stream junctions, and similar features (Figure 11.21). Although identification of such points may seem to be an easy task, in fact, the difficulties that might emerge during this step can form a serious roadblock to the entire analytical process, as another procedure may depend on completion of an accurate registration.

Typically, it is relatively easy to find a rather small or modest-sized set of control points. However, in some scenes, the analyst finds it increasingly difficult to expand this set, as one has less and less confidence in each new point added to the set of GCPs. Thus, there may be a rather small set of "good" GCPs, points that the analyst can locate with confidence and precision both on the image and on an accurate map of the region.

The locations may also be a problem. In principle, GCPs should be dispersed throughout the image, with good coverage near edges. Obviously, little is to be gained from having a large number of GCPs if they are all concentrated in a few regions of the image. Analysts who attempt to expand areal coverage to ensure good dispersion are forced to consider points in which it is difficult to locate GCPs with confidence. Therefore, the desires to select "good" GCPs and to achieve good dispersion may work against each other such that the analyst finds it difficult to select a judicious balance. Analysts should anticipate difficulties in selecting GCPs as they prepare subsets early in the analytical process. If subsets are too small, or if they do not encompass important landmarks, the analysts may later find that the subset region of the image does not permit selection of a sufficient number of high-quality GCPs.

Bernstein (1983) presents information that shows how registration error decreases as the number of GCPs is increased. Obviously, it is better to have more rather than fewer GCPs. But, as explained above, the quality of GCP accuracy may decrease as their number increases because the analyst usually picks the best points first.

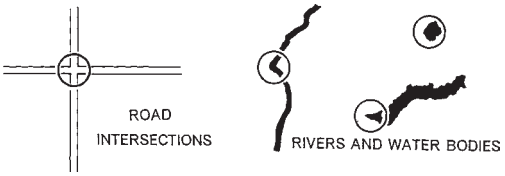


FIGURE 11.20 Selection of distinctive ground control points.

FIGURE 11.21 Examples of ground control points (GCPs). GCPs must be identifiable both on the image and on a planimetrically correct reference map.

Many image processing programs permit the analyst to anticipate the accuracy of the registration by reporting errors observed at each GCP if a specific registration has been applied. The standard measure of the location error is the *root mean square error* (RMSE), which is the standard deviation of the difference between the actual positions of GCPs and their calculated positions (i.e., after registration). These differences are known as the *residuals*. Usually, RMSE is reported in units of image pixels for both north–south and east–west directions (Table 11.11). Points with abnormally high residuals should be reevaluated, but deleted or moved only if, upon further inspection, there is a clear problem with the GCP quality. Many image processing programs can suggest candidate GCP locations on the target image once there are enough GCPs to create a model. This feature should be used with caution, as novice analysts often seek to minimize the RMSE of each succeeding point based on what may be a very poor initial model. To assess the overall accuracy of the registration, some of the GCPs should be withheld from the registration procedure (*check points*) and then used to evaluate its success. Registration accuracy can also be calculated using a cross-validation procedure employing the same points used to build the model, though this capability is rarely available in software solutions.

Some images distributed by commercial enterprises or governmental agencies have been georeferenced by standardized processing algorithms to meet specified standards. Some users may therefore find that the positional accuracy of such images is satisfactory for their needs. Such images are designated by specific processing levels. For example, Landsat 8 OLI data are terrain corrected using cubic convolution, high-quality digital elevation models, inputs from the sensor and spacecraft, and an established archive of GCPs (USGS, 2016). Other image providers employ similar procedures.

TABLE 11.11 Sample Tabulation of Data for Ground Central Points

Point no.	Image *X* pixel	*X* pixel residual	Image *Y* pixel	*Y* pixel residual
1	1269.75	−0.2471E+00	1247.59	0.1359E+02
2	867.91	−0.6093E+01	1303.90	0.8904E+01
3	467.79	−0.1121E+02	1360.51	0.5514E+01
4	150.52	0.6752E+02	1413.42	−0.8580E+01
5	82.20	−0.3796E+01	163.19	0.6189E+01
6	260.89	0.2890E+01	134.23	0.5234E+01
7	680.59	0.3595E+01	70.16	0.9162E+01
8	919.18	0.1518E+02	33.74	0.1074E+02
9	1191.71	0.6705E+01	689.27	0.1127E+02
10	1031.18	0.4180E+01	553.89	0.1189E+02
11	622.44	−0.6564E+01	1029.43	0.8427E+01
12	376.04	−0.5964E+01	737.76	0.6761E+01
13	162.56	−0.7443E+01	725.63	0.8627E+01
14	284.05	−0.1495E+02	1503.73	0.1573E+02
15	119.67	−0.8329E+01	461.59	0.4594E+01
16	529.78	−0.2243E+00	419.11	0.5112E+01
17	210.42	−0.1558E+02	1040.89	−0.1107E+01
18	781.85	−0.2915E+02	714.94	−0.1521E+03
19	1051.54	−0.4590E+00	1148.97	0.1697E+02
20	1105.95	0.9946E+01	117.04	0.1304E+02

Note: *X* root mean square error (RMSE) error = 18.26133. *Y* RMSE = 35.33221. Total RMSE = 39.77237

Point no.	Error	Error contribution by point
1	13.5913	0.3417
2	10.7890	0.2713
3	12.4971	0.3142
4	68.0670	1.7114
5	7.2608	0.1826
6	5.9790	0.1503
7	9.8416	0.2474
8	18.5911	0.4674
9	13.1155	0.3298
10	12.6024	0.3169
11	10.6815	0.2686
12	9.0161	0.2267
13	11.3944	0.2865
14	21.6990	0.5456
15	9.5121	0.2392
16	5.1174	0.1287
17	15.6177	0.3927
18	154.8258	3.8928
19	16.9715	0.4267
20	16.3982	0.4123

11.6 IMAGE DATA PROCESSING STANDARDS

The availability of analysis-ready data from agencies such as the USGS, NOAA, NASA, and the European Space Agency has limited the availability of lower-order products except for internal users developing or testing the preprocessing algorithms. For example, data from the MSI on Sentinel-2a have the following processing levels (Table 11.12), but only levels 1c and 2a are made available to users.

11.7 SUMMARY

Understanding the remotely sensed data being considered for or used in an analysis is vital. The increasing availability of analysis-ready data is decreasing the need for radiometric and geometric preprocessing by the image end user. However, no analysis should proceed without a visual and statistical understanding of the image data, including its covariance structure. Feature extraction, whether by principal components analysis, tasseled cap transformation, or other means, is often helpful both to reduce dimensionality and to reveal hidden structure and meaning in the data.

REVIEW QUESTIONS

1. How can an analyst know if preprocessing is advisable? Suggest how you might make this determination?
2. How can an analyst determine if specific preprocessing procedures have been effective?
3. Discuss the merits of preprocessing techniques that improve the visual appearance of an image but do not alter its basic statistical properties. Are visual qualities important in the context of image analysis?
4. Examine images and maps of your region to identify prospective GCPs. Evaluate the pattern of GCPs. Is the pattern even, or is it necessary to select questionable points to attain an even distribution?
5. Are optimum decisions regarding preprocessing likely to vary according to the subject of the investigation? For example, would optimum preprocessing decisions for a land cover analysis differ from those for a hydrologic or geologic analysis?
6. Can you identify analogies for preprocessing in other contexts?

TABLE 11.12 Sentinel-2 MSI (European Space Agency) Processing Levels

Level	Characteristics
0	Compressed raw data
1a	Uncompressed raw data
1b	Radiometrically corrected radiance data
1c	Orthorectified TOA reflectance
2a	BOA reflectance

7. Images from unmanned aerial vehicles (UAVs) are increasingly available to analysts and decision makers and often require substantial preprocessing. What factors lead to the need for increased geometric and radiometric preprocessing of UAV-acquired data when compared to, say, data from a satellite-based program like Sentinel?

REFERENCES

Baig, M. H. A., L. Zhang, T. Shuai, and Q. Tong. 2014. Derivation of a Tasseled Cap Transformation Based on Landsat 8 At-Satellite Reflectance. *Remote Sensing Letters,* Vol. 5, pp. 423–431.

Bernstein, R. 1983. Image Geometry and Rectification. Chapter 21 in *Manual of Remote Sensing* (R. N. Colwell, ed.). Falls Church, VA: American Society of Photogrammetry, pp. 873–922.

Chavez, P. S. 1975. Atmospheric, Solar, and M.T.F. Corrections for ERTS Digital Imagery. Proceedings, American Society of Photogrammetry, Bethesda, MD: American Society for Photogrammetry and Remote Sensing, pp. 69–69a.

Cohen, W. B., and S. N. Goward. 2004. Landsat's Role in Ecological Applications of Remote Sensing. *Bioscience,* Vol. 54, pp. 535–545.

Crist, E. P. 1983. The Thematic Mapper Tasseled Cap: A Preliminary Formulation. Machine Processing of Remotely Sensed Data: Natural Resources Evaluation; Proceedings of the Ninth International Symposium, West Lafayette, Indiana, June 21–23, 1983 (A84–13004 03–43). New York, Institute of Electrical and Electronics Engineers, pp. 357–364.

Crist, E. P, and R. C. Cicone. 1984. A Physically-Based Transformation of Thematic Mapper Data—The TM Tasseled Cap. *IEEE Transactions on Geoscience and Remote Sensing,* Vol. GE-22, pp. 256–263.

Davis, J. C. 2002. *Statistics and Data Analysis in Geology* (3rd ed.). New York: John Wiley, 656 pp.

Estes, J. E., E. J. Hajic, and L. R. Tinney. 1983. Fundamentals of Image Analysis: Analysis of Visible and Thermal Infrared Data. Chapter 24 in *Manual of Remote Sensing* (R. N. Colwell, ed.). Falls Church, VA: American Society of Photogrammetry, pp. 987–1124.

Jackson, R. 1983. Spectral Indices in N-Space. *Remote Sensing of Environment,* Vol. 13, No. 5, pp. 409–421.

Kauth, R. J., and G. S. Thomas. 1976. The Tasseled Cap—A Graphic Description of the Spectral–Temporal Development of Agricultural Crops as Seen by Landsat. In LARS: Proceedings of the Symposium on Machine Processing of Remotely Sensed Data. West Lafayette, IN: Purdue University Press, pp. 4B-41–4B-51.

Mather, P. M., and M. Koch. 2011. *Computer Processing of Remotely Sensed Images: An Introduction* (4th ed.). Chichester, UK: John Wiley, 460 pp.

Mayer, B., and A. Kylling. 2005. Technical Note: The libRadtran Software Package for Radiative Transfer Calculations: Description and Examples of Use. *Atmospheric Chemistry and Physics,* Vol. 5, pp. 1855–1877.

Richards, J. A. (2013). *Remote Sensing Digital Image Analysis: An Introduction* (5th ed.). Heidelberg: Springer, 494 pp.

Richter, R., and D. Schlaepfer. 2011. *Atmospheric/Topographic Correction for Satellite Imagery: ATCOR-2/3 User Guide* Vers. 8.0.2. DLR—German Aerospace Center, Remote Sensing Data Center.

Schlens, J. 2003. A Tutorial on Principal Component Analysis: Derivation, Discussion and Singular Value Decomposition. Available at *www.cs.princeton.edu/picasso/mats/PCA-Tutorial-Intuition_jp.pdf.*

Schlens, J. 2014. A Tutorial on Principal Component Analysis. Available at *http://adsabs.harvard.edu/abs/2014arXiv1404.1100S.*

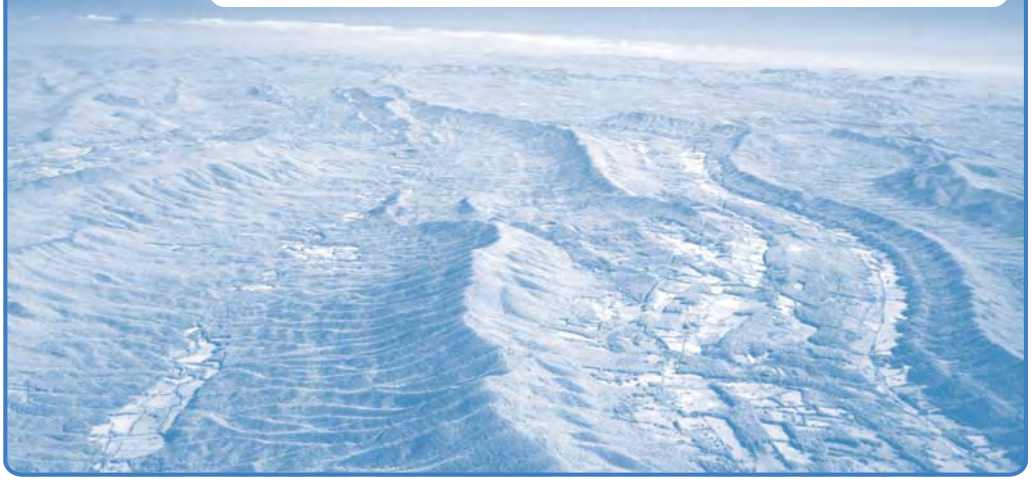
Tobler, W. R. 1970. A Computer Movie Simulating Urban Growth in the Detroit Region. *Economic Geography,* Vol. 46, p. 234.

U.S. Geological Survey (USGS). 2016. *Landsat 8 Data Users Handbook*. LSDS-1574 Version 2.0. Sioux Falls, SD. Available at *https://usgs.gov/landsat-missions/landsat-8-data-users-handbook*.

Vermote, E., D. Tanre, J. L. Deuze, M. Herman, and J. J. Morcrette. 1997. Second Simulation of the Satellite Signal in the Solar Spectrum (6S). 6S User Guide Version 2. Appendix III: Description of the Subroutines. *IEEE Transactions on Geoscience and Remote Sensing,* Vol. 35, pp. 675–686.

Zhai, Y., D. P. Roy, V. S. Martins, H. K. Zhang, L. Yan, and Z. Li. 2022. Conterminous United States Landsat-8 Top of Atmosphere and Surface Reflectance Tasseled Cap Transformation Coefficients. *Remote Sensing of Environment*, Vol. 274, No. 112992.

12 Image Classification



MAJOR TOPICS TO UNDERSTAND

- Informational Classes and Spectral Classes/Samples
- Unsupervised Classification
- Supervised Classification

12.1 INTRODUCTION

Digital image classification is the process of assigning single multiband pixels (Figure 12.1a) or multipixel objects (Figure 12.1b) to classes (categories). An example of this type of analysis is determining whether a given pixel or object is a forest or some other type of land cover. Analyses that result in estimates of continuous variables for pixels or objects are also common. For example, even if we already know that an area is a forest, managers may need maps of tree height, canopy cover, leaf area index, or biomass on a per-pixel or per-object basis. Our focus in this chapter is on classification.

Usually, each pixel is treated as a vector composed of values in several spectral bands at one point in time (Figure 12.2), multiple points in time (more commonly), or some equivalent vector derived from nonoptical remotely sensed data. By comparing pixels to one another and to pixels of known identity, it is possible to assemble groups of similar

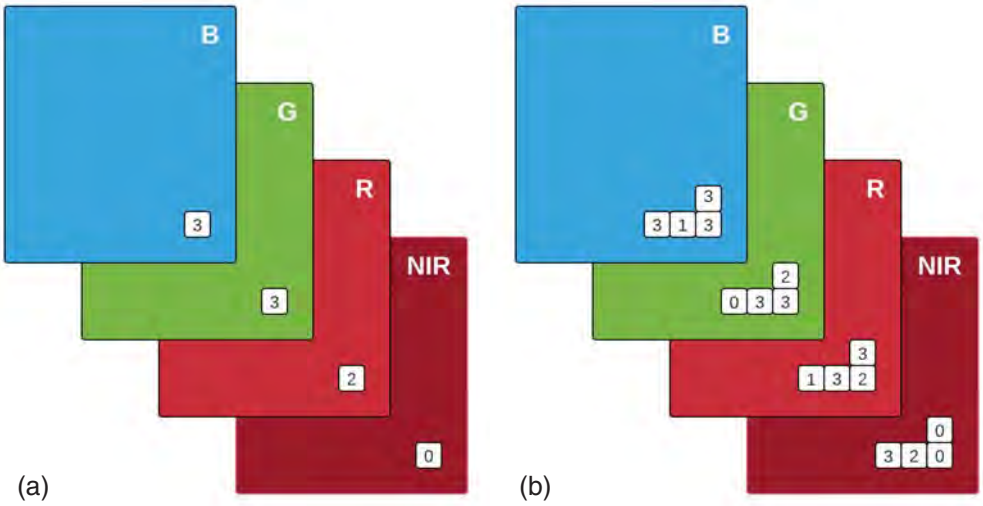


FIGURE 12.1 (a) Pixel as a single set of spectral values considered in isolation from its neighbors. (b) Object as a group of similar adjacent pixels.

pixels into classes that are associated with the informational categories of interest to users of remotely sensed data. These classes form regions on a map or an image, so that after classification the digital image is presented as a mosaic of uniform parcels, each identified by a color or symbol (Figure 12.3). These classes are, in theory, homogeneous: pixels within classes are spectrally more similar to one another than they are to pixels in other classes. In practice, of course, each class will display some diversity.

Image classification is an important part of the fields of remote sensing, image analysis, and pattern recognition. In some instances, the classification itself may be the object of the analysis. For example, classification of land use from remotely sensed data (Chapter 21) produces a map-like image as the final product of the analysis. In other instances, the classification may be only an intermediate step in a more elaborate analysis, in which the classified data form one of several data layers in a GIS. For example, in a study of water quality, an initial step may be to use image classification to identify wetlands and open water within a scene. Later steps may then focus on more detailed study of these areas to

$$\begin{bmatrix} \text{Blue} \\ \text{Green} \\ \text{Red} \\ \text{NIR} \end{bmatrix} \quad \begin{bmatrix} 3 \\ 3 \\ 2 \\ 0 \end{bmatrix} \quad \begin{bmatrix} 7713 \\ 6865 \\ 6013 \\ 5345 \end{bmatrix}$$

(a) (b) (c)

FIGURE 12.2 (a) Spectral bands shown in parts b and c. (b) Vector derived from pixel shown in **Figure 12.1a.** (c) Scaled Landsat 8 Operational Land Imager (OLI) top-of-atmosphere reflectance vector from Leech Lake, Minnesota, on September 29, 2015 (397227E, 52223510N, UTM zone 15N, scene ID LC80280272015272LGN00).

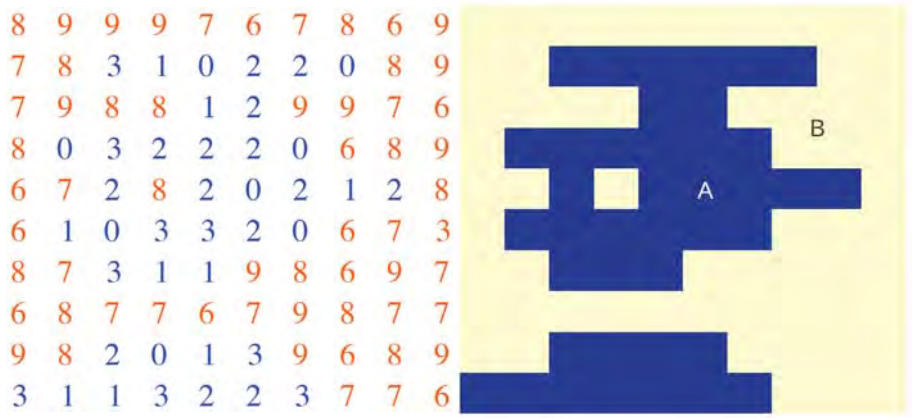


FIGURE 12.3 (a) Numeric image. (b) Classified image. The classified image is defined by examining the numeric image and then grouping together those pixels that have similar spectral values. Here class "A" is formed from values 0, 1, 2, and 3 and class "B" from values 6, 7, 8, and 9.

identify influences on water quality and to map variations in water quality. Image classification therefore forms an important tool for examination of digital images—sometimes to produce a final product, other times as one of several analytical procedures applied to derive information from an image.

The term *classifier* refers loosely to a computer program that implements a specific procedure for image classification. Over the years scientists have devised many classification strategies. The analyst must select a classification method that will best accomplish a specific task. At present it is not possible to state that a given classifier is "best" for all situations because the characteristics of each image and the circumstances for each study vary so greatly. Therefore, it is essential that each analyst understand the alternative strategies for image classification so that she or he is prepared to select the most appropriate classifier for the task at hand.

The simplest form of digital image classification is to consider each pixel individually, assigning it to a class based on its reflectance vector. Sometimes such classifiers are referred to as spectral or point classifiers because they consider each pixel as a "point" observation. Although point classifiers offer the benefits of simplicity and economy, they are not capable of exploiting the information contained in relationships between each pixel and those that neighbor it. Human interpreters, for example, could derive little information using the point-by-point approach because humans derive less information from the brightnesses of individual pixels than they do from the context and the patterns of brightnesses, of groups of pixels, and from the sizes, shapes, and arrangements of parcels of adjacent pixels. These are the same qualities that we discussed in the context of manual image interpretation.

As an alternative, classification of multipixel objects (a key facet of *GEOgraphic-Object-Based Image Analysis [GEOBIA]*) enables use of texture, shape, size, context (site/association), and other characteristics that are so important for the human interpreter (**Figures 12.1b, 12.4, and 12.5**). Because of these advantages, object-based classifi-

(a)	(b)
m_{Blue}	2.5
m_{Green}	2.0
m_{Red}	2.3
m_{NIR}	1.3
s_{Blue}	1.0
s_{Green}	1.4
s_{Red}	1.0
s_{NIR}	1.5
A	4.0
P	10.0
$A:P$	0.4

FIGURE 12.4 (a) Vector positions of object-specific band means (m), standard deviations (s), and shape variables (area, A; perimeter, P; area:perimeter ratio, $A:P$). (b) Object vector example using object shown in **Figure 12.1b**.

ers are now the de facto standard for many analyses, and an informed analyst must always carefully consider whether pixels or objects will be most appropriate for a given task.

While there are no uniformly applicable rules for when objects will be the better choice of analysis unit, many analysts will gravitate toward the use of objects when the following conditions apply:

1. The image is composed of distinct entities.
2. The spatial resolution is much finer than the size of the entities.

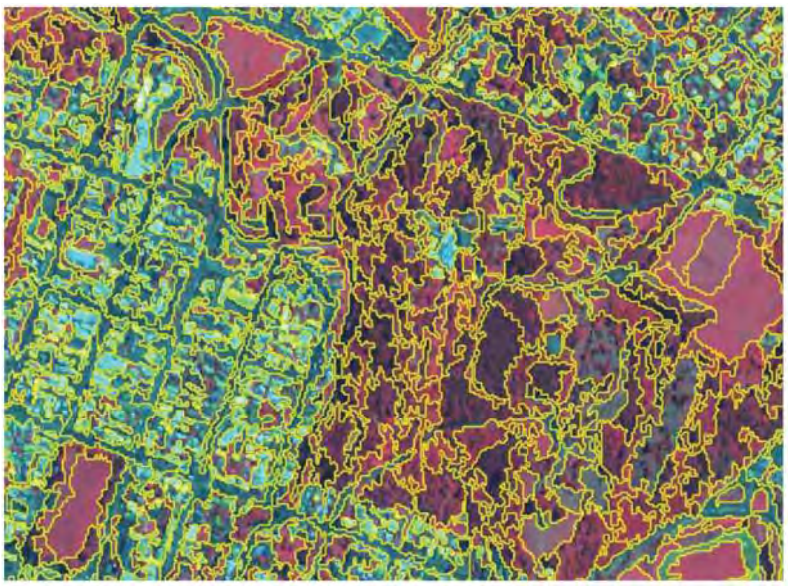


FIGURE 12.5 Segmentation of standard false-color composite IKONOS image of Dunedin City and the surrounding rural areas, New Zealand, acquired February 20, 2005. Yellow lines delineate the image objects. From Mathieu et al. (2007).

3. Shape, size, texture, and other characteristics of image interpretation uniquely available from GEOBIA are essential to an accurate classification.
4. There is an automated means by which useful objects can be created in a minimally heuristic manner.

Another kind of distinction in image classification separates supervised classification from unsupervised classification. Supervised classification procedures require considerable interaction with the analyst, who must guide the classification by identifying areas on the image that are known to belong to each category (i.e., the analyst identifies *spectral classes* within known *informational classes*). Unsupervised classification, on the other hand, proceeds with only minimal interaction with the analyst, in a search for natural groups of pixels present within the image (i.e., the algorithm identifies spectral classes—also known as clusters—typically without prior knowledge of the informational class to which they belong). The distinction between supervised and unsupervised classification is useful, especially for students who are first learning about image classification. But the two strategies are not as clearly distinct as these definitions suggest, for some methods do not fit neatly into either category. These so-called hybrid classifiers share characteristics of both supervised and unsupervised methods.

12.2 INFORMATIONAL CLASSES AND SPECTRAL CLASSES/SAMPLES

Informational classes are the categories of interest to the users of the data. Informational classes are, for example, the different kinds of geological units, forest, or the different kinds of land use that convey information to planners, managers, administrators, and scientists who use information derived from remotely sensed data. These classes are the information that we wish to derive from the data—they are the object of our analysis. Unfortunately, these classes are not directly recorded on remotely sensed images; we can derive them only indirectly, using the evidence contained in brightnesses recorded by each image. For example, the image cannot directly show geological units, but rather only the differences in topography, vegetation, soil color, shadow, and other factors that lead the analyst to conclude that certain geological conditions exist in specific areas.

Spectral classes are groups of pixels that are uniform with respect to the brightnesses in their several spectral channels. The statistics derived from spectral classes (e.g., mean and covariance matrix) are used in many classifiers. Others use reflectance vectors from individual pixels or objects (spectral samples) rather than descriptive statistics from spectral classes.

Remote sensing classification proceeds by matching spectral categories to informational categories. If the match can be made with confidence, then the information is likely to be reliable. If spectral and informational categories do not correspond, then the image is unlikely to be a useful source for that particular form of information. Seldom can we expect to find exact one-to-one matches between informational and spectral classes. Any informational class includes spectral variations arising from natural variations within the class (Figure 12.6). For example, a region of the informational class *forest* is still *forest*, even though it may display variations in age, species composition, density, and vigor, all of which lead to differences in the spectral appearance of a single informational class.

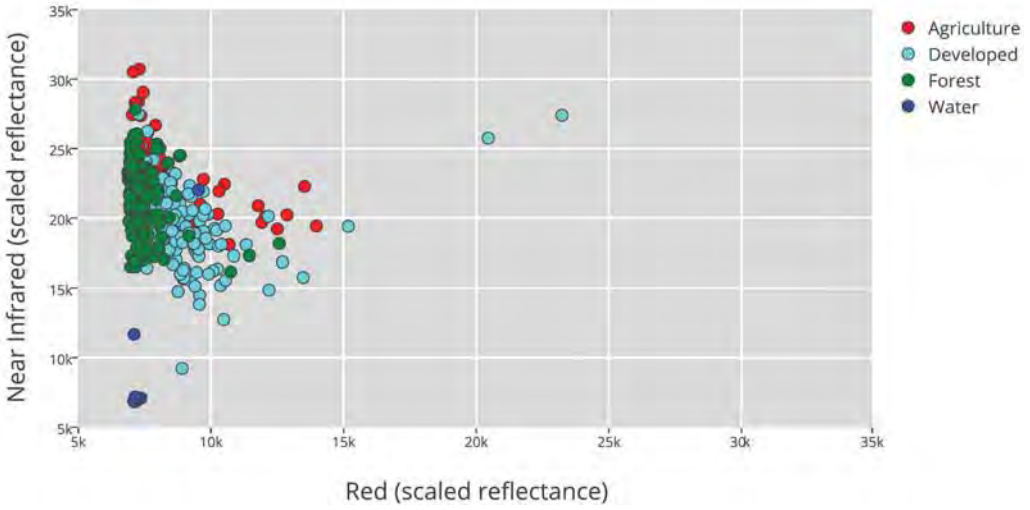


FIGURE 12.6 Two-dimensional scatter plot (near-infrared vs. red) showing pixels with known informational classes within the feature space.

Furthermore, other factors, such as variations in illumination and shadowing, may produce additional variations even within otherwise spectrally uniform classes. Fundamentally, nearly every informational class is composed of more than one spectral class, a one-to-many relationship, and this is not a concern with respect to classifier performance. However, a spectral class composed of multiple informational classes will decrease classification accuracy, since in typical "hard" classifications it will still have to be assigned to one informational class in the final recoding step.

Thus, informational classes are typically composed of numerous spectral subclasses, spectrally distinct groups of pixels that together may be assembled to form an informational class. In digital classification, we must often treat spectral subclasses as distinct units during classification but then recode spectral classes to their corresponding informational classes for the final image or map to be used by planners or administrators (who are, after all, interested only in the informational categories, not in the intermediate steps required to generate them).

In subsequent sections, we will be interested in several properties of spectral classes. For each band, each class is characterized by a mean, or an average, value that of course represents the typical brightness of each class. In nature, all classes exhibit some variability around their mean values; some pixels are darker than the average, others a bit brighter. These departures from the mean are measured as the variance or sometimes as the standard deviation (the square root of the variance; Chapter 11).

12.3 UNSUPERVISED CLASSIFICATION

Unsupervised classification can be defined as the identification of natural groups, or structures, within multispectral data (Figures 12.7 and 12.8). The notion of the existence of natural, inherent groupings of spectral values within a scene may not be intuitively

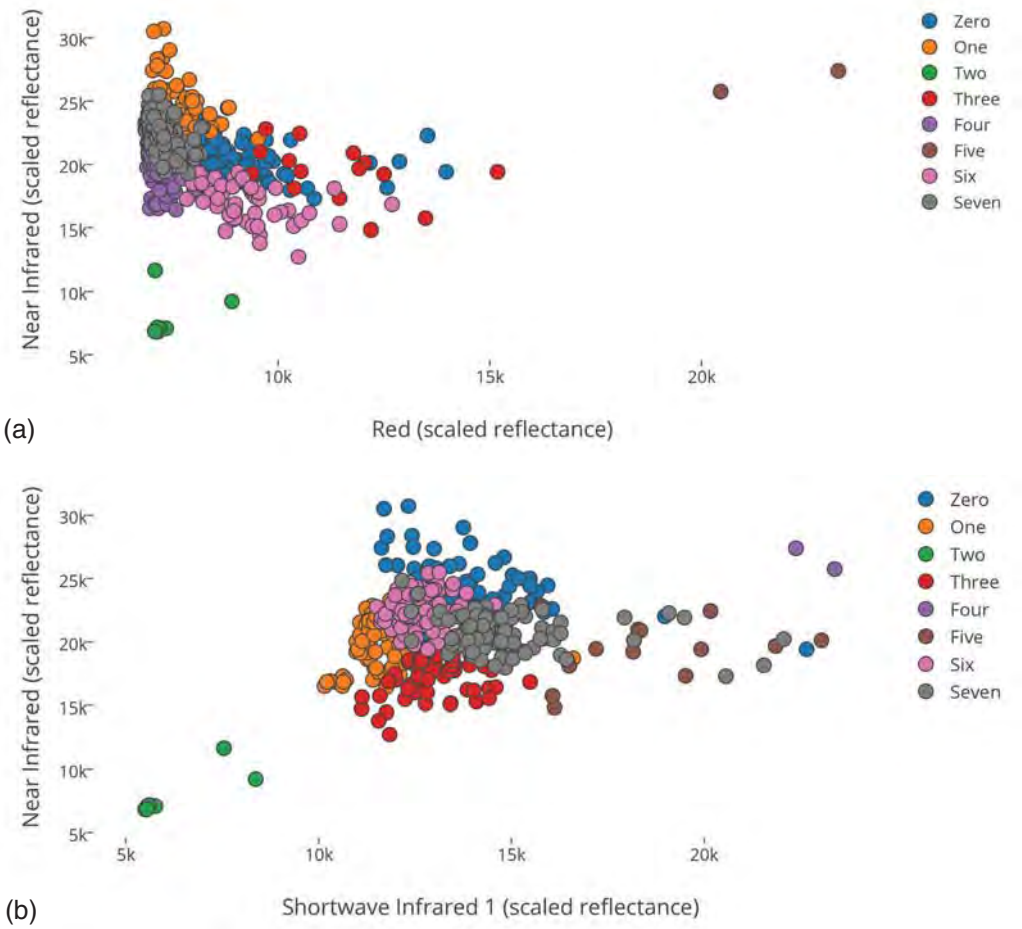


FIGURE 12.7 (a) NIR versus red scatter plot showing the result of an unsupervised classification into seven spectral classes. (b) NIR versus SWIR1 scatter plot showing the result of an unsupervised classification into seven spectral classes.

obvious, but it can be demonstrated that remotely sensed images are usually composed of spectral classes that are reasonably uniform internally in respect to brightnesses in several spectral channels. Unsupervised classification is the definition, identification, labeling, and mapping of these natural classes, also called *clusters*.

Advantages

The advantages of unsupervised classification (relative to supervised classification) can be enumerated as follows:

- No extensive prior knowledge of the region is required. Or, more accurately, the nature of knowledge required for unsupervised classification differs from that required for supervised classification. To conduct supervised classification, detailed knowledge of

the area to be examined is required to select representative examples of each class to be mapped. To conduct unsupervised classification, no detailed prior knowledge is required, but knowledge of the region is required to evaluate and label machine-generated spectral classes.

- Opportunity for human error (at least with respect to identifying spectral classes) is minimized. To conduct unsupervised classification, the operator might specify only the number of categories desired (or possibly, minimum and maximum limits on the number of categories), and sometimes constraints governing the distinctness and uniformity of groups. Many of the detailed decisions required for supervised classification are not required for unsupervised classification, so the analyst is presented with less opportunity for error. If the analyst has inaccurate preconceptions regarding the region, she or he will have little opportunity to influence the classification. Machine-generated spectral classes (clusters) also typically partition the feature space well, with minimal within-class heterogeneity.

- Unique classes are recognized as distinct units. Such classes (perhaps of very small areal extent, or widely dispersed, or both) may remain unrecognized in the process of supervised classification and could inadvertently be incorporated into other classes, generating error and imprecision throughout the entire classification.

Disadvantages and Limitations

The disadvantages and limitations of unsupervised classification arise primarily from reliance on "natural" groupings and difficulties in matching these groups to the informational categories that are of interest to the analyst.

- Unsupervised classification identifies spectrally homogeneous classes within the data that do not necessarily correspond to the informational categories that are of interest to the analyst. As a result, the analyst is faced with the problem of matching spectral classes generated by the classification to the informational classes that are required by the ultimate user of the information. Seldom is there a simple one-to-one correspondence between the two sets of classes (e.g., Figure 12.9). Some hybrid approaches (see below) focus specifically on developing informationally homogeneous spectral classes.

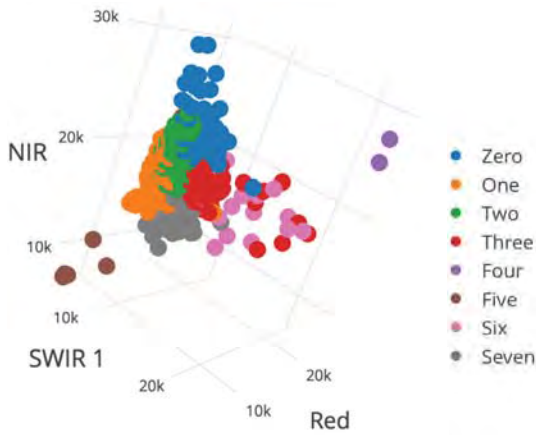


FIGURE 12.8 Three-dimensional scatter plot showing the same seven spectral classes shown in **Figure 12.7**.

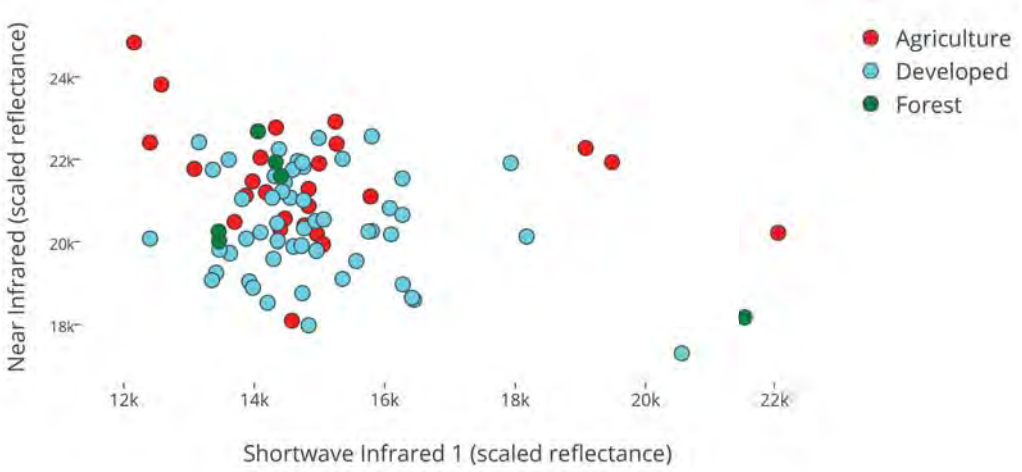


FIGURE 12.9 Informational classes contained within spectral class six in **Figure 12.7b**. Note that this spectral class, while a natural grouping of "like" pixels, is not homogeneous with respect to informational class composition.

- The analyst has limited control over the menu of classes and their specific identities, though some algorithms do permit "seeding" with known informational classes and reflectance vectors. Supervised identification of initial cluster centers, followed by unsupervised identification of spectrally homogeneous spectral classes, has characteristics of both supervised and unsupervised methods (i.e., "hybrid" classification), and provides evidence that the distinction between the two approaches is not as clear as idealized descriptions imply. Generally, however, if it is necessary to generate a specific menu of informational classes (e.g., to match to other classifications for other dates or adjacent regions), the use of unsupervised classification may be unsatisfactory.

- Spectral properties of specific informational classes will change over time (on a seasonal basis, as well as over the years). As a result, relationships between informational classes and spectral classes are not constant, and relationships defined for one image usually cannot be extended to others.

Distance Measures

Unsupervised classification of an entire image must consider many thousands of pixels. But the clustering process is always based on the answer to the same question: "Do the two pixels belong to the same group?" Answering this question requires finding the distances between pairs of pixels.

A number of methods for finding distances in multidimensional feature space are available. One of the simplest is Euclidean distance, which is the square root of the inner product of the difference vectors, as follows:

$$\sqrt{(\mathbf{x}_1 - \mathbf{x}_2)^T (\mathbf{x}_1 - \mathbf{x}_2)} \qquad \text{(EQ. 12.1)}$$

where $\mathbf{x}$ is a vector from a pixel or object, and 1 and 2 are two different pixels or objects.

This is the square root of the sum of the band-specific squared differences. This measure can be applied to as many dimensions (spectral channels) as might be available. Using, for example, the reflectance vectors in Table 12.1, we obtain the Euclidean distances shown in Table 12.2.

The Euclidean distance has little significance in itself, but in relation to other Euclidean distances it forms a means of defining similarities between pixels. For example, we can see that the two forested pixels are more similar to each other than to the agriculture pixel and that both are comparably distant from the agriculture pixel.

Not all distance measures are based on Euclidean distance. Other possibilities include the L1 distance (the sum of the absolute differences between values in individual bands; Swain and Davis, 1978) or the Pearson distance.

k-Means

Remote sensing analysts use numerous clustering algorithms, and exploring the nuances of each is beyond the scope of this book. However, one of the most useful and widely implemented algorithms is *k*-means. The *k* in *k*-means refers to the number of clusters, which must be specified a priori by the analyst. While numerous formalized methods have been proposed by which *k* can be specified, it is most commonly chosen heuristically to address two fundamental criteria, as follows: (1) finding the minimum number of clusters that maximize the variance explained (i.e., the point at which additional clusters explain very little additional variance); and (2) maximizing spectral class homogeneity with respect to informational classes. The latter tends to be the dominant factor in the selection of *k* for most remote sensing applications.

TABLE 12.1 Scaled Reflectance Vectors from Landsat 8 OLI

	For 1	For 2	Ag 1
Blue	8799	8876	9134
Green	8493	8666	9200
Red	6949	7108	7617
NIR	20581	20324	26159
SWIR1	11807	12353	13635
SWIR2	7744	8248	8501

Note: For = forested; Ag = agriculture. Scene (WRS–2 row 17 path 34) acquired May 11, 2015 (Blacksburg, VA, and vicinity).

TABLE 12.2 Euclidean Distances between Pairs of Vectors Shown in **Table 12.1**

	For 1	For 2	Ag 1
FOR 1	0	824	6007
FOR 2	824	0	6030
AG 1	6007	6030	0

Note: For = forested; Ag = agriculture.

The objective of *k*-means is to separate the data into groups of minimized, equal variance. It will always converge with enough iterations but possibly to a local rather than global minimum. The sequence is as follows:

1. Identify *k* initial cluster centers, preferably located throughout the multidimensional feature space.
2. Assign each reflectance vector to its nearest cluster center based on Euclidean distance.
3. Calculate new cluster centers as the mean of the reflectance vectors assigned to each cluster.
4. Repeat steps 2 and 3 until cluster centers no longer move or the maximum number of specified iterations is reached (**Figure 12.10**).

ISODATA

The ISODATA classifier (Ball and Hall, 1965; Duda and Hart, 1973) is a modification of *k*-means. ISODATA, unlike *k*-means, does not require a priori identification of the number of clusters, but other analyst-specified parameters (with some variation by implementation) affect the resulting clustering, including the following: (1) the initial number of clusters, (2) the maximum number of iterations, (3) the minimum number of pixels per cluster, (4) the maximum standard deviation of pixels from their cluster center, and (5) the minimum required distance between two cluster centers.

Initial cluster centers can be prespecified by the analyst (see above). More commonly, however, they are statistically initialized throughout the feature space, often along the first principal component axis. The algorithm flow is as follows (again, with some variation across implementations):

1. Assign all pixels to the nearest cluster center.
2. Delete clusters with fewer than the minimum number of pixels; reassign pixels within deleted clusters.
3. Calculate cluster means, within-cluster standard deviations (even iterations), and among-cluster Euclidean distances (odd iterations).
4. Merge clusters that are closer than the minimum required distance between two cluster centers (odd iterations) and recalculate cluster centers, or
5. Split clusters that exceed the maximum standard deviation (even iterations). The two new class centers are plus or minus one standard deviation from the previous center.
6. Check to see if the maximum number of iterations has been reached. If so, stop; otherwise repeat. (Some implementations also allow specification of the degree of convergence between successive iterations as an additional stopping criterion.)

Mather (1976) suggests resolving ISODATA's occasional difficulties by first creating a superset of clusters (i.e., specifying a very large initial *k*) and then implementing only the merge function. (Fortran code for this *k*-means modification is available in Mather [1976]; a Python implementation is in Whitebox Tools [Lindsay, 2017].)

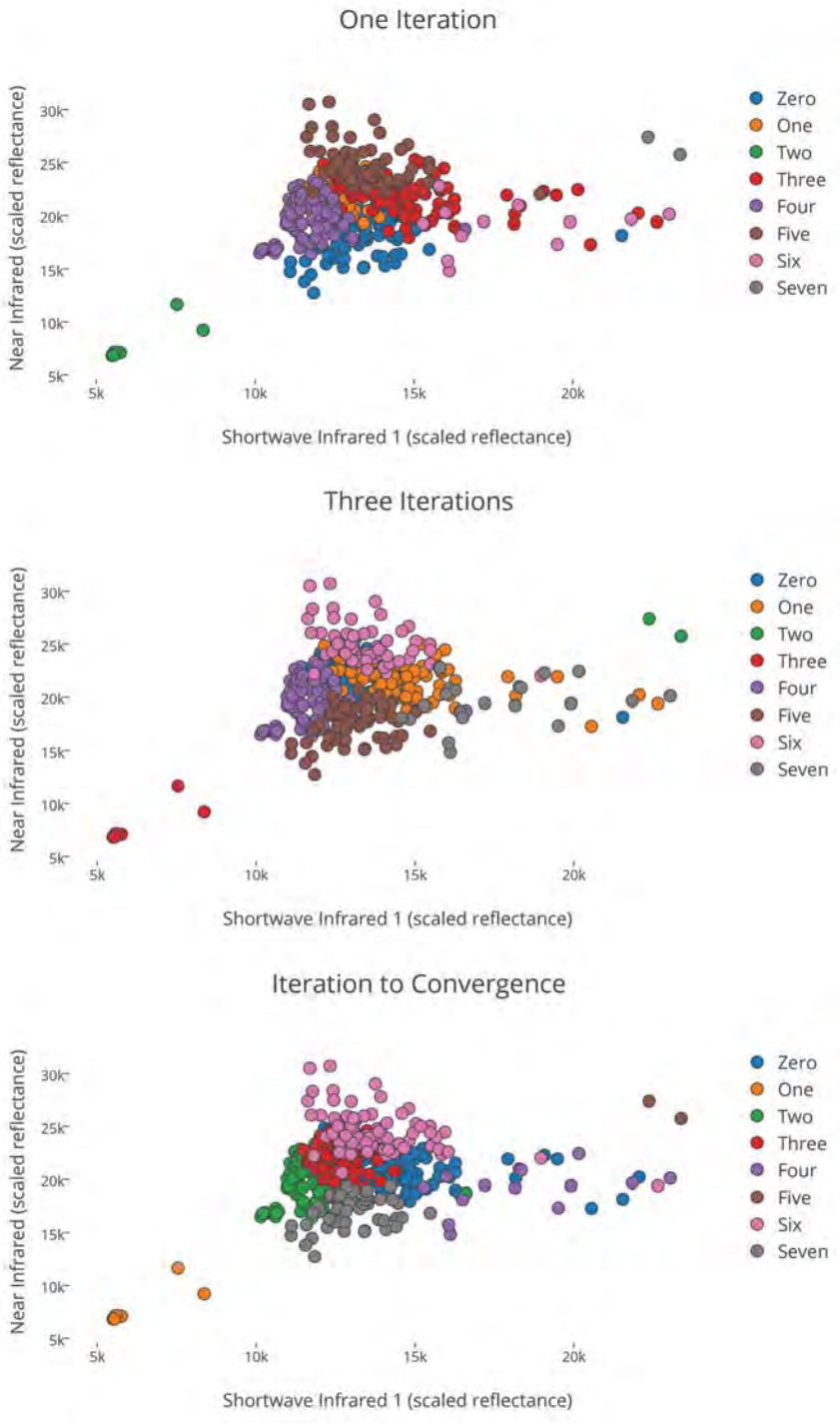


FIGURE 12.10 Unsupervised classification of Landsat 8 OLI data with k-means with $k = 7$ at first, third, and final iterations.

Assignment of Spectral Categories to Informational Categories

Regardless of whether *k*-means, ISODATA, or some other clustering algorithm is chosen, the resulting spectral classes have to be labeled with an informational class. This is often an analyst-intensive process, with difficulty increasing as the number of spectral and informational classes increases. A serious practical problem with unsupervised classification is that clear matches between spectral and informational classes are not always possible; some informational categories may not have direct spectral counterparts, and vice versa. Thus, despite the advantages that unsupervised classification has with respect to feature space partitioning, finding nonspatially contiguous classes, and the like, lack of one-to-one correspondence between spectral and informational classes can preclude its use. However, hybrid techniques have successfully addressed the issue of spectral classes being mixed with respect to informational classes. Some examples of these hybrid techniques are detailed in the section "Hybrid Parametric Classifiers" later in this chapter.

12.4 SUPERVISED CLASSIFICATION

Supervised classification can be defined informally as the process of using samples of known identity (i.e., pixels already assigned to informational classes) to classify pixels of unknown identity (i.e., to assign unclassified pixels to one of several informational classes). While most supervised classification algorithms use sample pixels or objects, some decision rules use descriptive statistics from analyst-developed spectral classes (i.e., a spectrally homogeneous group of pixels or objects). Regardless, it is impossible to overemphasize how vital high-quality training data are to classification outcomes. Analysts will often go to great lengths to specify a state-of-the-art classification algorithm, but algorithm performance is completely dependent on the quality of the training data, defined as follows:

1. It is a representative sample that completely partitions the feature space. This typically requires a strong element of (or complete) randomness and a large number of observations.
2. Each observation in the sample is properly labeled with the appropriate informational class.
3. All necessary algorithm-specific criteria are met (e.g., independent and identically distributed observations).
4. When spectral classes, rather than sample pixels or objects, are used, they should be spectrally separable and distinct, especially between spectral classes that are assigned to different informational classes.

Advantages

The advantages of supervised classification, relative to unsupervised classification, can be enumerated as follows. First, the analyst has control of a selected menu of informational categories tailored to a specific purpose and geographic region. This quality may be vitally important if it becomes necessary to generate a classification for the specific purpose of comparison with another classification of the same area at a different date or if the classification must be compatible with those of neighboring regions. Under such

circumstances, the unpredictable (with respect to number, identity, size, and pattern) qualities of categories generated by unsupervised classification may be inconvenient or unsuitable. Second, supervised classification is tied to specific areas of known identity, determined through the process of selecting training areas or sample observations. Third, the analyst using supervised classification is not faced with the problem of matching spectral categories on the final map with the informational categories of interest (this task has, in effect, been addressed during the process of selecting training data). Finally, the operator may be able to detect serious errors in classification by examining training data to determine if they have been correctly classified by the classification procedure—inaccurate classification of training data indicates serious problems in the classification or selection of training data, although correct classification of training data does not always indicate correct classification of other data.

Disadvantage

The primary disadvantage of supervised classification is that conscientious selection of a high-quality training data set is often a subjective, time-consuming, expensive, and tedious undertaking, even if ample resources are at hand. Furthermore, as noted earlier, prior studies (e.g., Scholz, Fuhs, and Hixson, 1979; Hixson, Scholz, and Fuhs, 1980; and many others in the ensuing decades) have shown that training data quality is as important as or more important than the choice of classification algorithm in most cases (really just a remote sensing variant on the well-known adage, "garbage in, garbage out").

Specific Methods for Supervised Classification

A variety of different methods have been devised to implement the basic strategy of supervised classification. The following sections outline only a few of the most commonly used methods of supervised classification. The most widely used methods (Classification and Regression Tree analysis [CART], random forests, k-nearest neighbor [kNN], and artificial neural networks [ANNs]) are considered nonparametric in that they "rely on no or few assumptions about the shape or parameters of the population distribution from which the sample was drawn" (Hoskin, 2012, p. 2). Some methods (maximum likelihood/Bayes, minimum distance to means, and the hybrid parametric classifiers) do assume an approximately normal distribution. In most instances, as noted above, it is the quality of the training sample that drives classification quality, but choice of algorithm is an important second-order decision.

Classification and Regression Tree Analysis

CART (Breiman et al., 1984; Lawrence and Wright, 2001) is one of the most widely used (and most understandable) of all classification techniques. CART (also known as *decision trees*) uses binary recursive splitting to divide the feature space into nonoverlapping (hyper)rectangles, such that each rectangle is as pure (homogeneous) as possible with respect to an informational class. Each split is made by examining all possible split values for each variable to determine the extent to which impurity is reduced, normally by computing the difference between the pre-split purity and the sum of the two post-split purities. Each split is *greedy* in that it only maximizes purity at that step. Purity can be defined in numerous ways, but one of the most common is called the Gini rule:

$$G = \sum_{k=1}^{C} p_k \left(1 - p_k\right) \quad \text{(EQ. 12.2)}$$

where G is the total cost over all classes, C is the number of classes, k is a given class, and p is the proportion of the training observations in a rectangle that are in a given class.

Let's look at this with a couple of examples. Assume in both instances that we have a classification scheme consisting of *forest* (k = 1), *nonforest* (k = 2), and *water* (k = 3). We have 10 observations, all in class *forest*. In that case, *p(k)* for both *nonforest* and *water* is 0, and that for *forest* is 1 (10 of 10). As such,

$$\begin{aligned} G &= \sum_{k=1}^{C} p_k \left(1 - p_k\right) \\ G &= 1(1-1) + 0(1-0) + 0(1-0) \\ G &= 0 \end{aligned} \quad \text{(EQ. 12.3)}$$

Contrast this pure rectangle with one in which we have 4 *forest* observations, 4 *nonforest* observations, and 2 *water* observations:

$$\begin{aligned} G &= \sum_{k=1}^{C} p_k \left(1 - p_k\right) \\ G &= 0.4(1-0.4) + 0.4(1-0.4) + 0.2(1-0.2) \\ G &= 0.64 \end{aligned} \quad \text{(EQ. 12.4)}$$

This is very close to the maximum value of C (1 – C), which is 0.67 in this three-informational class case.

Particularly using the Gini rule, application of CART is sensitive to variations in numbers of pixels; it performs best when numbers of pixels in training data sets are approximately equal. Often the use of a large number of spectral and ancillary variables leads to very accurate results, but results that are tailored to specific data sets—a condition known as *overfitting*. As a result, the application of CART often includes a pruning step, which reduces the tree size to create a more concise, more robust, and generally applicable solution. Many CART implementations improve the robustness of the pruning step by assessing node error rates through implementing a V-fold cross-validation (that also helps assess possible overfitting when a separate hold-out test sample is not feasible). This process divides the data into V subsets, with $V - (1/V)$ observations used for training and $1/V$ observations used for validation for each of V realizations. Most studies have found V = 5 to 10 optimal, though increasing V to 20 is feasible when the number of observations per class is sufficient to allow a test sample to have the analysis-specified minimum number of samples at both leaf (terminal) and split nodes.

For many of the subsequent illustrations in this chapter, classifications use scaled reflectances from a 36-band Landsat OLI time series stack from scene WRS–2 row 17 path 34 (Blacksburg, Virginia, and vicinity). Blue, green, red, NIR, SWIR1, and SWIR2 bands are used from six dates: March 21, 2014; May 8, 2014; May 24, 2014; January 19,

2015; February 4, 2015; and May 11, 2015. The training data set consists of 500 random points (excepting a few purposive water samples) labeled as either *forest* or *nonforest*.

If we constrain a CART analysis to only one level using these data, we find that a single split, on SWIR2 from May 8, 2014, enables separation of forest from nonforest with an overall accuracy of 89% (**Figures 12.11 and 12.12**).

Random Forests

As noted above, one mechanism by which overfitting can be addressed in part is through pruning from the maximal tree. Another, more robust method to address overfitting is by not relying on a single tree but rather an ensemble (forest) of trees (typically ranging in number from several hundred to several thousand). In the resulting *random forest,* each tree is created using the same techniques but using random subsets of both the samples (for each tree) and explanatory variables (for each split of each tree) (**Figure 12.13**). Using a small subset of explanatory variables at each node decorrelates the trees.

To further clarify how random forests works, we need to define two more sampling terms, *with replacement* and *without replacement*. A simple example will show the difference. Consider a set of eight colors, C = *{red, blue, green, yellow, cyan, magenta, black, white}*. We choose a random element from this set, *blue*. If we sample without replacement, *blue* is now not in the set of possible colors, so the new set (which is a subset of C) is D = *{red, green, yellow, cyan, magenta, black, white}* with seven elements. It is now not possible to choose *blue* as another element is drawn randomly. However, if we sample with replacement, and *blue* is the element selected randomly, the set of possible colors does not change; it is still C = *{red, blue, green, yellow, cyan, magenta, black, white}*.

Returning to random forests, the subset of samples for a given tree is selected randomly with replacement (*bootstrapping*), meaning it is possible for a given element to

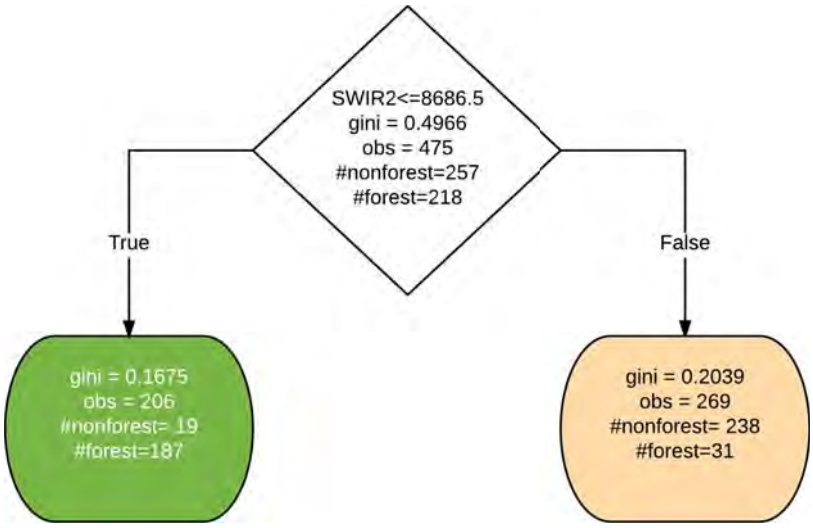


FIGURE 12.11 Classification tree from a single-level classification of forest from nonforest that uses only a spring SWIR2 band from OLI scene WRS–2 row 17 path 34 acquired May 8, 2014.

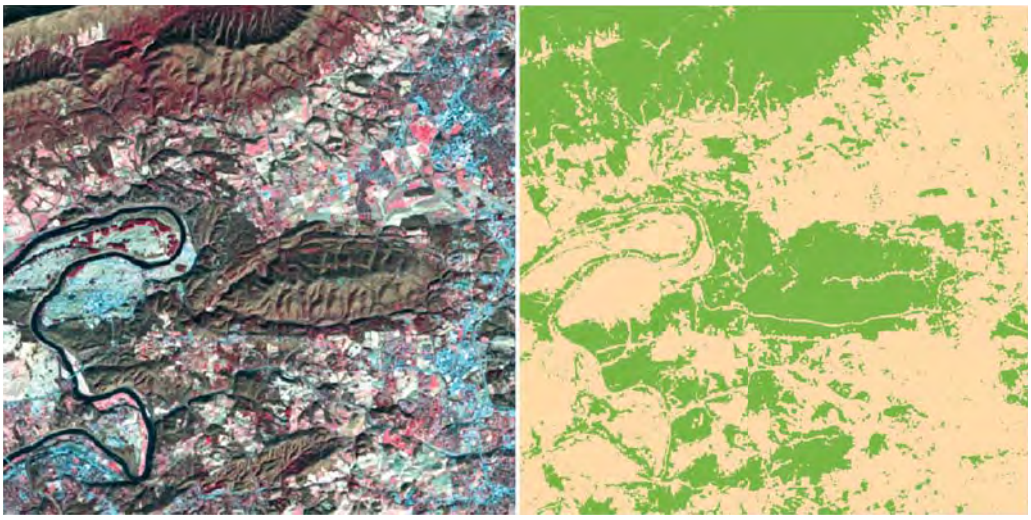


FIGURE 12.12 Classification results using the tree shown in **Figure 12.11**. On the left is a standard false-color composite from the Landsat image acquired January 19, 2015. On the right is the resulting classification. Tan is nonforest, green is forest.

be selected for that tree more than once. The subset of explanatory variables for a given node is selected without replacement. While the size of these subsets is an adjustable model parameter, typical defaults (for classification) are as follows: (1) the square root of the number of original explanatory variables is selected for each node, and (2) two-thirds (usually) to four-fifths of the data are selected for each tree. The data that are not sampled for a given tree are called the *out-of-bag* sample. Conversely, the data not in the bootstrap sample are known as the *in-bag* sample (**Figure 12.13**).

The resulting trees are different and, as such, the results for a given set of input data will not be uniform. A majority vote is common among implementations, though other options are available. To get a feel for how this would work in practice, consider a forest of 100 decision trees and a given pixel reflectance vector to which a class must be assigned in a simple three-class scheme: *forest, nonforest,* and *water.* If, for that pixel, 60 trees classify it as *forest,* 25 as *water,* and 15 as *nonforest,* then the class label will become *forest.*

The out-of-bag sample is particularly important for random forests, as it is used to calculate (1) an unbiased estimate of error as well as (2) variable importance. Remember that each tree in the forest has its own out-of-bag samples that were not used to develop the tree but for which the class is known. The flip side of this is that every training sample has a set of trees for which it was not used to train. The *out-of-bag error estimate* is then calculated as follows:

1. For a given sample, determine the majority class that is assigned for the trees that did not use that sample for construction.
2. Compare the assigned (majority) class for a given sample to the known class. If different, it is considered an error.

3. Repeat steps (1) and (2) for all samples.
4. The out-of-bag error estimate is then just the number of misclassified samples divided by the total number of samples.

Variable importance measures differ, but the most robust and widely used is based on random permutation of the values of individual features (in our case, usually spectral bands) for the out-of-bag samples for each tree. The steps are as follows:

1. For a given tree, run the out-of-bag samples down the tree and count how many out-of-bag samples were correctly classified.
2. For a given variable in a given tree, randomly permute its values (i.e., typically, scramble the existing values across pixel/object vectors).

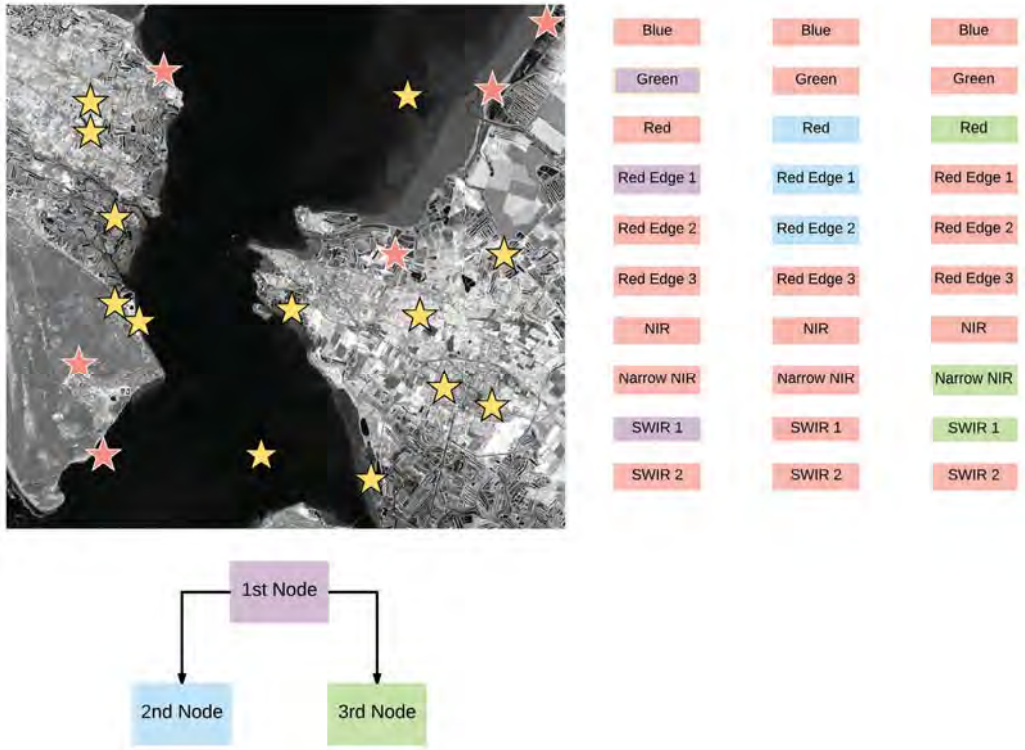


FIGURE 12.13 Sampling with replacement for a given tree in the ensemble. Eighteen sample locations are shown on the 10-m resolution near-infrared band of a Sentinel-2 MultiSpectral Instrument image acquired over western France on May 4, 2016. Twelve of these samples are the in-bag samples chosen at random for this tree, indicated by yellow stars, leaving six out-of-bag samples, indicated by pink stars. (Since 18 samples were drawn with replacement, some of the 12 samples shown in yellow will be repeated, perhaps multiple times.) For each of these samples, there are 10 spectral bands with 20-m or finer resolution. The square root of 10, rounded down, is 3; thus, three of these bands are randomly selected (without replacement) for each node, color coded here. Note that a spectral band can be selected for more than one node, as is the case with Red Edge 1 for the first two nodes and Red for nodes 2 and 3. Also recollect that only one of the selected subsamples of bands can be used for the split at a given node.

3. Run the modified out-of-bag samples (with the permuted variable) back down the tree and count how many of the out-of-bag samples with the permuted variable are correctly classified.
4. Subtract the result obtained from Step 3 from that obtained in Step 1 to get the (raw) importance of that variable. (Note that if little changes, then the variable is not very important!)
5. Repeat steps 2–4 across all trees. Average the results and calculate the standard error.
6. Compute *z*-scores by dividing the average by the standard error.
7. Repeat steps 1–6 for every variable.

A random forests classification is shown as Figure 12.14.

It should be noted that while highly correlated variables do not adversely affect the quality of random forest models, they do reduce the quality of permutation-based variable importance.

k-*Nearest Neighbors*

The kNN classifier is a simple but effective means of assigning a class to a pixel or object using the training data points closest to that pixel or object in the feature space. It requires a dense, representative (usually random) sample of labeled training data. The kNN procedure examines each pixel to be classified (Figure 12.15), then identifies the *k*-nearest training samples, as measured in feature space. Typically, *k* is set to be a relatively small integer divisor. The candidate pixel is then assigned to the class that is represented by the most samples among the *k* neighbors. The optimal *k* varies by data set and can be optimized using iterative bootstrapping (e.g., Budreski et al., 2007). Smaller *k* values sharpen boundaries; larger *k* values are needed for noisy data sets.

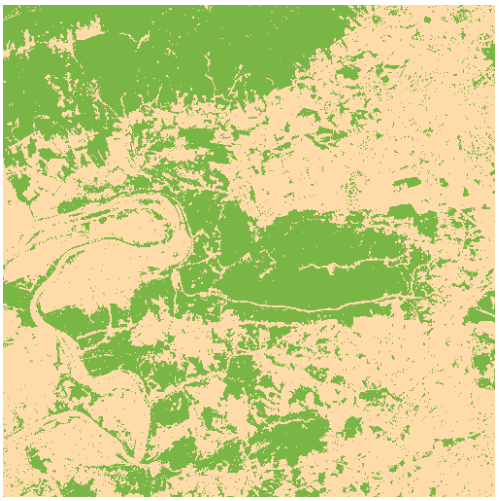


FIGURE 12.14 Classification results using a random forest. Compare to **Figure 12.12**. Tan is nonforest, green is forest.

Artificial Neural Networks

ANNs are computer programs designed to simulate human learning processes through establishment and reinforcement of linkages between input data and output data. It is these linkages, or pathways, that form the analogy with the human learning process, in that repeated associations between input and output in the training process reinforce the connections that can then be employed to link input and output, in the absence of training data.

ANNs are often represented as being composed of three elements: an input layer, an output layer, and a hidden layer. An input layer consists of the source data, which in the context of remote sensing are the multispectral observations, perhaps in several bands and from several dates. ANNs are designed to work with large volumes of data, including many bands and dates of multispectral observations, together with related ancillary data.

The output layer consists of the classes required by the analyst. There are few restrictions on the nature of the output layer, although the process will be most reliable when the number of output labels is small, or modest, with respect to the number of input channels. Included are training data in which the association between output labels and input data is clearly established. During the training phase, an ANN establishes an association between input and output data by establishing weights within one or more hidden layers (Figure 12.16) composed of virtual "neurons." Connections among neurons are, by extension, virtual "synapses." As the number of hidden layers increases, the network becomes "deeper," from which the term *deep learning* (which typically, though not always, refers to a deep neural network) arises. In the context of remote sensing, repeated associations between classes and digital values, as expressed in the training data, strengthen weights (*synaptic strength*) within hidden layers that permit the ANN to assign correct labels when given spectral values in the absence of training data.

ANNs are designed using less severe statistical assumptions than many of the usual classifiers (e.g., maximum likelihood), although in practice successful application requires careful application, particularly in the training phase. ANNs have been found to be accurate in the classification of remotely sensed data, and deep learning is at the time of this writing increasingly widely used in remote sensing, especially when both copious,

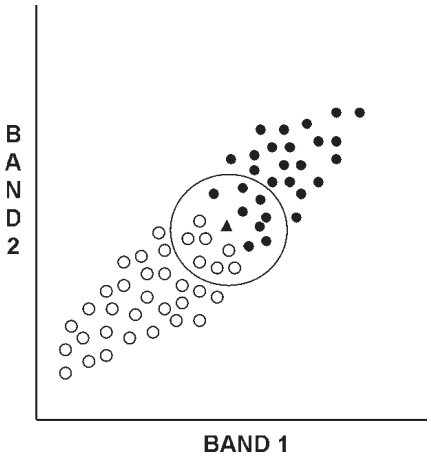


FIGURE 12.15 *k*-nearest-neighbors classifier (kNN). kNN assigns candidate pixels according to a "vote" of the *k*-neighboring pixels, with *k* determined by the analyst.

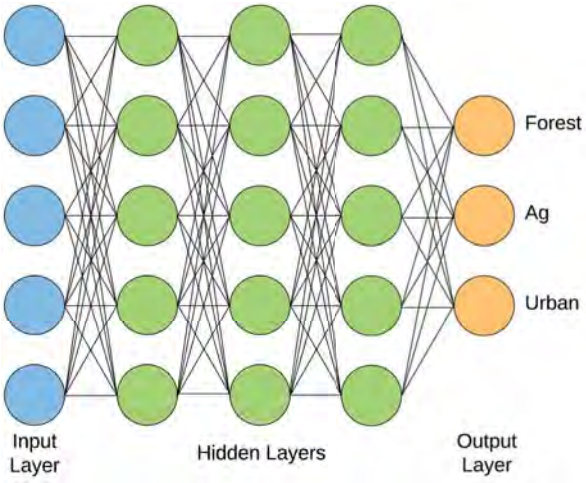


FIGURE 12.16 Artificial neural network.

high-quality training data and the necessary computational resources are available. An example classification is shown as Figure 12.17.

Parallelepiped Classification

Parallelepiped classification, sometimes also known as the box decision rule, or level-slice procedures, uses ranges of values within the training data to define regions within a multidimensional feature space. The spectral values of unclassified pixels are projected into feature space; those that fall within the regions defined by the training data are assigned to the appropriate categories.

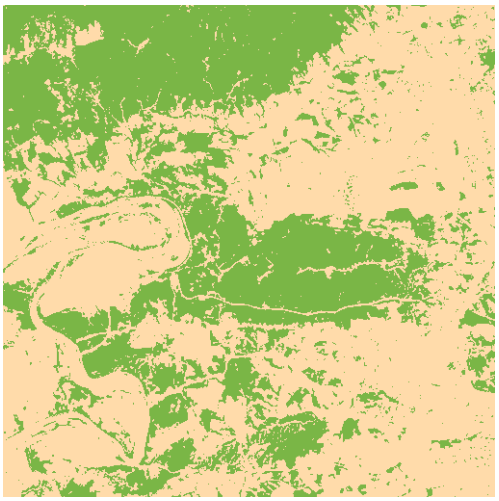


FIGURE 12.17 Classification results using a sequential artificial neural network with four epochs. The input layer has 500 neurons, the hidden layer 250 neurons, and the output layer 2 neurons. The activation for the input and hidden layers is a rectified linear unit. The activation for the output layer is softmax. Twenty percent of the neurons were randomly dropped in the input and hidden layers to avoid overfitting. Tan is nonforest, green is forest.

An example can be formed from data presented in Table 12.3. Here Landsat MSS bands 5 (Red) and 7 (NIR, 0.8–1.1 um) are selected from a larger data set to provide a concise, easily illustrated example. In practice, four or more bands can be used. The ranges of values with respect to band 5 can be plotted on the horizontal axis in Figure 12.18. The extremes of values in band 7 training data are plotted on the vertical axis and then projected to intersect with the ranges from band 5. The polygons thus defined represent regions in feature space that are assigned to categories in the classification. As pixels of unknown identity are considered for classification, those that fall within these regions are assigned to the category associated with each polygon, as derived from the training data. The procedure can be extended to as many bands, or as many categories, as necessary. In addition, the decision boundaries can be defined by the standard deviations of the values within the training areas rather than their ranges. This kind of strategy is useful because fewer pixels will be placed in an "unclassified" category (a special problem for parallelepiped classification), but it also increases the opportunity for classes to overlap in feature space.

Although this procedure for classification has the advantages of accuracy, directness, and simplicity, some of its disadvantages are obvious. Spectral regions for informational categories may intersect. The covariance among spectral bands present in most spectral classes is poorly handled. Training data may underestimate actual ranges of classification and leave large areas in feature space and on the image unassigned to informational categories. Also, the regions as defined in feature space are not uniformly occupied by pixels in each category; those pixels near the edges of class boundaries may belong to other classes. Also, if training data do not encompass the complete range of values encountered in the image (as is frequently the case), large areas of the image remain unclassified, or the

TABLE 12.3 Data for Example Shown in **Figure 12.15**

	Group A				Group B			
	Band				Band			
	1	2	3	4	1	2	3	4
	34	28	22	3	28	18	59	35
	36	35	24	6	28	21	57	34
	36	28	22	6	28	21	57	30
	36	31	23	5	28	14	59	35
	36	34	25	7	30	18	62	28
	36	31	21	6	30	18	62	38
	35	30	18	6	28	16	62	36
	36	33	24	2	30	22	59	37
	36	36	27	10	27	16	56	34
Low	34	28	18	3	27	14	56	28
High	36	36	27	10	30	22	62	38

Note: These data have been selected from a larger data set to illustrate parallelepiped classification.

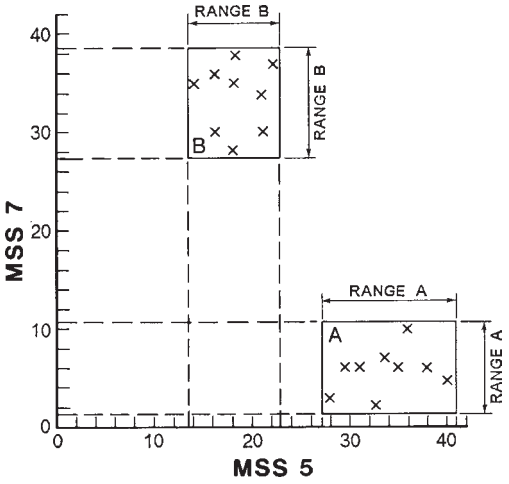


FIGURE 12.18 Parallelepiped classification. Ranges of values within training data (**Table 12.3**) define decision boundaries. Here only two spectral bands are shown, but the method can be extended to several spectral channels. Other pixels, not included in the training data, are assigned to a given category if their positions fall within the polygons defined by the training data.

basic procedure described here must be modified to assign these pixels to logical classes. This strategy was among the first used in the classification of Landsat data and is still used, although it may not always be the most effective choice for image classification.

Minimum Distance Classification

Another approach to classification uses the central values of the spectral data that form the training data to assign pixels to informational categories. The spectral data from training fields can be plotted in multidimensional feature space in the same manner illustrated previously for unsupervised classification. Values in several bands determine the positions of each pixel within the clusters that are formed by training data for each category (Figure 12.19). These clusters may appear to be the same as those defined earlier for unsupervised classification. However, in unsupervised classification, these clusters of pixels were defined according to the "natural" structure of the data. Now, for super-

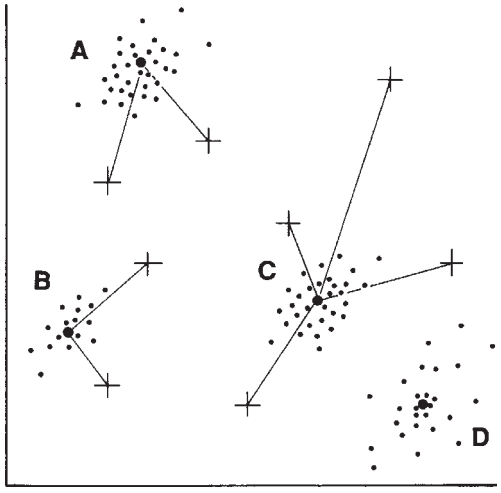


FIGURE 12.19 Minimum distance classifier. Here the small dots represent pixels from training fields, and the crosses represent examples of large numbers of unassigned pixels from elsewhere on the image. Each of the pixels is assigned to the closest group, as measured from the centroids (represented by the larger dots) using distance measures like Equation 12.1.

vised classification, these groups are formed by values of pixels within the training fields defined by the analyst.

Each cluster can be represented by its centroid, which is usually defined as its mean value. As unassigned pixels are considered for assignment to one of the several classes, the multidimensional (Euclidean) distance to each cluster centroid is calculated and the pixel is then assigned to the closest cluster. Thus, the classification proceeds by always using the "minimum distance" from a given pixel to a cluster centroid defined by the training data as the spectral manifestation of an informational class.

Minimum distance classifiers are direct in concept and in implementation but are not widely used in remote sensing work. In its simplest form, minimum distance classification is not always accurate; there is no provision for accommodating differences in variability of classes, and some classes may overlap at their edges. It is possible to devise more sophisticated versions of the basic approach just outlined by using different distance measures and different methods of defining cluster centroids.

Maximum Likelihood Classification

In nature, the classes that we classify exhibit natural variation in their spectral patterns (Figure 12.20). Further variability is added by the effects of haze, topographic shadowing, system noise, and the effects of mixed pixels. As a result, remote sensing images seldom record spectrally pure classes; more typically, they display a range of brightnesses in each band. The classification strategies considered thus far do not consider variation that may be present within spectral categories and do not address problems that arise when frequency distributions of spectral values from separate categories overlap. For example, for application of a parallelepiped classifier, the overlap of classes is a serious problem because feature space cannot then be neatly divided into discrete units for classification.

This kind of situation arises frequently because often our attention is focused on classifying those pixels that tend to be spectrally similar rather than those that are distinct enough to be easily and accurately classified by other classifiers.

Assume that we examine a digital image representing a region composed of three-fourths forested land and one-fourth cropland, and that (extremely atypically) one spectral class suffices to represent each informational class. The two spectral classes "Forest" and "Cropland" are distinct with respect to average brightness, but extreme values (very

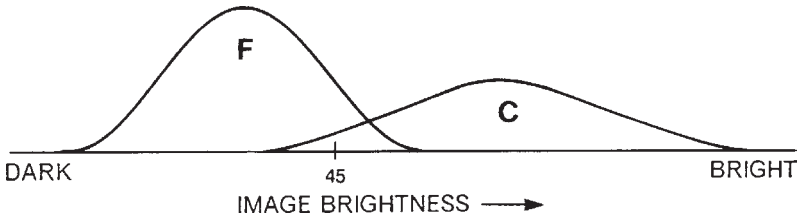


FIGURE 12.20 Maximum likelihood classification. These frequency distributions represent pixels from two training fields; the zone of overlap depicts pixel values common to both categories. The relation of the pixels within the region of overlap to the overall frequency distribution for each class defines the basis for assigning pixels to classes. Here, the relationship between the two histograms indicates that the pixel with the value "45" is more likely to belong to the *Forest* ("F") class rather than the *Crop* ("C") class.

bright forest pixels or very dark crop pixels) are similar in the region where the two frequency distributions overlap. (For clarity, Figure 12.20 shows data for only a single spectral band, although the principle extends to values observed in several bands and to more than the two classes shown here.) Brightness value "45" falls into the region of overlap, where we cannot make a clear assignment to either "Forest" or "Cropland."

In this situation, an effective classification would consider the relative likelihoods of "45 as a member of Forest" and "45 as a member of Cropland." We could then choose the class that would maximize the probability of a correct classification, given the information in the training data. This kind of strategy is known as maximum likelihood classification: it uses the training data as a means of estimating means and variances of the classes, which are then used to estimate the probabilities. Maximum likelihood classification considers not only the mean, or average, values in assigning classification, but also the variability of brightness values in each class, which need not be the same from class to class.

The maximum likelihood decision rule, implemented quantitatively to consider several classes and several spectral channels simultaneously, forms a powerful classification technique. It requires intensive calculations, so it has the disadvantage of requiring more computer resources than do most of the simpler techniques mentioned above. Also, it is sensitive to variations in the quality of training data—even more so than most other supervised techniques. Computation of the estimated probabilities is based on the assumption that both training data and the classes themselves display multivariate normal (Gaussian) frequency distributions. (This is one reason that training data should exhibit unimodal distributions, as discussed above.) Data from remotely sensed images often do not strictly adhere to this rule, although the departures are often small enough that the usefulness of the procedure is preserved. Nonetheless, training data that are not carefully selected may introduce error.

Bayes' Classification

The classification problem can be expressed more formally by stating that we wish to estimate the "probability of Forest (F), given that we have an observed digital value 45," and the "probability (P) of Cropland (C), given that we have an observed digital value 45." These questions are a form of conditional probabilities, written as "P(F|45)" and "P(C|45)," and read as "The probability of encountering category Forest, given that digital value 45 has been observed at a pixel," and "The probability of encountering category Cropland, given that digital value 45 has been observed at a pixel." That is, they state the probability of one occurrence (finding a given category at a pixel), given that another event has already occurred (the observation of digital value 45 at that same pixel). Whereas estimation of the probabilities of encountering the two categories at random (without a conditional constraint) is straightforward (here P[F] = 0.50, and P[C] = 0.50, as mentioned above), conditional probabilities are based on two separate events. From our knowledge of the two categories as estimated from our training data, we can estimate P(45|F) ("the probability of encountering digital value 45, given that we have category Forest"), and P(45|C) ("the probability of encountering digital value 45, given that we have category Cropland"). For this example, P(45|F) = 0.75, and P(45|C) = 0.25.

However, what we want to know are values for probabilities of "Forest, given that we observe digital value 45" [P(F|45)], and "Cropland, given that we observe digital value 45" [P(C|45)], so that we can compare them to choose the most likely class for the pixel. These probabilities cannot be found directly from the training data. From a purely

intuitive examination of the problem, there would seem to be no way to estimate these probabilities.

But, in fact, there is a way to estimate $P(F|45)$ and $P(C|45)$ from the information at hand. Thomas Bayes (1702–1761) defined the relationship between the unknowns $P(F|45)$ and $P(C|45)$, and the known $P(F)$, $P(C)$, $P(45|F)$, and $P(45|C)$. His relationship, now known as *Bayes' theorem,* is expressed as follows for our example:

$$P(F|45) = \frac{P(F)P(45|F)}{P(F)P(45|F) + P(C)P(45|C)} \quad \text{(EQ. 12.5)}$$

$$P(C|45) = \frac{P(C)P(45|C)}{P(C)P(45|C) + P(F)P(45|F)} \quad \text{(EQ. 12.6)}$$

In a more general form, Bayes' theorem can be written:

$$P(b_1|a_1) = \frac{P(b_1)P(a_1|b_1)}{P(b_1)P(a_1|b_1) + P(b_2)P(a_1|b_2) + \ldots} \quad \text{(EQ. 12.7)}$$

where a_1 and a_2 represent alternative results of the first stage of the experiment, and where b_1 and b_2 represent alternative results for the second stage. For our example, Bayes' theorem can be applied as follows:

$$P(F|45) = \frac{P(F)P(45|F)}{P(F)P(45|F) + P(C)P(45|C)}$$

$$= \frac{\frac{1}{2} \times \frac{3}{4}}{(\frac{1}{2} \times \frac{3}{4}) + (\frac{1}{2} \times \frac{1}{4})} = \frac{\frac{3}{8}}{\frac{4}{8}} = \frac{3}{4} \quad \text{(EQ. 12.8)}$$

$$P(C|45) = \frac{P(C)P(45|C)}{P(C)P(45|C) + P(F)P(45|F)}$$

$$= \frac{\frac{1}{2} \times \frac{1}{4}}{(\frac{1}{2} \times \frac{1}{4}) + (\frac{1}{2} \times \frac{3}{4})} = \frac{\frac{1}{8}}{\frac{4}{8}} = \frac{1}{4}$$

So we conclude that this pixel is more likely to be "Forest" than "Cropland." Almost always there will be multiple spectral classes for any given informational class, data for multiple spectral channels/dates are considered, and there are more than two categories, so this example is greatly simplified. We can extend this procedure to as many bands or as many categories as may be necessary, although the expressions become more complex than can be discussed here.

For remote sensing classification, application of Bayes' theorem is especially effective when classes are indistinct or overlap in feature space. It can also form a convenient vehicle for incorporating ancillary data into the classification, as the added information can be expressed as a conditional probability. In addition, it can provide a means of introducing the costs of misclassification into the analysis. (Perhaps an error in misassignment of a pixel to Forest is more serious than a misassignment to Cropland.) Furthermore, we can combine Bayes' theorem with other classification procedures, so, for example, most of the pixels can be assigned using a parallelepiped classifier, and then a Bayesian classifier can be used for those pixels that are not within the decision boundaries or within a region of overlap. Some studies have shown that such classifiers are very accurate (Story, Campbell, and Best, 1984).

Thus, Bayes' theorem is an extremely powerful means of using information at hand to estimate the probabilities of outcomes related to the occurrence of preceding events. The weak points of the Bayesian approach to classification are as follows: (1) prior probabilities are needed for each *spectral* rather than informational class, a much more difficult proposition in most instances, and (2) the spectral classes must (a) partition the multidimensional feature space well, (b) be separable, and (c) have multivariate normal frequency distributions. In the hands of a skilled analyst, this remains one of the most effective and accurate means of classification, albeit increasingly less common in the operational domain.

Hybrid Parametric Classifiers

The clustering algorithms used for unsupervised classification have numerous advantages, primarily (1) partitioning the feature space well with separable spectral classes that meet the distributional assumptions for parametric classifiers and (2) identifying spatially noncontiguous but spectrally homogeneous spectral classes. They have two primary disadvantages, which are the (1) difficulty in labeling the resulting spectral classes, especially as they become more numerous, and (2) nonhomogeneity with respect to informational classes (i.e., a spectral class might inherently "belong" to two or more informational classes). Hybrid classifiers retain all the benefits associated with unsupervised classification while improving (1) labeling and (2) spectral class homogeneity.

Many hybrid classification methods have been developed. Phillips, Watson, and Wynne (2007) encapsulate the general thrust of the historical development, to which the reader is referred. However, there are two methods for which development activity and/or operational use remains high: guided clustering (Bauer et al., 1994; **Figure 12.21**) and iterative guided spectral class rejection (IGSCR; Wayman et al., 2001; Musy et al., 2006; Phillips et al., 2007; **Figure 12.22**).

Guided Clustering

Guided clustering (**Figure 12.21**) requires the analyst to develop informational class (rather than spectral class) areas of interest. The polygons that constitute, in the aggregate, a given informational class, must encompass the spectral variability associated with that informational class. However (unlike supervised training), the individual spectral classes within an informational class need not be identified, greatly simplifying the process. Clusters are developed using only the pixels that are circumscribed or intersected by the polygons that constitute a given informational class. For example, if eight polygons

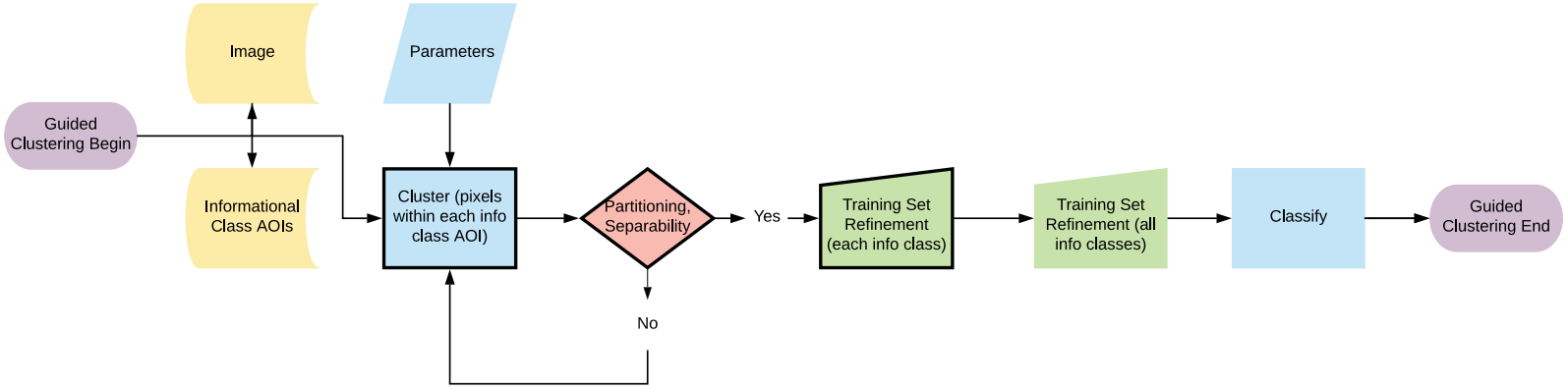


FIGURE 12.21 Guided clustering flow. Black boundaries denote informational-class specific processing.

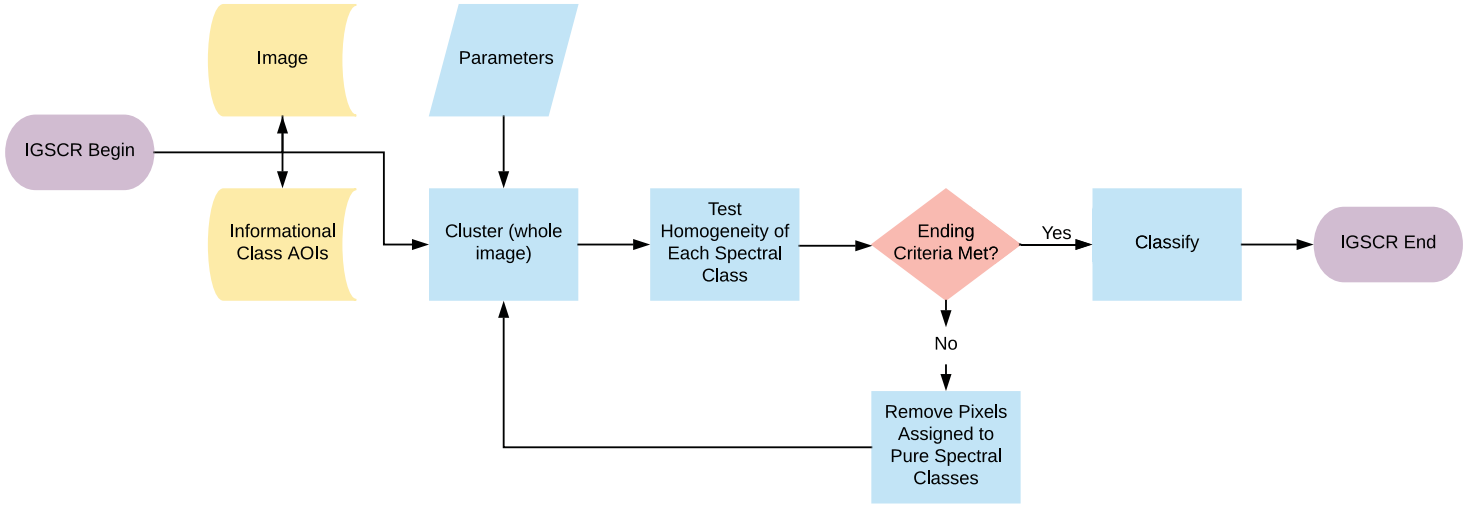


FIGURE 12.22 Iterative guided spectral class rejection (IGSCR) flow.

together encapsulate all the spectral variability within informational class Agriculture, then only the pixels that fall within (or touch) those eight polygons are used in the clustering stage. The principal user-supplied parameter is the desired number of spectral classes. The analyst is not constrained to a given clustering algorithm, though both *k*-means and ISODATA have clearly demonstrated prior utility. The resulting spectral classes are then evaluated using an informational class-specific feature space image (constructed using the same pixels) to evaluate feature space partitioning and class separability. If either of these is lacking, then the analyst is generally advised to recluster using a different number of clusters. Once the clustering result is satisfactory, then manual refinement of the clusters within an informational class is possible, though seldom necessary. After a similar process is complete for each informational class, the resulting spectral signatures are then combined. Spectral class refinement by the analyst (particularly careful deletion of confused classes across informational class boundaries) is often necessary at this stage, though it is normally quite minimal. Attempts have been made to automate even the relatively minimal user input required (e.g., Podger, 2004), but the process can be quickly implemented by an experienced analyst.

Iterative Guided Spectral Class Rejection

IGSCR also uses informational class-specific areas of interest. However, the pixels undergirding the polygons in a given informational class are used only to label clusters derived from the whole image rather than individual identified informational classes. A hypothesis test is used to determine whether each resulting spectral class is homogeneous with respect to one informational class (**Figure 12.23**). If a spectral class is not "pure" by this criterion, all the pixels in the image that were assigned to that spectral class (and others that are similarly "impure") will be reclustered. Since the specified number of spectral classes does not change between iterations, the net result is usually a continual reduction in the size of spectral classes at each iteration, particularly at the informational class boundary. The process ends (**Figure 12.24**) when (1) only pure classes are identified in a given iteration, (2) the analyst-provided maximum number of iterations is reached, or (3) there are no longer enough remaining pixels for hypothesis tests at a prespecified level of homogeneity. Analyst-provided inputs include, as noted above, the maximum number of iterations and the homogeneity threshold. The number of spectral classes to be identified at each iteration is also required. Since the method is iterative, there is no need to specify a large number of spectral classes. A sensitivity analysis (Kelly et al., 2004) revealed three spectral classes as best for their application, along with a 90 percent homogeneity threshold. Successful application of IGSCR does require that the number of training pixels in each class roughly approximates the proportionality of informational classes in the image or study area. Also, as the homogeneity threshold increases, the number of required training samples also increases, often substantially, so a lower threshold is often more practical (and, as noted above, has been shown to be as or more accurate than higher thresholds).

Ancillary Data

Ancillary, or collateral, data can be defined as data acquired by means other than remote sensing that are used to assist in the classification or analysis of remotely sensed data. There is a long tradition of both implicit and explicit use of such information in the pro-

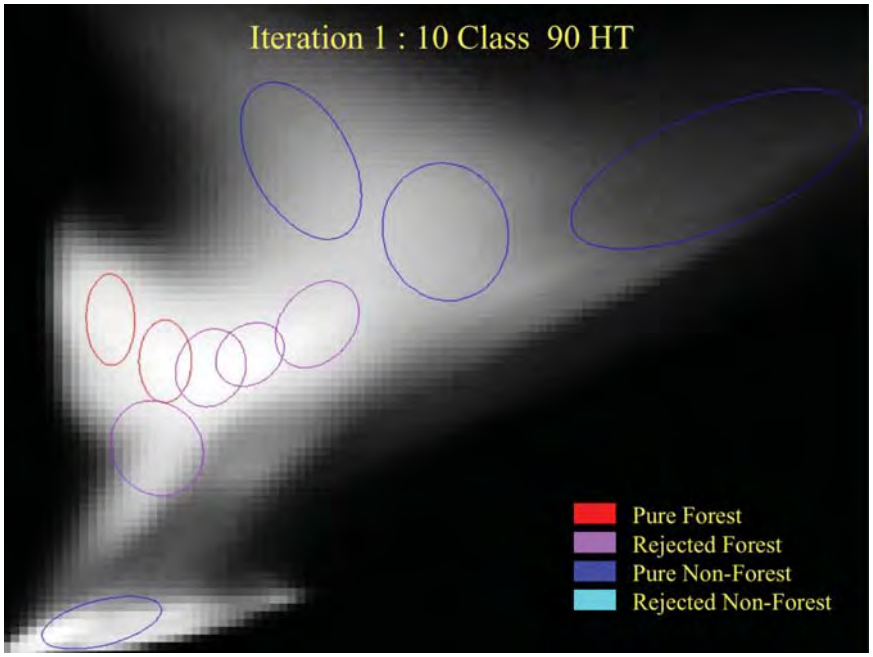


FIGURE 12.23 Feature space display of the spectral classes created as a result of the first iteration of an IGSCR classification of a Landsat Thematic Mapper (TM) image using 10 ISODATA classes and a 90% homogeneity threshold.

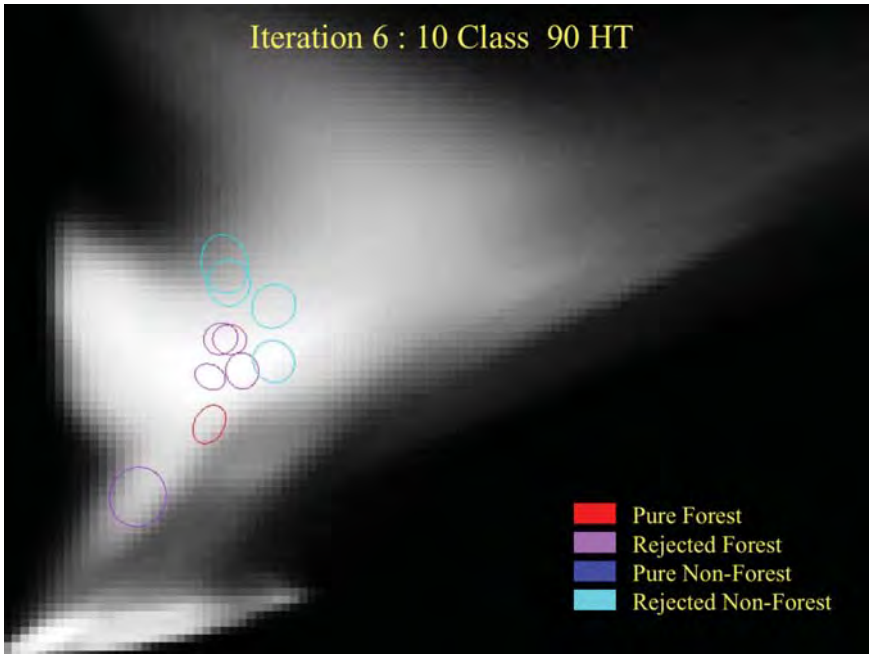


FIGURE 12.24 Feature space display of the spectral classes created as a result of the sixth iteration of an IGSCR classification of a Landsat TM image using 10 ISODATA classes and a 90% homogeneity threshold. Note the substantially smaller size of the classes at higher iterations.

cess of manual interpretation of images, including data from maps, photographs, field observations, reports, and personal experience. For digital analysis, ancillary data often consist of data available in formats consistent with the digital spectral data, or in forms that can be conveniently transformed into usable formats. Examples include digital elevation data or digitized soil maps (Anuta, 1976).

Ancillary data can be used in either of two ways. They can be added to the spectral data to form a single multiband image; the ancillary data are treated simply as additional channels of data. Or the analysis can proceed in two steps using a layered classification strategy. In either instance, ancillary data are vital to the success of many classifications, and they are integral to the production of operational products derived from remotely sensed data such as the U.S. National Land Cover Data (Homer et al., 2004; Wickham et al., 2014).

Fuzzy Classification

Fuzzy classification addresses a problem implicit in much of the preceding material: pixels must be assigned to a single discrete class. However, many processes prevent clear matches between pixels and classes, as noted by Robinove (1981) and Richards and Kelly (1984). Therefore, the focus on finding discrete matches between pixels and informational classes ensures that many pixels will be incorrectly or illogically labeled. Fuzzy logic attempts to address this problem by applying a different classification logic.

Fuzzy logic (Kosko and Isaka, 1993) has applications in many fields but has special significance for remote sensing. Fuzzy logic permits partial membership, a property that is especially significant in the field of remote sensing, as partial membership translates closely to the problem of mixed pixels. So whereas traditional classifiers must label pixels as either "Forest" or "Water," for example, a fuzzy classifier is permitted to assign a pixel a membership grade of 0.3 for "Water" and 0.7 for "Forest," in recognition that the pixel may not be properly assigned to a single class. Membership grades typical vary from 0 (nonmembership) to 1.0 (full membership), with intermediate values signifying partial membership in one or more other classes (Table 12.4).

A fuzzy (aka "soft") classifier assigns membership to pixels based on a membership function. Membership functions for classes are determined either by general relationships or by definitional rules describing the relationships between data and classes. Or,

TABLE 12.4 Partial Membership in Fuzzy Classes

	Pixel						
Class	**A**	**B**	**C**	**D**	**E**	**F**	**G**
Water	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Urban	0.00	0.01	0.00	0.00	0.00	0.00	0.85
Transportation	0.00	0.35	0.00	0.00	0.99	0.79	0.14
Forest	0.07	0.00	0.78	0.98	0.00	0.00	0.00
Pasture	0.00	0.33	0.21	0.02	0.00	0.05	0.00
Cropland	0.92	0.30	0.00	0.00	0.00	0.15	0.00

as is more likely in the instance of remote sensing classification, membership functions are derived from experimental (i.e., training) data for each specific scene to be examined. In the instance of remote sensing data, a membership function describes the relationship between class membership and brightness in several spectral bands.

Figure 12.25 provides contrived examples showing several pixels and their membership grades. (Actual output from a fuzzy classification is likely to form an image that shows varied levels of membership for specific classes.) Membership grades can be hardened (Table 12.5) by setting the highest class membership to 1.0 and all others to 0.0. Hardened classes are equivalent to traditional classifications: each pixel is labeled with a single label and the output is a single image labeled with the identity of the hardened class. Programs designed for remote sensing applications (Bezdek, Ehrilich, and Full, 1984) provide the ability to adjust the degree of fuzziness and thereby adjust the structures of classes and the degree of continuity in the classification pattern.

Fuzzy clustering has been judged to improve results, at least marginally, with respect to traditional classifiers, although the evaluation is difficult because the usual evaluation methods require the discrete logic of traditional classifiers. Thus, the improvements

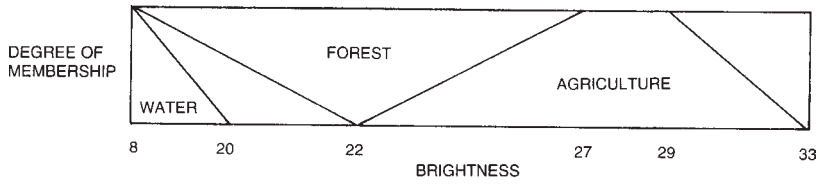


FIGURE 12.25 Membership functions for fuzzy clustering. This example illustrates membership functions for the simple instance of three classes considered for a single spectral band, although the method is typically applied to multibands. The horizontal axis represents pixel brightness; the vertical axis represents degree of membership, from low, near the bottom, to high, at the top. The class "Water" consists of pixels darker than brightness 20, although only pixels darker than 8 are likely to be completely occupied by water. The class "Agriculture" can include pixels as dark as 22 and as bright as 33, although pure "Agriculture" is found only in the range 27–29. A pixel of brightness 28, for example, can only be agriculture, although a pixel of brightness of 24 could be partially forested, partially agriculture. (Unlabeled areas on this diagram are not occupied by any of the designated classes.)

TABLE 12.5 "Hardened" Classes for Example Shown in **Table 12.4**

	Pixel						
Class	**A**	**B**	**C**	**D**	**E**	**F**	**G**
Water	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Urban	0.00	0.00	0.00	0.00	0.00	0.00	1.00
Transportation	0.00	1.00	0.00	0.00	1.00	1.00	0.00
Forest	0.00	0.00	1.00	1.00	0.00	0.00	0.00
Pasture	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Cropland	1.00	0.00	0.00	0.00	0.00	0.00	0.00

noted for hardened classifications are probably conservative, as they do not reveal the full power of fuzzy logic.

Contextual Classification

Contextual information is derived from spatial relationships among pixels or objects (with the latter being particularly common) within a given image. Context is determined by positional relationships among objects, either classified or unclassified, anywhere within the scene (Swain, Vardeman, and Tilton, 1981; Gurney and Townshend, 1983).

Although contextual classifiers can operate on either classified or unclassified data, it is convenient to assume that some initial processing has assigned a set of preliminary classes on a pixel-by-pixel basis without using spatial information. The function of the contextual classifier is to operate on the preliminary classification to reassign pixels as appropriate in the light of contextual information.

Context can be defined in several ways, as illustrated in Figure 12.26. In each instance, the problem is to consider the classification of a pixel or a set of pixels (represented by the shaded pattern) using information concerning the classes of other, related pixels. Several kinds of links define the relationships between the two groups. The simplest link is that of distance. Perhaps the unclassified pixels are agricultural land, which is likely to be "Irrigated Cropland" if positioned within a certain distance of a body of open water; if the distance to water exceeds a certain threshold, the area might be more likely to be assigned to "Rangeland" or "Unirrigated Cropland." The second example in Figure 12.26 illustrates the use of both distance and direction. Contiguity (Figure 12.26c) may be an important classification aid. For example, in urban regions, specific

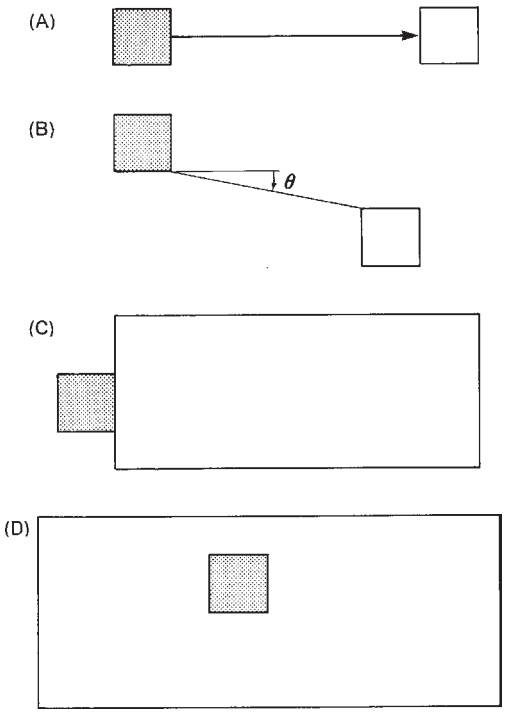


FIGURE 12.26 Contextual classification. The shaded regions depict pixels to be classified; the open regions represent other pixels considered in the classification decision. (a) Distance is considered; (b) direction is considered; (c) contiguity forms a part of the decision process; (d) inclusion is considered. From Gurney and Townshend (1983). Copyright © 1983 the American Society of Photogrammetry and Remote Sensing. Used by permission.

land uses may be found primarily in locations adjacent to a specific category. Finally, specific categories may be characterized by their positions within other categories, as shown in Figure 12.26d.

Contextual classifiers are efforts to simulate some of the higher order interpretation processes used by human interpreters, in which the identity of an image region is derived in part from its location in relation to other regions of specified identity. For example, a human interpreter considers the sizes and shapes of parcels in identifying land use, as well as the identities of neighboring parcels. The characteristic spatial arrangement of the central business district, industrial, residential, and agricultural land in an urban region permits the interpreter to identify parcels that might be indistinct, if considered with conventional classifiers.

Contextual classifiers can also operate upon classified data to reclassify erroneously classified pixels or to reclassify isolated pixels (perhaps correctly classified) that form regions so small and so isolated that they are of little interest to the user. Such uses may be essentially cosmetic operations, but they could be useful in editing the results for final presentation.

12.5 SUMMARY

This chapter has described a few specific classifiers as a means of introducing the student to the variety of classification strategies that are available today. Possibly the student may have the opportunity to use some of these procedures, so these descriptions may form the first step in a more detailed learning process. It is more likely, however, that the student who uses this book will never use many of the specific classifiers described here. Nonetheless, those procedures that are available for student use are likely to be based on the same principles outlined here using specific classifiers as examples. Therefore, this chapter should not be regarded as a complete catalog of image classification methods, but rather as an effort to illustrate some of the primary methods of image classification. Specific details and methods will vary greatly, but if the student has mastered the basic strategies and methods of image classification, he or she will recognize unfamiliar methods as variations on the fundamental approaches described here.

REVIEW QUESTIONS

1. This chapter mentions only a few of the many strategies available for image classification. Why have so many different methods been developed? Why not use just one?
2. Why might the decision to use or not to use preprocessing be especially significant for image classification?
3. Image classification is not necessarily equally useful for all fields. For a subject of interest to you (geology, forestry, etc.), evaluate the significance of image classification by citing examples of how classification might be used. Also list some subjects for which image classification might be more or less useful.
4. Speculate on the course of further developments in image classification. Can you suggest relationships between sensor technology and the design of image classification strategies?

5. Why might an analyst choose to use CART rather than random forests if the latter is better at preventing overfitting?
6. What is the difference between a parametric and nonparametric classifier? How does training differ between these two sets of classifiers?
7. Why, for parametric classifiers, should you not use a single spectral class to represent an informational class?
8. Given that the quality of training data is normally much more important than the choice of classification algorithm, how can you ensure their quality?

REFERENCES

Anuta, P. E. 1976. Digital Registration of Topographic Data and Satellite MSS Data for Augmented Spectral Analysis. In Proceedings, 42nd Annual Meeting, American Society of Photogrammetry. Falls Church, VA: American Society of Photogrammetry, pp. 180–187.

Ball, G. H., and D. J. Hall. 1965. ISODATA: A Novel Method of Data Analysis and Pattern Classification. Technical Report AD699616, Stanford Research Institute, Menlo Park, CA, 79 pp.

Bauer, M. E., T. E. Burk, A. R. Ek, P. R. Coppin, S. D. Lime, T. A. Walsh, . . . and D. F. Heinzen. 1994. Satellite Inventory of Minnesota Forest Resources. *Photogrammetric Engineering and Remote Sensing,* Vol. 60, pp. 287–298.

Bezdek, J. C., R. Ehrilich, and W. Full. 1984. FCM: The Fuzzy c-Means Clustering Algorithm. *Computers and Geosciences,* Vol. 10, pp. 191–203.

Breiman, L., J. Friedman, C. J. Stone, and R. A. Olshen. 1984. *Classification and Regression Trees*. Boca Raton, FL: Chapman and Hall/CRC, 368 pp.

Budreski, K. A., R. H. Wynne, J. O. Browder, and J. B. Campbell. 2007. *Photogrammetric Engineering and Remote Sensing,* Vol. 73, pp. 813–827.

Duda, R. O., and P. E. Hart. 1973. *Pattern Classification and Scene Analysis*. New York: John Wiley, 482 pp.

Gurney, C. M., and J. R. Townshend. 1983. The Use of Contextual Information in the Classification of Remotely Sensed Data. *Photogrammetric Engineering and Remote Sensing,* Vol. 49, pp. 55–64.

Hixson, M., D. Scholz, and N. Fuhs. 1980. Evaluation of Several Schemes for Classification of Remotely Sensed Data. *Photogrammetric Engineering and Remote Sensing,* Vol. 46, pp. 1547–1553.

Homer, C., C. Huang, L. Yang, B. Wylie, and M. Coan. 2004. Development of a circa 2000 Landcover Database for the United States. *Photogrammetric Engineering and Remote Sensing,* Vol. 70, pp. 829–840.

Hoskin, T. 2012. *Parametric and Nonparametric: Demystifying the Terms*. Rochester, MN: Mayo Clinic Department of Health Sciences Research, 5 pp. Available at *www.mayo.edu/research/documents/parametric-and-nonparametric-demystifying-the-terms/doc-20408960*.

Kelly, M., D. Shaari, Q. H. Guo, and D. S. Liu. 2004. A Comparison of Standard and Hybrid Classifier Methods for Mapping Hardwood Mortality in Areas Affected by "Sudden Oak Death." *Photogrammetric Engineering and Remote Sensing,* Vol. 70, pp. 1229–1239.

Kosko, B., and S. Isaka. 1993. Fuzzy Logic. *Scientific American,* Vol. 271, pp. 76–81.

Lawrence, R. L., and A. Wright. 2001. Rule-Based Classification Systems Using Classification and Regression Tree Analysis. *Photogrammetric Engineering and Remote Sensing,* Vol. 67, pp. 1137–1142.

Lindsay, J. S. 2017. Whitebox Tools modified_k_means_clustering.rs. Available at *https://github.com/jblindsay/whitebox-tools/blob/master/whitebox-tools-app/src/tools/image_analysis/modified_k_means_clustering.rs*.

Mather, P. M. 1976. *Computational Methods of Multivariate Analysis in Physical Geography.* Chichester, UK: John Wiley, 532 pp.

Mathieu, R., J. Aryal, and A. K. Chong. 2007. Object-Based Classification of IKONOS Imagery for Mapping Large-Scale Vegetation Communities in Urban Areas. *Sensors,* Vol. 7, pp. 2860–2880.

Musy, R. F., R. H. Wynne, C. E. Blinn, J. A. Scrivani, and R. E. McRoberts, 2006. Automated Forest Area Estimation Using Iterative Guided Spectral Class Rejection. *Photogrammetric Engineering and Remote Sensing,* Vol. 72, pp. 949–960.

Phillips, R. D., L. T. Watson, and R. H. Wynne. 2007. Hybrid Image Classification and Parameter Selection Using a Shared Memory Parallel Algorithm. *Computers and Geosciences,* Vol. 33, pp. 875–897.

Podger, N. E. 2004. *Classification, Change-Detection and Accuracy Assessment: Toward Fuller Automation.* Unpublished dissertation, Department of Environmental Monitoring, University of Wisconsin—Madison, 232 pp.

Richards, J. A., and D. J. Kelly. 1984. On the Concept of the Spectral Class. *International Journal of Remote Sensing,* Vol. 5, pp. 987–991.

Robinove, C. J. 1981. The Logic of Multispectral Classification and Mapping of Land. *Remote Sensing of Environment,* Vol. 11, pp. 231–244.

Scholz, D., N. Fuhs, and M. Hixson. 1979. An Evaluation of Several Different Classification Schemes, Their Parameters, and Performance. In Proceedings of the Thirteenth International Symposium on Remote Sensing of the Environment. Ann Arbor: University of Michigan Press, pp. 1143–1149.

Story, M. H., J. B. Campbell, and G. Best. 1984. An Evaluation of the Accuracies of Five Algorithms for Machine Classification of Remotely Sensed Data. Proceedings of the Ninth Annual William T. Pecora Remote Sensing Symposium. Silver Spring, MD: IEEE, pp. 399–405.

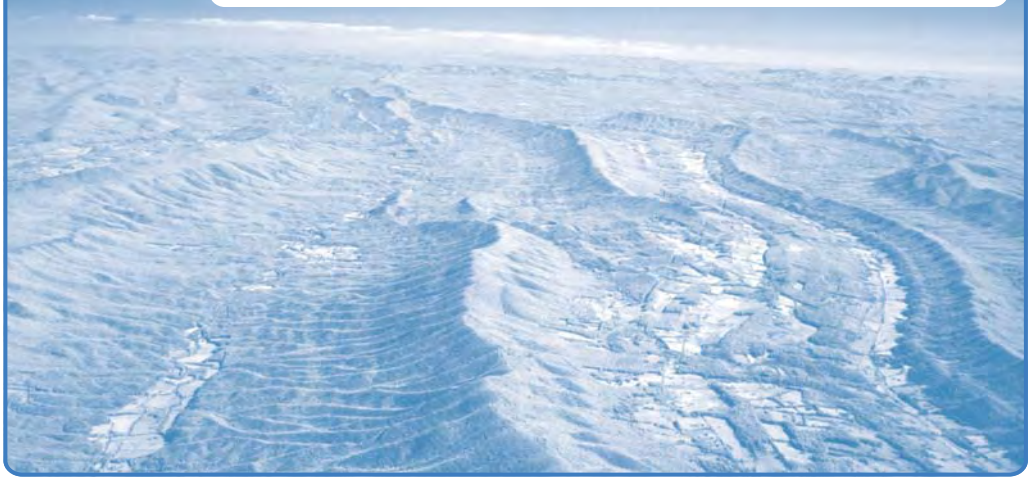
Swain, P. H., and S. M. Davis. 1978. *Remote Sensing: The Quantitative Approach.* London: McGraw-Hill, 396 pp.

Swain, P. H., S. B. Vardeman, and J. C. Tilton. 1981. Contextual Classification of Multispectral Image Data. *Pattern Recognition,* Vol. 13, pp. 429–441.

Wayman, J. P., R. H. Wynne, J. A. Scrivani, and G. A. Burns. 2001. Landsat TM-Based Forest Area Estimation Using Iterative Guided Spectral Class Rejection. *Photogrammetric Engineering and Remote Sensing,* Vol. 67, pp. 1155–1166.

Wickham, J., C. Homer, J. Vogelmann, A. McKerrow, R. Mueller, N. Herold, and J. Coulston. 2014. The Multi-Resolution Land Characteristics (MRLC) Consortium—20 Years of Development and Integration of USA National Land Cover Data. *Remote Sensing,* Vol. 6, pp. 7424–7441.

13 Accuracy Assessment



MAJOR TOPICS TO UNDERSTAND

- Sources of Classification Error
- Error Characteristics
- Measurement of Map Accuracy
- Sampling Scheme
- Cross Validation
- Sample Size
- Comparing Maps
- Area Estimation

13.1 INTRODUCTION

Prospective users of maps and data derived from remotely sensed images quite naturally ask about the accuracy of the information they will use. Yet questions concerning accuracy are surprisingly difficult to address in a convincing manner. This chapter describes how the accuracy of a thematic map can be evaluated and how two maps can be compared to determine if they are statistically different from one another.

Accuracy and Precision

Accuracy defines "correctness"; it measures the agreement between a standard assumed to be correct and a classified image of unknown quality. If the image classification corresponds closely with the standard, it is said to be "accurate." There are several methods for measuring the degree of correspondence, all of which are described in later sections.

Precision defines a measure of the sharpness (or certainty) of a measurement (Figure 13.1). In remote sensing, precision has two connotations: (1) categorical specificity of a thematic map and (2) the confidence interval within which estimates of map accuracy or area are likely to be contained. For continuous variables such as aboveground biomass or Secchi depth (a measure of water transparency), this latter connotation can also be expressed by the root mean square error or standard error of the estimator. With respect to categorical specificity, one may be able to increase accuracy by decreasing precision—that is, by being vague in the classification. For example, consider a stand of trees labeled simply as *forest*. If that class is further subdivided into *coniferous forest, mixed forest,* and *pine forest,* detail increases. If the pines are further subdivided into *shortleaf pine forest, Virginia pine forest,* and *loblolly pine forest,* detail increases even more. Each time categorical specificity increases, so does the opportunity for error. (It is clearly more difficult to assign detailed classes correctly than to assign general classes.) Evaluation of accuracy seldom explicitly considers precision, yet we must always ask if the precision is appropriate for the purpose at hand. Accuracy of 95% in separating *water* and *forest* is unlikely to be useful if we need to know the distributions of *evergreen* and *deciduous* categories.

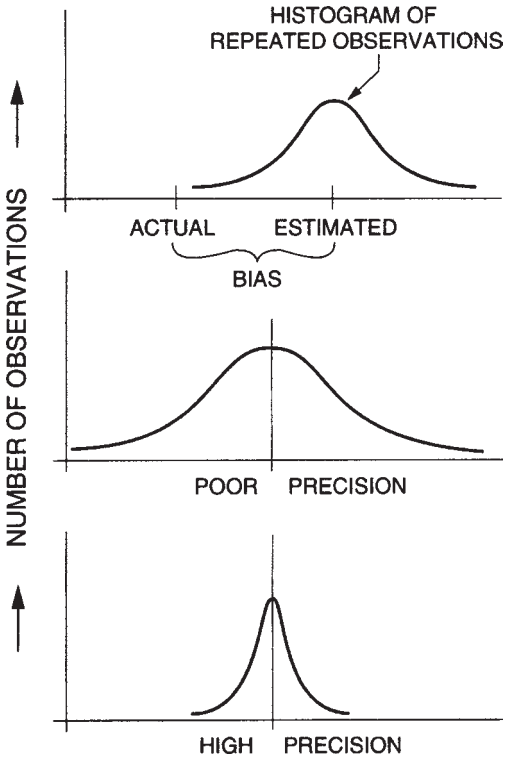


FIGURE 13.1 Bias and precision. Accuracy consists of bias and precision. Consistent differences between estimated values and true values create bias (top diagram). The lower diagram illustrates the concept of precision. High variability of estimates leads to poor precision; low variability of estimates creates high precision.

In a statistical context, high accuracy means that bias is low (that estimated values are consistently close to an accepted reference value). High precision means that the variability of estimates (independent of their bias) is low (Figure 13.1). The usefulness of a map is related not only to its correctness, but also to the precision with which the user can make statements about specific points depicted on the map. A map that offers only general classes (even if correct) enables users to make only vague statements about any given point represented on the map; one that uses detailed classes permits the user to make more precise statements (Webster and Beckett, 1968).

Significance

Accuracy has many practical implications: for example, it affects the legal standing of maps and reports derived from remotely sensed data, the operational usefulness of such data for land management, and their validity as a basis for scientific research. Analyses of accuracies of alternative classification strategies have significance for everyday uses of remotely sensed data. There have, however, been few systematic investigations of relative accuracies of manual and machine classifications, of different individuals, of the same interpreter at different times, of alternative preprocessing and classification algorithms, or of accuracies associated with different images of the same area. As a result, accuracy studies would be valuable research in both practical and theoretical aspects of remote sensing.

Often people assess accuracy from the appearance of a map, from past experience, or from personal knowledge of the areas represented. These can all be misleading, as overall accuracy may be unrelated to the map's cosmetic qualities, and often personal experience may be unavoidably confined to a few unrepresentative sites. Instead, accuracy should be evaluated through a well-defined effort to assess the map in a manner that permits quantitative measure of accuracy and comparisons with alternative images of the same area.

Evaluation of accuracies of information derived from remotely sensed images has long been of interest, but a new concern regarding the accuracies of digital classifications has stimulated research on accuracy assessment. As digital classifications were first offered in the 1970s as replacements for more traditional products, many found the methods of machine classification to be abstract and removed from the direct control of the analyst; their validity could not be accepted without evidence. This concern prompted much of the research outlined in this chapter.

Users should not be expected to accept at face value the validity of any map, regardless of its origin or appearance, without supporting evidence. We shall see in this chapter how difficult it can be to compile the data necessary to support credible statements concerning map accuracy.

13.2 SOURCES OF CLASSIFICATION ERROR

Errors are present in any classification. In manual interpretations, errors are caused by misidentification of parcels, excessive generalization, errors in registration, variations in detail of interpretation, and other factors. Perhaps the simplest causes of error are related to the mis-assignment of informational categories to spectral categories (Chapter 12). Bare granite in mountainous areas, for example, can be easily confused with the

spectral response of concrete in urban areas. However, most errors are probably more complex. Mixed pixels occur as resolution elements fall on the boundaries between landscape parcels; these pixels may well have digital values unlike either of the two categories represented and may be misclassified even by the most robust and accurate classification procedures. Such errors may appear in digital classification products as chains of misclassified pixels that parallel the borders of large, homogeneous parcels (Figure 13.2).

In this manner, the character of the landscape contributes to the potential for error through the complex patterns of parcels that form the scene. A very simple landscape composed of large, uniform, distinct categories is likely to be easier to classify accurately than one with small, heterogeneous, indistinct parcels arranged in a complex pattern. Key landscape variables are likely to include:

- Parcel size
- Variation in parcel size
- Parcel identities
- Number of categories
- Arrangement of categories
- Number of parcels per category
- Shapes of parcel
- Radiometric and spectral contrast with surrounding parcels

These variables change from one region to another (Podwysocki, 1976; Simonett and Coiner, 1971) and within a given region from season to season. As a result, errors present in a given image are not necessarily predictable from previous experience in other regions or on other dates.

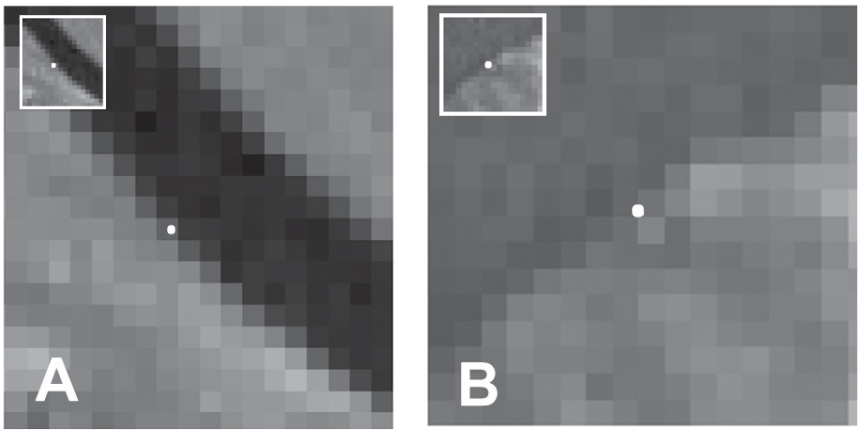


FIGURE 13.2 Incorrectly classified border pixels at the edges of parcels. Both examples, based on Musy et al. (2006), show validation points that were classified as nonforest but were labeled as forest in the reference data. The images are black-and-white infrared composites (Landsat Thematic Mapper bands 4, 3, and 2). In both cases, the forest lies at the edge of a water body.

13.3 ERROR CHARACTERISTICS

Classification error is the assignment of a pixel belonging to one category (as, for example, determined by ground observation) to another category during the classification process. There are few if any systematic studies of geographic characteristics of these errors, but experience and logic suggest that errors are likely to possess at least some of the characteristics listed below.

Errors are not distributed over the image at random but display a degree of systematic, ordered occurrence in space. Likewise, errors are not assigned at random to the various categories on the image but are likely to be preferentially associated with certain classes. Often erroneously assigned pixels are not spatially isolated but occur grouped in areas of varied size and shape (Campbell, 1981). Errors may have specific spatial relationships to the parcels to which they pertain; for example, they may tend to occur at the edges or in the interiors of parcels.

Figure 13.3 shows error patterns from estimation of a continuous variable, tree canopy cover, using random forest regression (Coulston et al., 2016). Note the following: (1) error is not random but in fact is higher in heterogeneous areas or at land cover boundaries; (2) modern analysis methods can discern the spatial distribution of error (albeit using techniques beyond the scope of this book). Figure 13.3 shows the quantity, not sources, of errors, but quantifying the magnitude of uncertainty from various sources is also now possible using modern data assimilation techniques (e.g., Das et al., 2016).

13.4 MEASUREMENT OF MAP ACCURACY

The accuracy assessment task can be defined as one of comparing two sources of information, one based on analysis of remotely sensed data (the map) and another based on a different source of information, assumed to be more accurate (the reference data). The reference data are of obvious significance; if they are in error, then the attempt to measure accuracy will be in error. For information that varies seasonally or over time, it is important that the reference data match with respect to time. Otherwise, the comparison may detect differences that are not caused solely by inaccuracies in the classification. For example, some of the differences may not really be errors, but simply changes that have occurred during the interval that elapsed between image acquisition and reference data acquisition. In some instances, we may examine two maps simply to decide if there is a difference, without concluding that one is more accurate than the other. For example, we may compare maps of the same area made from data acquired by different sensors or using different classification protocols. More commonly in such cases, however, we will use reference data applicable to both classifications to determine whether one map is more accurate than the other.

To assess the accuracy of a map, it is necessary that the map and reference data be co-registered, that they both use the same classification system and minimum mapping unit, and that they have been classified at comparable levels of detail. The strategies described here are not appropriate if the two data sources differ with respect to detail, number of categories, or meanings of the categories.

The resulting quantification of *site-specific accuracy* is based on the detailed assessment of agreement between the map and reference data at specific locations (Figure

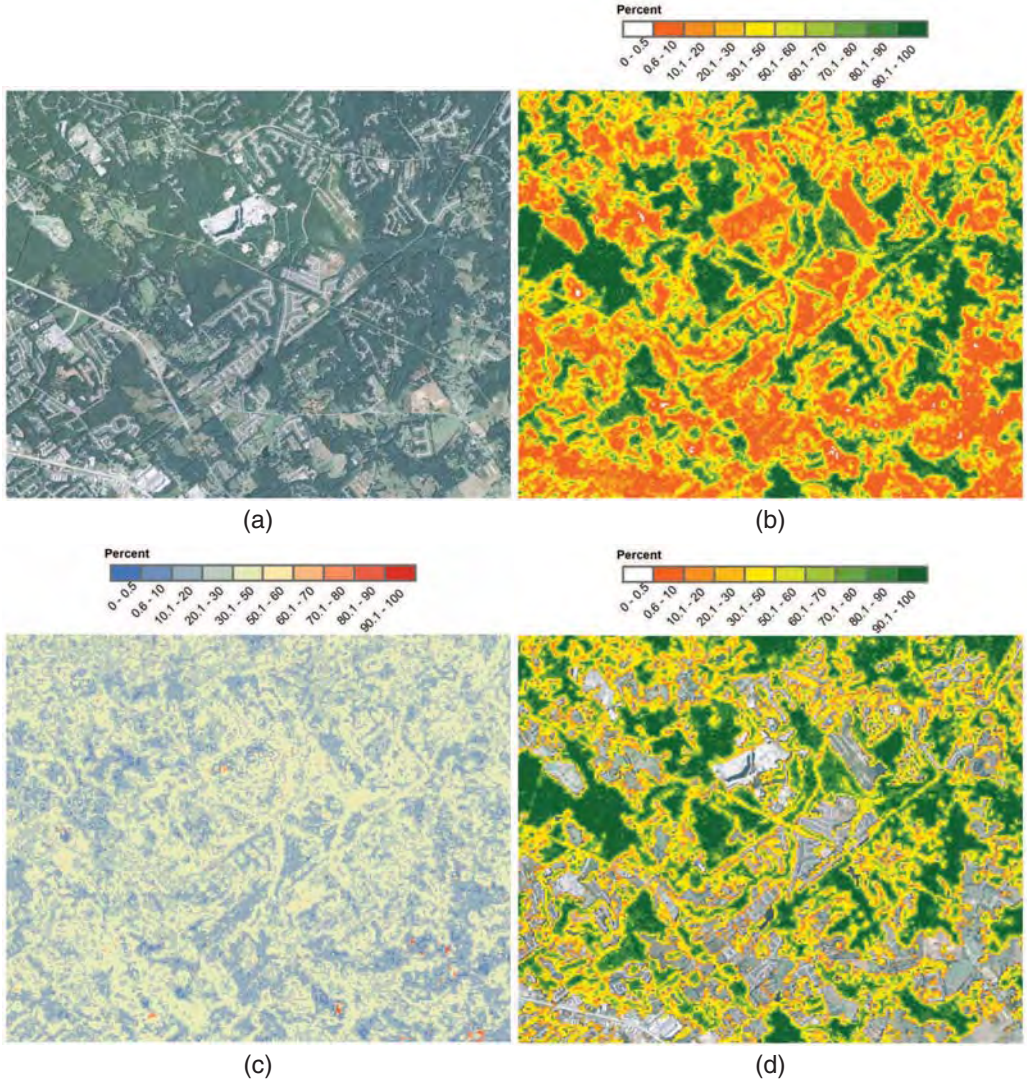


FIGURE 13.3 (a) National Agriculture Imagery Program (NAIP) imagery for an area east of Atlanta, Georgia. (b) Percent tree canopy cover based on a random forests model. (c) Half-width of the 95% prediction interval for percent tree canopy cover. (d) Masked predicted percent tree canopy cover with 5% error rate for area of no canopy cover overlaid on NAIP imagery. From Coulston et al. (2016). Copyright © 2016 the American Society for Photogrammetry and Remote Sensing. Used by permission.

13.4). Gersmehl and Napton (1982) refer to this kind of error as classification error. In many accuracy assessments of maps derived from moderate resolution imagery, the samples are pixels, but this strategy is impractical for very large (and usually mixed) or very small pixels (for which, among other issues, accurate registration between the mapped pixel and the reference data becomes more difficult). Therefore, instead of pixel-by-pixel comparisons, accuracy comparison often uses pixel groups (commonly a square 3 × 3 window) or polygons.

13.5 SAMPLING SCHEME

Once the sample unit has been chosen, the next two questions that immediately arise are: (1) What sampling scheme should be used? and (2) How many samples should be collected? With regard to sampling scheme, the most critical issue is that it needs to have a random element (i.e., be probabilistic). Remember, the goal of accuracy assessment is to estimate the accuracy of the image classification (the population, in this case) using a sample of reference data. With a large enough sample size, a simple random sample is best. However, as sample size decreases, categories that are less prevalent on the landscape can be missed using a simple random sample. For this and other reasons, stratification based on map or change categories (resulting in a stratified random sample) is often used and is a recommended best practice (Olofsson et al., 2014). However, modern machine learning workflows often collect training and validation (test) data concomitantly, making a priori stratification infeasible (see also Section 13.6). Furthermore, the stratification will

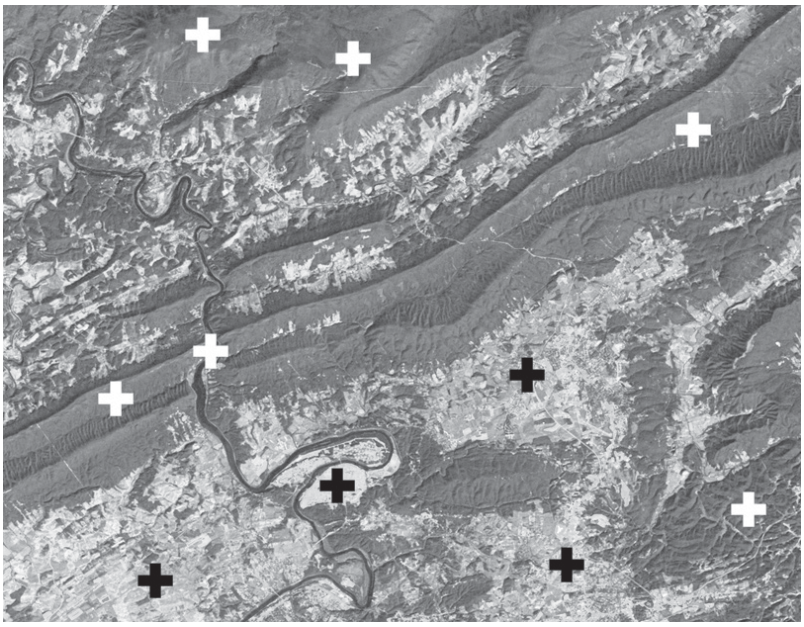


FIGURE 13.4 Site-specific accuracy. Reference data (shown here as black or white crosses) with known locations are used to assess map accuracy.

be most appropriate for the map or change pairs from which strata were developed, making the validation set potentially less appropriate for other maps of the same area (such as when one is comparing maps made by different organizations, analyses, or classification algorithms; see Section 13.8).

13.6 CROSS VALIDATION

It is possible to use the training data to assess accuracy using a process known as *cross validation*. The training data set is split into smaller sets (usually five or ten) called *folds*, in which the number of folds is denoted V (Pedregosa et al., 2011; Abu-Mostafa et al., 2017).[1] The following steps are then repeated V times, one for each separate fold:

1. Remove a fold from the training data set.
2. Use the remaining folds to train the model.
3. Assess the model accuracy using the removed fold.

The accuracies from each of the V folds are then averaged, as formalized in Equation 13.1.

$$O_{CV} = \frac{1}{V}\sum_{v=1}^{V} o_v \qquad \text{(EQ. 13.1)}$$

where v is the fold number, V = total number of folds, o_v = accuracy of a given fold, O_{CV} = the cross validation accuracy (after Abu-Mostafa et al., 2017).

While conceptually straightforward, we illustrate the process with an example in which the objective is to choose the appropriate k (how many proximate training data points are used to label an unclassified pixel or segment) in the k-nearest-neighbors classification algorithm. Ten samples were chosen from a Sentinel-2 MultiSpectral Instrument (MSI) image acquired over Blacksburg, Virginia, on November 7, 2020 (Table 13.1; Figure 13.5), and assigned (using an equalized random sampling scheme) to one of five folds (Table 13.1, Figure 13.5).

For each fold, the Euclidean distance between the two samples in that fold and the other eight samples not held out is calculated (Table 13.2).

From the computed distances, the three nearest neighbors (the three samples closest spectrally) are identified. For sample 2 in fold 1, these are, in order, sample 3 (distance = 188.1), sample 6 (distance = 493.8), and sample 7 (distance = 561.5). From Table 13.1, we note that these were labeled as *forest, forest,* and *water,* respectively (Figure 13.6). The closest neighbor corresponds to k = 1, which is *forest*. The class assignment for k = 3 is the class with the majority of the three nearest neighbors. Since two of the three closest neighbors are *forest*, the class assignment for k = 3 is also *forest*. The results from this process, completed for the other nine samples as well, is shown as Table 13.3.

Notice in Table 13.3 for fold 1 that the accuracy is 50% for k = 1. There are two samples in the fold (ID = 2 and ID = 5; Table 13.1), and only one (ID = 2; Figure 13.6)

[1]Many treatments use k, rather than V for the number of folds. Here we follow Abu-Mostafa et al. (2017), in part because of the overabundant use of k in remote sensing (k-means clustering, k-nearest neighbors, k = the number of pixels in an image, and the like).

is labeled correctly using the nearest neighbor (*1st Cls* column). This is denoted by the nonblack (blue, in this case) *F* in the *1st Cls* column. As such, the accuracy for fold 1 is 0.5 (since one sample out of the two in the fold was classified correctly).

Using the same procedure, we can obtain the accuracies for the other four folds and then use Equation 13.1 to calculate the cross-validation accuracy for $k = 1$, as shown in Equation 13.2.

TABLE 13.1 Ten Water or Forest Sample Pixels from a Sentinel-2 MSI Scene

	X	Y	Red	NIR	Class	Fold
1	513384	4144281	389	204	water	5
2	511646	4140311	761	1693	forest	1
3	511623	4140266	896	1824	forest	3
4	516345	4134132	353	585	forest	3
5	510773	4133535	311	393	forest	1
6	504604	4141436	962	2144	forest	4
7	518507	4152852	643	1144	water	4
8	513399	4152556	382	230	water	2
9	521162	4162914	558	1123	water	2
10	532973	4174855	383	274	water	5

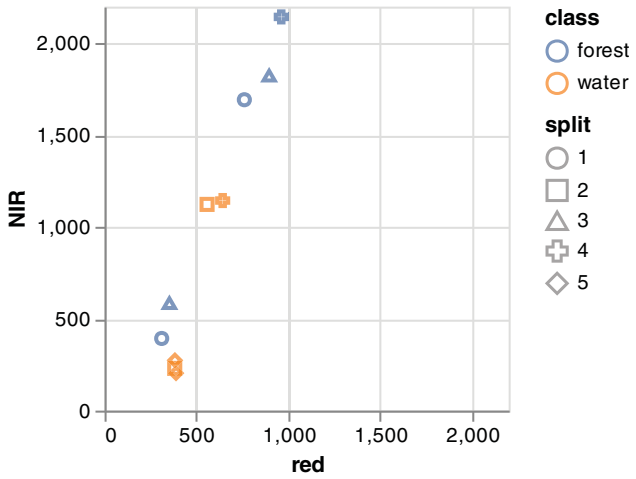


FIGURE 13.5 Two-band scatter plot of the 10 brightness value vectors from Sentinel-2 MSI image T17SNB_20201107T161459 shown in **Table 13.1.** Each pixel has been identified as either water or forest, and then the ten pixels were randomly allocated to one of five 2-pixel splits.

$$O_{CV} = \frac{1}{V}\sum_{v=1}^{V} o_v = \frac{1.0 + 0.5 + 1.0 + 1.0 + 1.0}{5} = \frac{4.5}{5} = 0.9 \quad \text{(EQ. 13.2)}$$

The calculation for $k = 3$ is shown as Equation 13.3.

$$O_{CV} = \frac{1}{V}\sum_{v=1}^{V} o_v = \frac{0.5 + 0.5 + 0.5 + 0.5 + 0.0}{5} = \frac{2.0}{5} = 0.4 \quad \text{(EQ. 13.3)}$$

As such, based on the cross-validation accuracies determined from the training data, the best choice of k is 1.

TABLE 13.2 Euclidean Distances

	1	2	3	4	5	6	7	8	9	10
1		1534.8	1697.5	382.7	204.5	2022.9	973.7	26.9	934.4	
2	1534.8		188.1	1180.7		493.8	561.5	1511.3	605.1	1468.5
3	1697.5	188.1			1546.0	326.7	725.5	1674.8	778.2	1632.7
4	382.7	1180.7			196.5	1673.7	629.7	356.2	575.7	312.4
5	204.5		1546.0	196.5		1868.1	821.1	177.8	770.7	139.1
6	2022.9	493.8	326.7	1673.7	1868.1			1999.9	1098.0	1957.6
7	973.7	561.5	725.5	629.7	821.1			950.5	87.6	908.0
8	26.9	1511.3	1674.8	356.2	177.8	1999.9	950.5			44.0
9	934.4	605.1	778.2	575.7	770.7	1098.0	87.6			866.8
10		1468.5	1632.7	312.4	139.1	1957.6	908.0	44.0	866.8	

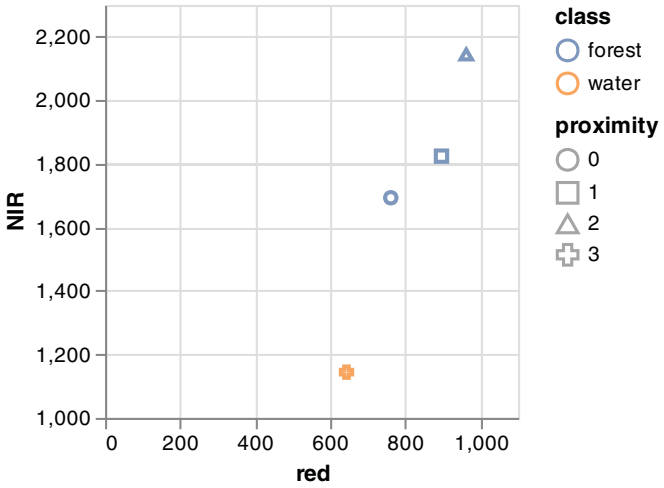


FIGURE 13.6 Three nearest neighbors of sample 2 (with proximity 0).

TABLE 13.3 Nearest Neighbors and Corresponding Class Assignments for the Samples Identified in **Table 13.1**

			Nearest neighbors			Corresponding class assignments for each neighbor			
ID	Cls	Split	1st	2nd	3rd	1st cls	2nd cls	3rd cls	$k = 3$
1	W	5	8	5	4	W	F	F	F
2	F	1	3	6	7	F	F	W	F
3	F	3	2	6	7	F	F	W	F
4	F	3	5	10	8	F	W	W	W
5	F	1	10	8	4	W	W	F	W
6	F	4	3	2	9	F	F	W	F
7	W	4	9	2	4	W	F	F	F
8	W	2	1	10	5	W	W	F	W
9	W	2	7	4	2	W	F	F	F
10	W	5	8	5	4	W	F	F	F

Note: Cls = class; F = forest; W = water. If the modeled *k*-nearest neighbor ($k = 1$ or $k = 3$) is the same as the actual class, then it is colored with respect to its fold, with the following scheme: $V = 1$, $V = 2$, $V = 3$, $V = 4$, $V = 5$.

So can cross validation replace the traditional held-out or separately created validation data set? The best practice guideline response to this question is "no," as illustrated in Figure 13.7. Cross validation is most commonly used to either select from possible machine learning algorithms or to parameterize a specific algorithm (as we just illustrated with choosing the best k between two possibilities for the k-nearest neighbors algorithm). The same data used to optimize an algorithm cannot be used to assess its accuracy. We further caution that the characteristics of the held-out samples (red test data used for final model evaluation in Figure 13.7) often must be different than the characteristics of training data used for a particular algorithm (see Budreski et al., 2007, for how this can be handled in practice). A likely scenario, for example, is that an analyst will need an equalized random sample for a random forest classification and then a stratified random sample for the accuracy assessment.

Every rule has its exceptions, of course. In the rare instance that algorithm parameterization did not use cross validation and the sampling scheme required for training of the algorithm matches that required for a robust accuracy assessment, it is possible to use cross validation for accuracy assessment. This situation, however, is increasingly unlikely given modern machine learning workflows used for advanced remote sensing analysis.

13.7 SAMPLE SIZE

So what is a "sufficient" sample size? Since a key objective of most accuracy assessments is to estimate proportional areas of each class, stratified random sampling in which map classes (or change categories) are the strata is a recommended best practice. Estimating

the needed sample size requires estimating the likely user's accuracy of each class. While at some level this is a classic "chicken-and-egg" dilemma, in practice we often have either prior experience or prior studies from which to base our estimates of user's accuracies.

Following Cochran (1977) and Olofsson et al. (2014), we can then compute the standard deviation of a given stratum as shown in Equation 13.4.

$$S_i = \sqrt{U_i\left(1-U_i\right)} \tag{EQ. 13.4}$$

where S_i is the standard deviation of the stratum (map class) i and U_i is the estimated user's accuracy for stratum (map class) i.

Assuming that the classified image being assessed has a very large number of pixels or objects, the required sample size can then be estimated using Equation 13.5.

$$n = \left(\frac{\sum_{i=1}^{M} W_i S_i}{S\left(\hat{O}\right)}\right)^2 \tag{EQ. 13.5}$$

where n = sample size, i = map class (which is also the stratum), M = total number of classes (categories), W_i is the map marginal proportion for map class i, S_i is the standard deviation of stratum i, $\hat{O}$ is the (estimate of) overall accuracy, and $S(\hat{O})$ is the desired standard error of the estimated overall accuracy.

Assume mapped area (marginal) proportions of nondisturbed forest (0.852) and disturbed forest (0.147) following Vogeler et al. (2020). From prior experience we know we have more trouble mapping the disturbances (estimated user's accuracy of 0.70) than the undisturbed forest (estimated user's accuracy of 0.90). Table 13.4 shows the computation of the stratum-specific metrics, assuming the desired standard error for overall accuracy is 0.01. In the table, n is the sum over all strata, which per Equation 13.5 is $((0.2556 + 0.0673) \div 0.01)^2 = 1043$. If strata are sufficiently large, a proportional allocation of samples to strata is warranted. For this example, the number of samples allocated to *undisturbed* would be the product of the marginal proportion (0.852) and the total

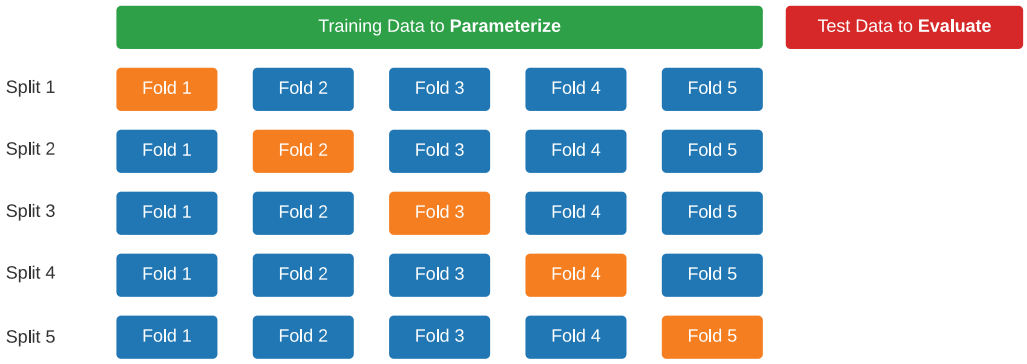


FIGURE 13.7 Cross validation and its role in model parameterization. Based on Pedregosa et al. (2011).

TABLE 13.4 Example of Sample Size Calculation for Two Map Classes

	$S(\hat{O})$	U_i	S_i	W_i	S_iW_i
Undisturbed	0.01	0.9	0.300	0.852	0.2556
Disturbed	0.01	0.7	0.458	0.147	0.0673

Note: Variables are the same as those in Equation 13.5.

sample size (1043), or 0.852 × 1043 = 889, with the other 154 samples allocated to *disturbed*.

Foody (2009) notes that one of the principal objectives of accuracy assessment is typically comparison, whether it be among classification results or against a reference standard. Statistically valid comparisons must have a sample size appropriate to the type of test, effect size (which Foody defines as the "minimum meaningful difference in accuracy"), power (see the discussion in Section 13.8, later in this chapter), significance level, and confidence limits of the comparison. In many studies, there are too few samples to permit statistically robust comparisons.

The Error Matrix

The standard form for reporting site-specific error is the error matrix in which each element is expressed as estimated proportional areas (Table 13.5; Olofsson et al., 2014; Morales-Barquero et al., 2019). Sometimes the error matrix is referred to as the confusion matrix because it identifies not only overall errors for each category, but also misclassifications (due to confusion between categories) by category. Compilation of a proportional error area matrix is required for any serious study or reporting of accuracy. The error matrix is an $M \times M$ array, where M represents the number of categories (classes) (following Richards, 2013).

The left-hand side is labeled with the categories on the map (i); the upper edge is labeled with the same M categories from the reference data (j). (Note that the meanings of the two axes can be reversed in some applications, as the convention is not universal.)

The error matrix contains a wealth of information. The diagonal from upper left to lower right (the main diagonal) shows the proportional area correctly classified for each class. Nondiagonal elements represent the areas in which the map class and reference

TABLE 13.5 Example of an Error Matrix from a Probabilistic Sample of 725 Pixels across Minnesota

	Undisturbed	Disturbed	Mapped area proportions
Undisturbed	*0.827*	0.025	0.852
Disturbed	0.002	*0.145*	0.147
Estimated area proportions	0.829	0.170	

Note: Proportional areas correctly classified are shown in italics. The forest disturbance map being assessed was created using an analysis of 6041 Landsat images spanning 1974 through 2018.
Source: Vogeler et al. (2020).

data class did not agree. Comparison of the relative magnitudes of the nondiagonal elements enables a scientist or manager to understand which classes are typically confused on the map in question. Further inspection of the matrix reveals other information. Row totals show the area proportions as mapped (also known as the *map marginal proportions*). Column totals show the *estimated area proportions* (using the reference data).

In the subsections that follow, we work through a step-by-step example on how to construct an error matrix and the descriptive statistics derived therefrom, starting with a necessary precursor, the sample count error matrix.

Precursor Sample Count Error Matrix

To construct the sample count error matrix, the analyst must compare two sources of data—the reference samples and the classified image—on a point-by-point basis to determine exactly how each of the validation samples is represented in the classification. It is very important that the validation samples and the map be well co-registered to one another. Errors in registration will appear as errors in classification, so registration problems will create errors in the assessment of accuracy. This is also true with respect to timing; the validation data and the image should be acquired as close in time as possible.

The categories for the map and reference data are compared systematically for each sample. The analyst or software maintains a count of the numbers in each reference category as they are assigned to classes on the map. A summation of this tabulation becomes the error matrix.

Let's walk through the whole process using a very simple data set and two classification examples from Chapter 12. A multitemporal stack of six Landsat-8 Operational Land Imager images has been classified into *nonforest* and *forest* using two different techniques, as follows: (1) random forests, labeled as *rf* (see **Figure 12.14**), and (2) nearest neighbor, labeled as *nn*. Forty-five samples, representing Landsat pixels at ground resolution, have been randomly chosen (as a random subset of the original random sample with an n of 475) and categorized using the same classification scheme and minimum mapping unit using a combination of aerial photography and field visits at or near the time of image acquisition.[2] The result is **Table 13.6**, where 1 = *nonforest,* and 2 = *forest*.

The resulting sample count error matrix (for *rf*) is shown as **Table 13.7**.

Image Statistics

Estimation of the sample size and conversion of the precursor sample count matrix to the estimated proportional area error matrix both require descriptive statistics from the classified image. The descriptive statistics from the random forests classification from which the samples shown in **Table 13.6** were drawn are shown as **Table 13.8**. Note that the *map marginal proportion* (W; Card, 1982; Wynne et al., 2000) for a given class i (W_i) is just the number of pixels mapped into that class divided by the total number of pixels in the classified image. In the case of nonforest in this example, $W_{nonforest}$ = 156,950 ÷ 262,144 = 0.5987. Note that the mapped area could also be divided by total area to yield the same proportion.

[2]Note that (1) computing the required number of samples (n) is a necessary precursor, and (2) a stratified random sample would be preferred over a random sample given constraints to the size of n.

TABLE 13.6 Accuracy Assessment Data Using a Probabilistic Sample

	X (m)	Y (m)	Reference	rf	nn
248	550869	4116222	1	1	2
378	552465	4112117	*1*	*2*	2
473	546894	4109051	1	1	1
118	546141	4120919	1	1	1
132	541389	4119978	1	1	2
365	543427	4112096	2	2	2
339	539828	4113160	1	1	1
198	546175	4117797	1	1	1
191	540752	4117591	1	1	1
73	542357	4121628	2	2	2
434	548706	4110531	*1*	*2*	1
321	541967	4113572	1	1	1
91	540608	4121110	2	2	1
21	537941	4123645	2	2	2
62	550445	4122835	1	1	1
392	547863	4111465	1	1	1
241	545841	4116400	2	2	2
341	541178	4112855	2	2	2
406	542431	4110950	2	2	1
228	552318	4117094	1	1	1
46	539466	4122622	1	1	1
117	544971	4120387	1	1	1
474	547151	4108705	2	1	2
16	549276	4124005	1	1	1
275	538865	4115341	2	2	2
304	545392	4114707	2	2	2
1	537854	4123848	2	2	2
17	550704	4123679	1	1	1
176	544227	4118681	2	2	2
289	549504	4114836	1	1	1
93	541801	4121164	2	2	2
144	550235	4119828	1	1	1

(continued)

TABLE 13.6 *(continued)*

	X (m)	Y (m)	Reference	rf	nn
158	545911	4118940	1	1	1
429	544421	4110387	*1*	*2*	1
246	549331	4116732	2	2	2
67	538587	4121630	2	2	2
106	551398	4121193	1	1	1
291	550672	4114842	2	2	2
416	549824	4110821	1	1	1
30	544182	4123431	2	2	2
160	548164	4119110	1	1	1
363	541527	4112665	2	2	1
303	544466	4114416	2	2	2
223	548214	4116993	2	2	2
53	544587	4122465	2	2	1

Note: Coordinates are shown in the WGS84 ellipsoid, Universal Transverse Mercator projection, zone 17 North. Note that retention of the object identifier and coordinates is important for quality control and reuse of the data set by other practitioners. 1 = nonforest; 2 = forest; rf = random forests; nn = nearest neighbor. Italicized class labels are used to create the italicized count in a cell of **Table 13.7.**

TABLE 13.7 Precursor Sample Count Error Matrix Derived from Random Forests Classification (rf) in Table 13.6

Map	Nonforest	Forest	Row totals
Nonforest	21	1	22
Forest	*3*	20	23
Column totals	24	21	45

Note: Italicized count is from samples with object IDs 378, 434, and 429 in **Table 13.6** (also italicized). Class row totals, shown in blue, are used to calculate the error matrix.

TABLE 13.8 Map Statistics from Random Forests Classification (Figure 12.14) from which Table 13.6 Was Derived

Map class (i)	Number of pixels in map class ($N_{i.}$)	Map marginal proportion (W_i)	Area (ha)
Nonforest	156950	0.5987167358398438	14125.5
Forest	105194	0.40128326416015625	9467.5
Total	262144	1.0000	23593.0

Note: For each category i, the map marginal proportion (W_i) is the number of pixels mapped into a given class ($N_{i.}$) divided by the total number of pixels in the classified image (k). The map marginal proportions, shown in red, are used in calculating the error matrix. While rounding the proportions to three or four places after the decimal point is typical for publication, these values are not rounded to prevent floating point error in subsequent calculations.

Constructing the Error Matrix

Each element of the error matrix is an unbiased estimator of the proportion of total area represented. As noted by Card (1982), for a given map class *i* this estimate is calculated as the product of the map marginal proportion (the proportion of a thematic map in a given category) for that class and the proportion of the reference data mapped as class *i* that is class *i* in the reference data (Equation 13.6).

$$\hat{p}_{ij} = W_i \frac{n_{ij}}{n_{i.}} \quad \text{(EQ. 13.6)}$$

where *i* is a map data class, *j* is a reference data class, W_i is the map marginal proportion for map data class *i*, n_{ij} is the corresponding element in the precursor sample count error matrix, and $n_{i.}$ is the row total (number of the reference samples mapped into class *i*; shown as blue in Table 13.7).

Let's look at an example using Table 13.7 and Table 13.8. For reference class *non-forest* ($j = 1$) and map class *forest* ($i = 2$), the calculation is shown as Equation 13.7. The row total is shown in blue, the map marginal proportion in red, and the corresponding element in the precursor sample count error matrix shown in italics.

$$\hat{p}_{2,1} = \pi_2 \frac{n_{2,1}}{n_{2.}} = 0.40128326416015625 \frac{3}{23} = 0.0523413 \quad \text{(EQ. 13.7)}$$

The estimate of the proportion of area in each class *j* is the column total for that class ($\hat{p}_{.j}$).

The entire error matrix is shown as Table 13.9.

Note that since the number of pixels is known, it is possible to use the *estimated area proportions* totals to estimate the number of pixels represented by each reference class *j*, ($\hat{N}_j^2$) as shown in Equation 13.8:

$$\hat{N}_j^2 = \hat{p}_{.j} k \quad \text{(EQ. 13.8)}$$

where *k* is the number of pixels in the image.

TABLE 13.9 Error Matrix Calculated Using Table 13.7 and the (Red) Map Marginal Proportions from Table 13.8

Map	Nonforest	Forest	Mapped area proportions (W_i)
Nonforest	0.5715020	0.0272144	0.59871673583984380
Forest	0.0523413	0.3489420	0.40128326416015625
Estimated area proportions ($\hat{p}_{.j}$)	0.6238440	0.3761560	1.0000

Note: While rounding the proportions to three or four places after the decimal point is typical for publication, these values are not rounded to prevent floating point error in subsequent calculations. Note that the row totals (red) are the map marginal proportions, also shown in red in **Table 13.8.** The column totals are the estimates of the proportion of area in each class.

User's and Producer's Accuracies

We are often interested in determining how well a particular category was identified in a classification. There are two measures of category-specific accuracy, user's and producer's accuracy, estimated from the error matrix using the following formulas (Equation 13.9 and Equation 13.10; Story and Congalton, 1986).

$$\hat{U}_i = \frac{\hat{p}_{ii}}{W_i} \quad \text{(EQ. 13.9)}$$

$$\hat{P}_j = \frac{\hat{p}_{jj}}{\hat{p}_{\cdot j}} \quad \text{(EQ. 13.10)}$$

where $\hat{U}_i$ is the estimate of user's accuracy (or *recall*) for map class i, $\hat{P}_j$ is the estimate of producer's accuracy (or *precision*) for reference data class j, $\hat{p}_{ii}$ and $\hat{p}_{jj}$ are the error matrix elements for a given map (i) or reference data (j) class, $\hat{p}_{ii}$ is the row total (total mapped area proportion) for map class i, and $\hat{p}_{\cdot j}$ is the column total (total estimated area proportion) for reference data class j.

Using the error matrix shown as **Table 13.9**, the user's and producer's accuracies for nonforest are shown as Equation 13.11 and Equation 13.12.

$$\hat{U}_i = \frac{\hat{p}_{ii}}{W_i} = \frac{0.5715020}{0.5987167} = 95.5\% \quad \text{(EQ. 13.11)}$$

$$\hat{P}_j = \frac{\hat{p}_{jj}}{\hat{p}_{\cdot j}} = \frac{0.5715020}{0.6238440} = 91.6\% \quad \text{(EQ. 13.12)}$$

The variances for the user's and producer's accuracies are given by the following equations (after Olofsson et al., 2014):

$$\hat{V}\left(\hat{U}_i\right) = \frac{\hat{U}_i\left(1-\hat{U}_i\right)}{n_{i\cdot}-1} \quad \text{(EQ. 13.13)}$$

$$\hat{V}\left(\hat{P}_j\right) = \frac{1}{\hat{N}_{\cdot j}^2}\left[\frac{N_{j\cdot}^2(1-\hat{P}_j)^2\,\hat{U}_j\left(1-\hat{U}_j\right)}{\left(n_{j\cdot}-1\right)} + \hat{P}_j^2\sum_{i\neq j}^{M}\frac{N_{i\cdot}^2\frac{n_{ij}}{n_{i\cdot}}\left(1-\frac{n_{ij}}{n_{i\cdot}}\right)}{n_{i\cdot}-1}\right] \quad \text{(EQ. 13.14)}$$

where each variable is as described in **Table 13.10**. The 95% confidence interval for a given category-specific accuracy is then approximated as ± 1.96 times the square root of the variance. Producer's accuracy would then be expressed as follows:

$$\hat{P}_j \pm 1.96\sqrt{\hat{V}\left(\hat{P}_j\right)}$$

with a parallel formulation for user's accuracy.

TABLE 13.10 Description of Variables Used in Variance Equations (Eq. 13.13 and Eq. 13.14)

Variable	Description
$\hat{V}(\hat{U}_i)$	Estimated variance for the user's accuracy for map class *i*
$\hat{U}_i$	Estimate of user's accuracy for map class *i*
n_i	Row total in the sample error matrix for map class *i*
$\hat{V}(\hat{P}_j)$	Estimated variance for the producer's accuracy for reference data class *j*
$\hat{P}_j$	Estimate of producer's accuracy for reference data class *j*
$\hat{N}_{\cdot j}$	Estimated marginal number of pixels of reference class *j*
$\hat{N}_{j\cdot}$	Number of pixels in map class *j*
$n_{j\cdot}$	Number of pixels in reference class *j*
M	Number of classes
$\hat{N}_{i\cdot}$	Number of pixels in map class *i*
n_{ij}	Element in sample error matrix

The estimated variance for the user's accuracy for nonforest is shown as Equation 13.15.

$$\hat{V}(\hat{U}_i) = \frac{\hat{U}_i(1-\hat{U}_i)}{n_{i\cdot}-1} = \frac{0.955(1-0.955)}{22-1} = 0.002046 \qquad \text{(EQ. 13.15)}$$

If we take the square root of the variance and multiply by 1.96, we then have the 95% confidence interval for nonforest user's accuracy, 8.9%. An example calculation for the estimated variance of the producer's accuracy is omitted for brevity.

The category-specific accuracies and their confidence intervals are shown as Table 13.11.

By examining relationships between the two kinds of errors (user's and producer's accuracies), the map user gains insight about the varied reliabilities of classes on the map, and the analyst learns about the performance of the process that generated the maps. Examined from the user's perspective, the matrix reveals user's accuracy (or *precision*). Examined from the analyst's point of view, the matrix reveals producer's accuracy (or *recall*). (The omission error is 1—producer's accuracy; the commission error is 1—user's accuracy.) The difference between the two lies in the base from which the error is assessed. For user's accuracy, the base is the area in each class on the final map. Thus, for the example in Table 13.9, producer's accuracy for nonforest is $\frac{0.5715020}{0.6238440} = 91.6\%$. For the same class, user's accuracy is $\frac{0.5715020}{0.5987167} = 95.5\%$. User's accuracy forms a guide to the reli-

TABLE 13.11 Category-Specific Accuracies with Their 95% Confidence Intervals

Category	User's accuracy	Producer's accuracy
Nonforest	95.5% ± 8.9%	91.6% ± 8.3%
Forest	87.0% ± 14.1%	92.8% ± 19.2%

ability of the map as a predictive device—it tells the user of the map that, in this example, of the area labeled *nonforest* on the map, 95.5% actually corresponds to *nonforest* on the ground. Producer's accuracy informs the analyst who prepared the classification that, of the actual nonforested landscape, 91.6% was correctly classified. In both instances, the error matrix permits identification of the nonforested areas erroneously labeled *forest* and forested areas mislabeled as *nonforest*.

F-measure

The *F-measure* (sometimes also known as the *F-score*) combines precision (user's accuracy) and recall (producer's accuracy; also known as *sensitivity*) into a single class-specific accuracy (Equation 13.16; Sokolova, Japkowicz, and Szpakowicz, 2006).

$$F = \frac{\left(\beta^2 + 1\right)(\text{recall})(\text{precision})}{\beta^2\left(\text{recall} + \text{precision}\right)} \quad \text{(EQ. 13.16)}$$

where $\beta = 1$ results in an evenly balanced harmonic mean. When $\beta > 1$, precision is favored; when $\beta < 1$, recall is favored (Sokolova et al., 2006). An evenly balanced *F-measure* (also known as *F1*) calculated for *nonforest* using the values in Table 13.11 is shown as Equation 13.17.

$$F = \frac{\left(\beta^2 + 1\right)(\text{recall})(\text{precision})}{\beta^2\left(\text{recall} + \text{precision}\right)} = \frac{\left(1^2 + 1\right)(0.916)(0.955)}{1^2\left(0.916 + 0.955\right)} = 0.935 \quad \text{(EQ. 13.17)}$$

Notice that the even balancing results in an F-measure between the producer's and user's accuracies.

Overall Accuracy

One of the most widely used measures of accuracy is the *overall accuracy*, a report of the overall proportion of area correctly classified. An estimate of overall accuracy is the sum of the diagonal elements (also called the trace) in the error matrix, not including row and column totals. Equation 13.18 expresses Table 13.9 as a matrix.

$$\mathbf{C} = \begin{bmatrix} 0.5715 & 0.0272 \\ 0.0523 & 0.3489 \end{bmatrix} \quad \text{(EQ. 13.18)}$$

The overall accuracy is the trace of the error matrix, as shown in Equation 13.19.

$$\hat{O} = \text{tr}(\mathbf{C}) \quad \text{(EQ. 13.19)}$$

where $\hat{O}$ is the (estimate of) overall accuracy and $\mathbf{C}$ is the error matrix. For our example, the overall accuracy can thus be calculated as shown in Equation 13.20.

$$\hat{O} = \text{tr}(\mathbf{C}) = 0.5715 + 0.3489 = 0.9204 \quad \text{(EQ. 13.20)}$$

Following Olofsson et al. (2014), the estimated variance for overall accuracy is given as Equation 13.21.

$$\hat{V}_{\hat{O}} = \sum_{i=1}^{M} \frac{W_i^2 \hat{U}_i \left(1 - \hat{U}_i\right)}{n_{i\cdot} - 1} \quad \text{(EQ. 13.21)}$$

where $\hat{V}_{\hat{O}}$ is the estimated variance of the overall accuracy, i = the map class, M = the number of classes, $\hat{U}_i$ is the user's accuracy, and n_i is the row total in the sample error matrix for map class i.

For the error matrix shown as Table 13.9, the computations to calculate the estimated variance for the overall accuracy are shown as Equation 13.22, Equation 13.23, and Equation 13.24 (class 1 is *nonforest* and class 2 is *forest*.)

$$\hat{V}_1 = \frac{W_1^2 \hat{U}_1 (1 - \hat{U}_1)}{n_{1\cdot} - 1} = \frac{(0.59872^2)(0.95454)(1 - 0.95454)}{22 - 1} = 0.000740716 \quad \text{(EQ. 13.22)}$$

$$\hat{V}_2 = \frac{W_2^2 \hat{U}_2 (1 - \hat{U}_2)}{n_{2\cdot} - 1} = \frac{(0.40128^2)(0.86956)(1 - 0.86956)}{23 - 1} = 0.0008302 \quad \text{(EQ. 13.23)}$$

$$\hat{V}_{\hat{O}} = \hat{V}_1 + \hat{V}_2 = 0.001570916 \quad \text{(EQ. 13.24)}$$

The 95% confidence interval is then 1.96 times the square root of the variance, or 7.8%. As such, the overall accuracy is 92.0% ± 7.8%.

Often, the overall accuracy is used alone, without the error matrix, as a simple measure of accuracy. By itself, the overall accuracy may suggest the relative effectiveness of a classification, but in the absence of an opportunity to examine the full error matrix, it cannot provide convincing evidence of the classification's accuracy. A full evaluation must consider the categories used in the classification. For example, it would be easy to achieve high values of overall accuracy by classifying a scene composed chiefly of open water—a class that is easy to classify correctly. Furthermore, variations in the accuracies of specific classes should be noted, as should the precision of the classes. A classification that used only broadly defined classes could achieve high accuracies but would not be useful for someone who required more detail.

Hay (1979) stated that it is necessary to consider five questions to thoroughly understand the accuracy of a classification:

- What proportion of the classification decision is correct?
- What proportion of assignments to a given category is correct?
- What proportion of a given category is correctly classified?
- Is a given category overestimated or underestimated?
- Are errors randomly distributed?

"Overall accuracy" can answer only the first of these questions; the others can be answered only by examination of the full error matrix.

Deprecation of Kappa

We caution that calculation of κ (a nominal adjustment for the effect of random chance on classification accuracy; Cohen, 1960) has been deprecated by the remote sensing community (Pontius and Millones, 2011; Oloffson et al., 2014). We no longer recommend its

use. The reasons that computation of κ is no longer a recommended practice include the following:

1. κ is redundant due to its strong correlation with overall accuracy (Pontius and Millones, 2011).
2. Chance agreement is often overestimated (Foody, 1992).
3. κ underestimates the probability that a given pixel or object is correctly classified (Olofsson et al., 2014).

13.8 COMPARING MAPS

There is often a need to compare maps, particularly when an analyst is ensuring the quality and representativeness of training data or comparing classification algorithms or their parameterizations (though cross validation can often be used for both algorithm comparison and parametrization within a given algorithm, as noted earlier). For example, we may ask if two maps are in agreement—a question that is difficult to answer because the notion of "agreement" may be difficult to define and implement. The error matrix is an example of a more general class of matrices, known as contingency tables, which summarize classifications analogous to those considered here. Some of the procedures that have been developed for analyzing contingency tables can be applied to examination of the error matrix.

The McNemar test (McNemar, 1947; Foody, 2004, 2009) is extensively used as a nonparametric method for comparing thematic maps using the same validation data set (since sample independence is not required). The test requires a particular "cross tabulation" of the thematic maps as shown in Table 13.12.

The McNemar calculation of z (the standard score) is then given as Equation 13.25.

$$z = \frac{b - c}{\sqrt{b + c}} \qquad \text{(EQ. 13.25)}$$

Let's look at an example using the data in Table 13.13 with a column added to denote the appropriate matrix element (denoted by letter *a*, *b*, *c*, or *d*) in Table 13.12 (with the caveat that the sample size is far too small). Let's look at the first and second rows as examples. In the first row, both maps are correct (both they and the validation point were placed in category 1, *nonforest*. Looking at Table 13.13, if both maps are correct, then the appropriate matrix element is *a*, so we place an *a* in the *Letter* column. In the second row, the first map is correct and the second map is not. In Table 13.13, this would make it matrix element *c*.

TABLE 13.12 McNemar's Test Cross Tabulation

		Map 1	
	Map 2	**Correct**	**Incorrect**
	Correct	a	b
	Incorrect	c	d

TABLE 13.13 Accuracy Assessment Data Labeled for McNemar Test

Point number	Map 1	Map 2	Reference	Letter
1	1	1	1	a
2	1	2	1	c
3	1	1	1	a
4	1	1	2	d
5	1	1	1	a
6	1	2	1	c
7	1	1	1	a
8	2	2	2	a
9	2	1	1	b
10	2	2	2	a
11	2	2	1	d
12	2	2	1	d
13	2	2	2	a
14	2	1	2	c
15	2	1	1	b
16	2	2	2	a
17	2	2	2	a
18	2	2	2	a
19	2	2	2	a
20	2	2	2	a

The McNemar statistic is now easily calculated using Equation 13.25 and the number of *b*'s (2) and *c*'s (3) we found in constructing Table 13.13, as shown in Equation 13.26.

$$z = \frac{b-c}{\sqrt{b+c}} = \frac{2-3}{\sqrt{2-3}} = -0.45 \qquad \text{(EQ. 13.26)}$$

Let's assume, per Foody (2009), that the maps will be stated to be statistically different from one another at α (the predetermined probability of making a Type I error, defined in this instance as incorrectly stating that the maps are different when in fact they are not) = 0.05 if $|z| > 1.96$ (two-tailed test). As such, since $|-0.45|$ is not greater than 1.96, the maps are not different from one another. It should be noted that while this is a common α level, there is no one level that is appropriate for all studies. Besides $\alpha = 0.05$, another common value for α is 0.01, in which case (for the two-tailed test) $|z|$ must be greater than 2.58.

While often ignored by researchers and practitioners, Type II error (its probability denoted by β) occurs when there really is a noteworthy difference between maps that is not detected. The power of the test is then 1 – β. As one increases the desired power and decreases the desired α, a larger sample size is required. Many, if not most, accuracy assessments contain too few samples, though it is possible (when, say, using another map as reference data) to have so many samples that very small differences between and among maps are declared to be statistically significant (Foody, 2009).

13.9 AREA ESTIMATION

Category-specific areas from thematic maps are used in many operational programs. For example, funds to fight forest fires are allocated in many U.S. states based on how much forest there is in a given county. However, in almost every conceivable instance, a thematic map will not be 100% accurate. As such, category-specific areas will not be either. Fortunately, there are long-established methods (e.g., Card, 1982, and Wynne et al., 2000; both based on Cochran, 1977) by which corrected area estimates (expressed in either areal units or proportion of the mapped area) and their precision (usually expressed as confidence intervals, Figure 13.8) can be calculated using a thematic map and its error matrix. These methods are well established and long utilized (e.g., Musy et al., 2006). Figure 13.8 illustrates a desired outcome, where either area or proportion of the total mapped area is represented along with a measure of confidence (here shown, as is typical, with a 95% confidence interval).

The now standard practice of expressing the error matrix as estimated proportional areas for each element (e.g., Table 13.5) makes estimation of the area proportions in each class a straightforward sum of the reference data columns ($\hat{p}_{\cdot j}$). The area represented by each class is then just the corresponding portion of the total area, shown as Equation 13.27 (after Olofsson et al., 2013).

$$\hat{A}_j = A_{tot}\,\hat{p}_{\cdot j} \qquad \text{(EQ. 13.27)}$$

where A_{tot} is the total area, $\hat{A}_j$ is the estimated area in class j, and $\hat{p}_{\cdot j}$ is the sum of the error matrix column for reference class j.

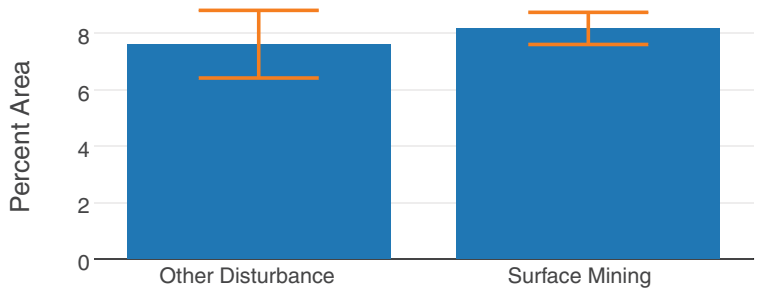


FIGURE 13.8 Proportions (± 95% confidence intervals) for the Landsat-mapped disturbance categories in Virginia's southwestern coalfields, 1984–2011. Data from Li et al. (2015).

The standard error of the area proportion (also assuming either simple random or stratified random sampling) is shown as Equation 13.28.

$$S\left(\hat{p}_{\cdot j}\right) = \sqrt{\sum_{i=1}^{M} \frac{W_i \hat{p}_{ij} - \hat{p}_{ij}^2}{n_{i\cdot} - 1}} \qquad \text{(EQ. 13.28)}$$

where $\hat{p}_{\cdot j}$ is the standard error of the area proportion, M = the number of classes, W_i is the map marginal proportion for map data class i, $\hat{p}_{ij}$ is an element of the error matrix (in which each element is expressed as estimated proportional areas) for map data class i and reference data class j, and n_i is the row total in the precursor sample count error matrix.

Figure 13.9 shows the area and 95% confidence intervals for the worked example.

13.10 SUMMARY

Accuracy assessment is a complex process. This chapter cannot address all the relevant topics in detail because even the most complete discussion leaves many issues unresolved. Research continues, and despite agreement on many important aspects of accuracy evaluation, other issues are likely to be debated for a long time before they are resolved. For example, there is disagreement concerning the best way to compare two thematic maps.

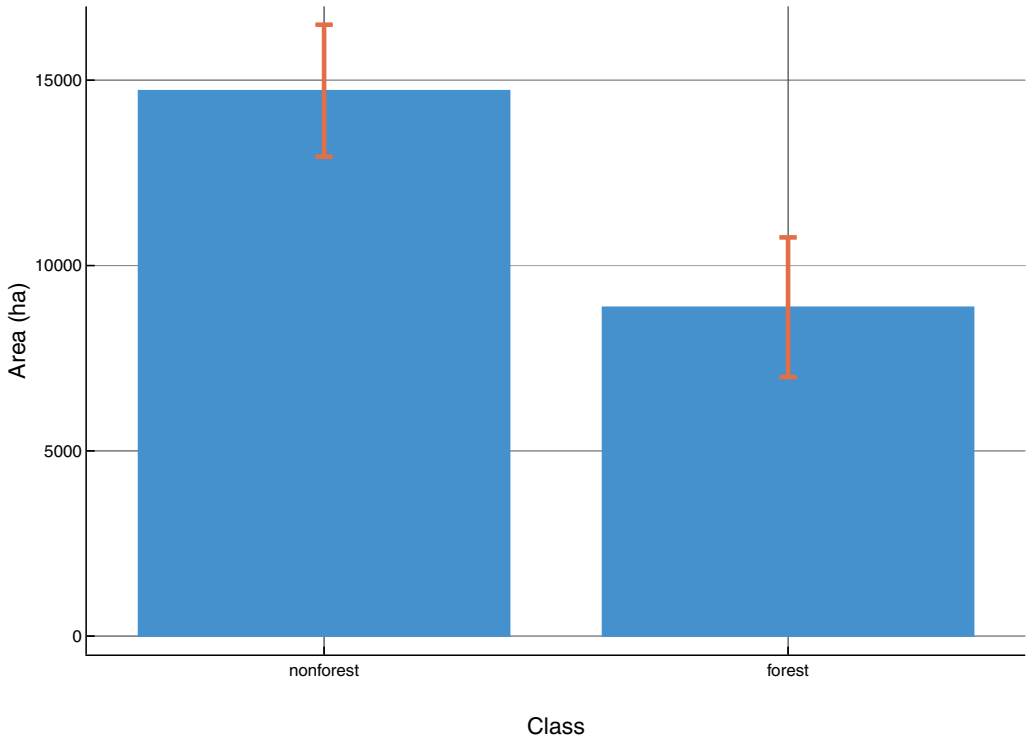


FIGURE 13.9 Area (± 95% confidence intervals) for the worked example (Landsat-mapped forest and nonforest) using the sample (**Table 13.6**) and image (**Table 13.8**) data.

For many problems, there may be no single "correct" way to conduct the analysis, but we may be able to exclude some alternatives and to speak of the relative merits and shortcomings of others.

This chapter should provide the background to assess the accuracies of classifications using procedures that, if not perfect, are at least comparable in quality to those in common use today. Furthermore, the student should now be prepared to read some of the current research on this topic and possibly to contribute to improvements in the study of accuracy assessment. Many of the problems in this field are difficult but are not beyond the reach of interested and informed students.

REVIEW QUESTIONS

1. Why is a probabilistic sample necessary for accuracy assessment?
2. Typically, one cannot use cross validation for accuracy assessment. Why?
3. What might be the consequence if we had no effective means of assessing the accuracy of a classification of a remotely sensed image?
4. Why is use of the kappa statistic no longer recommended?
5. What are the implications of knowing the confidence intervals for overall and class-specific accuracies?
6. Accuracy assessment and area estimation are now intertwined. Why was this an important development?
7. Discuss key considerations (with particular emphasis on the need to estimate user's accuracies beforehand) for deciding the sample size appropriate for accuracy assessment of a specific remote sensing project.
8. Why is the F-score a useful category-specific metric?

REFERENCES

Abu-Mostafa, Y. S., M. Magdon-Ismail, and H.-T. Lin. 2017. *Learning from Data*. Albany, NY: AMLBook, 213 pp.

Budreski, K. A., R. H. Wynne, J. O. Browder, and J. B. Campbell. 2007. Comparison of Segment and Pixel-Based Non-Parametric Land Cover Classification in the Brazilian Amazon Using Multitemporal Landsat TM/ETM+ Imagery. *Photogrammetric Engineering and Remote Sensing*, Vol. 73, pp. 813–827.

Campbell, J. B. 1981. Spatial Correlation Effects upon Accuracy of Supervised Classification of Land Cover. *Photogrammetric Engineering and Remote Sensing*. Vol. 47, pp. 355–363.

Card, D. H. 1982. Using Known Map Category Marginal Frequencies to Improve Estimates of Thematic Map Accuracy. *Photogrammetric Engineering and Remote Sensing*, Vol. 48. pp. 431–439.

Cochran, W. G. 1977. *Sampling Techniques* (3rd ed.). New York: John Wiley, 428 pp.

Cohen, J. 1960. A Coefficient of Agreement for Nominal Scales. *Educational and Psychological Measurement*, Vol. 20, No. 1, pp. 37–40.

Congalton, R. G., and K. Green. 2009. *Assessing the Accuracy of Remotely Sensed Data: Principles and Practices* (2nd ed.). Boca Raton, FL: CRC Press, 183 pp.

Coulston, J. W., C. E. Blinn, V. A. Thomas, and R. H. Wynne. 2016. Approximating Predic-

tion Uncertainty for Random Forest Regression Models. *Photogrammetric Engineering and Remote Sensing.* Vol. 82, pp. 189–197.

Das, N. N., D. Entekhabi, R. S. Dunbar, E. G. Njoku, and S. H. Yueh. 2016. Uncertainty Estimates in the SMAP Combined Active-Passive Downscaled Brightness Temperature. *IEEE Transactions on Geoscience and Remote Sensing,* Vol. 54, pp. 640–650.

Foody, G. M. 1992. On the Compensation for Chance Agreement in Image Classification Accuracy Assessment. *Photogrammetric Engineering and Remote Sensing,* Vol. 58, pp. 1459–1460.

Foody, G. M. 2004. Thematic Map Comparison: Evaluating the Statistical Significance of Differences in Classification Accuracy. *Photogrammetric Engineering and Remote Sensing,* Vol. 70, pp. 627–633.

Foody, G. M. 2009. Sample Size Determination for Image Classification Accuracy Assessment and Comparison. *International Journal of Remote Sensing,* Vol. 30, pp. 5273–5291.

Gersmehl, P. J., and D. E. Napton. 1982. Interpretation of Resource Data: Problems of Scale and Spatial Transferability. Proceedings of the Urban and Regional Information Systems Association, Minneapolis, Minnesota, pp. 471–482.

Hay, A. 1979. Sampling Designs to Test Land Use Map Accuracy. *Photogrammetric Engineering and Remote Sensing,* Vol. 45, pp. 529–533.

Li, J., C. E. Zipper, P. F. Donovan, R. H. Wynne, and A. J. Oliphant. 2015. Reconstructing Disturbance History for an Intensively Mined Region by Time-Series Analysis of Landsat Imagery. *Environmental Monitoring and Assessment,* Vol. 187, Article No. 557, 17 pp.

McNemar, Q. 1947. Note on the Sampling Error of the Difference Between Correlated Proportions or Percentages. *Psychometrika,* Vol. 12, pp. 153–157.

Morales-Barquero, L., M. Lyons, S. Phinn, and C. Roelfsema. 2019. Trends in Remote Sensing Accuracy Assessment Approaches in the Context of Natural Resources. *Remote Sensing,* Vol. 11, Article No. 2305.

Musy, R. F., R. H. Wynne, C. E. Blinn, J. A. Scrivani, and R. E McRoberts. 2006. Automated Forest Area Estimation Using Iterative Guided Spectral Class Rejection. *Photogrammetric Engineering and Remote Sensing,* Vol. 72, No. 8, pp. 949–960.

Olofsson, P., G. M. Foody, M. Herold, S. V. Stehman, and C. E. Woodcock. 2013. Making Better Use of Accuracy Data in Land Change Studies: Estimating Accuracy and Area and Quantifying Uncertainty Using Stratified Estimation. *Remote Sensing of Environment,* Vol. 129, pp. 122–131.

Olofsson, P., G. M. Foody, M. Herold, S. V. Stehman, C. E. Woodcock, and M. A. Wulder. 2014. Good Practices for Estimating Area and Assessing Accuracy of Land Change. *Remote Sensing of Environment,* Vol. 148, pp. 42–57.

Pedregosa, F., G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, . . . and E. Duchesnay, 2011. Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research,* Vol. 12, pp. 2825–2830.

Podwysocki, M. H. 1976. An Estimate of Field Size Distribution for Selected Sites in Major Grain Producing Countries. Publication No. X-923–76–93. Greenbelt, MD: Goddard Space Flight Center, 34 pp.

Pontius, R. G., and M. Millones. 2011. Death to Kappa: Birth of Quantity Disagreement and Allocation Disagreement for Accuracy Assessment. *International Journal of Remote Sensing,* Vol. 32, pp. 4407–4429.

Richards, J. A. 2013. *Remote Sensing Digital Image Analysis* (5th ed.). Berlin: Springer, 494 pp.

Simonett, D. S., and J. C. Coiner. 1971. Susceptibility of Environments to Low Resolution Imaging for Land Use Mapping. In Proceedings of the Seventh International Symposium on Remote Sensing of Environment. Ann Arbor: University of Michigan Press, pp. 373–394.

Sokolova, M., N. Japkowicz, and S. Szpakowicz. 2006. Beyond Accuracy, F-score and ROC: A Family of Discriminant Measures for Performance Evaluation. *AI 2006: Advances in Artificial Intelligence*. Berlin: Springer-Verlag, pp. 1015–1021.

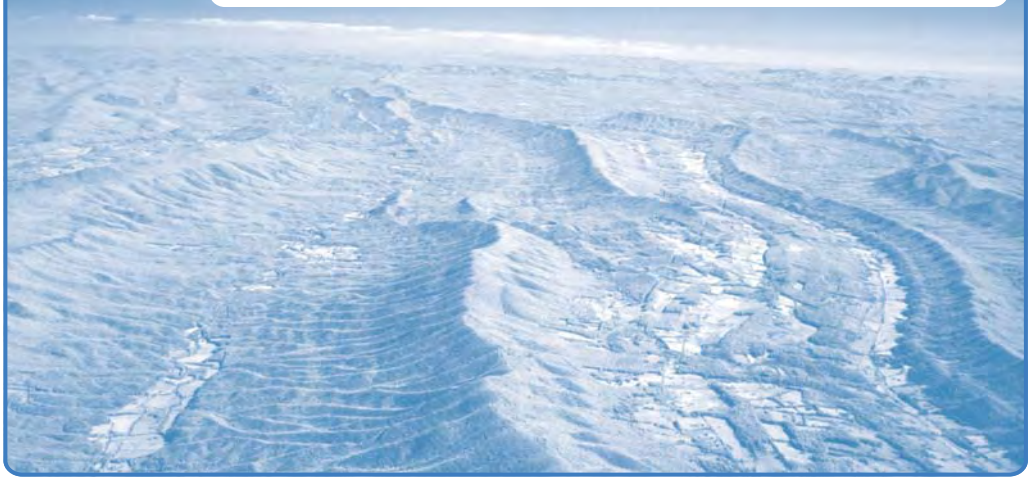
Story, M., and R. G. Congalton. 1986. Accuracy Assessment: A User's Perspective. *Photogrammetric Engineering and Remote Sensing,* Vol. 52, No. 3, pp. 397–399.

Vogeler, J., R. Slesak, P. Fekety, and M. Falkowski. 2020. Characterizing over Four Decades of Forest Disturbance in Minnesota, USA. *Forests,* Vol. 11, Article No. 362.

Webster, R., and P. H. T. Beckett. 1968. Quality and Usefulness of Soil Maps. *Nature,* Vol. 219, pp. 680–682.

Wynne, R. H., R. G. Oderwald, G. A. Reams, and J. A. Scrivani. 2000. Optical Remote Sensing for Forest Area Estimation. *Journal of Forestry,* Vol. 98, No. 5, pp. 31–36.

14 Hyperspectral Remote Sensing



MAJOR TOPICS TO UNDERSTAND

- Spectroscopy
- Hyperspectral Remote Sensing
- The Airborne Visible/Infrared Imaging Spectrometer
- The Image Cube
- Spectral Libraries
- Overview of Typical Abundance Mapping Processing Steps
- Spectral Mixing Analysis
- Spectral Angle Mapping
- Analyses
- Wavelet Analysis for Hyperspectral Imagery

14.1 INTRODUCTION

Remote sensing involves examination of features observed in several regions of the electromagnetic spectrum. Conventional remote sensing, as outlined in previous chapters,

is based on use of several rather broadly defined spectral regions. *Hyperspectral remote sensing* is based on examination of many narrowly defined spectral channels. Sensor systems such as the Multi-angle Imaging SpectroRadiometer (MISR) and Sentinel-2 MultiSpectral Instrument (MSI) provide 4 and 12 spectral channels, respectively. The hyperspectral sensors described below can provide 200 or more channels, each only 10 nm wide. In the context of the discussion of image resolution, hyperspectral sensors implement the concept of "spectral resolution" to the extreme. Although hyperspectral remote sensing applies the same principles and methods discussed previously, it requires such specialized data sets, instruments, field data, and software that it forms a specialized field of inquiry.

14.2 SPECTROSCOPY

Hyperspectral data have detail and accuracy that permit investigation of phenomena and concepts that greatly extend the scope of traditional remote sensing. For example, analysts can begin to match observed spectra to those recorded in spectral libraries and closely examine relationships among brightnesses in several spectral channels to estimate atmospheric effects using data within the image itself. Such capabilities present opportunities for much more precise identification of material types and conditions than is possible with broadband sensors, for investigation of phenomena such as the blue and red shifts, and, as noted, for correction of data in some bands using other bands that convey information about atmospheric transmission.

These capabilities extend the reach of remote sensing into the field of *spectroscopy,* the science devoted to the detailed examination of very accurate spectral data. Classical spectroscopy has its origins in experiments conducted by Isaac Newton (1642–1727), who used glass prisms to separate visible light into the spectrum of colors. Later, another English physicist, William Wollaston (1766–1828), noted that spectra displayed dark lines when light is projected through a narrow slit. The meaning of these lines was discovered through the work of Joseph Fraunhofer (1787–1826), a German glassmaker who discovered distinctive lines in spectra of light from the Sun and from stars. Dark lines (*absorption spectra*) are observed as radiation passes through gases at low pressure; bright lines (*emission spectra*) form as heated gases (e.g., in the Sun's atmosphere) emit radiation. These lines have origins in the chemical elements present in the gases, a discovery that has permitted astronomers to investigate differences in the chemical compositions of stars and planets. The Danish physicist Niels Bohr (1885–1962) found that the character of Fraunhofer lines is ultimately determined by the atomic structure of gases.

Instruments used in spectroscopy—*spectroscopes, spectrometers, spectrographs*—are designed to collect radiation with a lens and divide it into spectral regions (using a prism or diffraction grating) that are then measured electronically. This form of spectroscopy is now a standard method not only in astronomy, but also for laboratory analyses to characterize unidentified materials.

14.3 HYPERSPECTRAL REMOTE SENSING

Hyperspectral remote sensing uses the practice of spectroscopy to examine images of the Earth's surface. Although hyperspectral remote sensing sometimes applies the techniques

of classical spectroscopy to the study of atmospheric gases and pollutants, for example, more commonly it applies these techniques to the making of precise, accurate, detailed spectral measurements of the Earth's surface (*imaging spectroscopy*).[1] Such data have accuracy and detail sufficient to begin to match observed spectra to those stored in databases known as *spectral libraries*. Instruments for hyperspectral remote sensing differ from those of conventional spectroscopy in that they gather spectra not only for point targets, but for areas—not for stars or laboratory samples, but for regions of the Earth's surface. Instruments for hyperspectral remote sensing differ from other remote sensing instruments in terms of their extraordinarily fine spectral, spatial, and radiometric resolutions and their careful calibration. Some hyperspectral instruments collect data in 200 or more channels at 12 or more bits. Because of their calibration and ability to collect data having fine detail, such instruments greatly extend the reach of remote sensing, not only by extending the range of applications but also by defining new concepts and analytical techniques.

Although the techniques of classical spectroscopy can be used in hyperspectral remote sensing to examine, for example, atmospheric gases, hyperspectral remote sensing typically examines very detailed spectra for images of the Earth's surface, applies corrections for atmospheric effects, and matches them to spectra of known features.

14.4 THE AIRBORNE VISIBLE/INFRARED IMAGING SPECTROMETER

One of the first airborne hyperspectral sensors was designed in the early 1980s by the NASA Jet Propulsion Laboratory (JPL, Pasadena, California). The *airborne imaging spectrometer* (AIS) greatly extended the scope of remote sensing by virtue of its number of spectral bands; their fine spatial, spectral, and radiometric detail; and the accuracy of its calibration. AIS collected 128 spectral channels, each about 10 nm wide, in the interval 1.2 to 2.4 μm. The term *hyperspectral remote sensing* recognizes the fundamental difference between these data and those of the usual broadband remote sensing instruments. (Sensor systems with even finer spectral resolution, designed primarily to study atmospheric gases, are known as *ultraspectral* sensors.)

Although several hyperspectral instruments are now in operational use, an important pioneer in the field of hyperspectral remote sensing is the *airborne visible/infrared imaging spectrometer* (AVIRIS; *http://aviris.jpl.nasa.gov*). AVIRIS was developed by NASA and JPL from the foundations established by AIS. AVIRIS was first tested in 1987, was placed in service in 1989, and has since been modified at intervals to upgrade its reliability and performance. It has now acquired thousands of images.

Hyperspectral sensors necessarily employ designs that differ from those of the usual sensor systems. An objective lens collects radiation reflected or emitted from the scene. A collimating lens then projects the radiation as a beam of parallel rays through a diffraction grating that separates the radiation into discrete spectral bands (**Figure 14.1**).

[1]Other planetary bodies, can, of course be studied using similar instrumentation. For example, the Compact Reconnaissance Imaging Spectrometer for Mars (CRISM) is a hyperspectral imaging system aboard the Mars Reconnaissance Orbiter with 6.55 nm/channel spacing. Data quality, as of this writing, has been adversely impacted by the failure of two of the three cryocoolers.

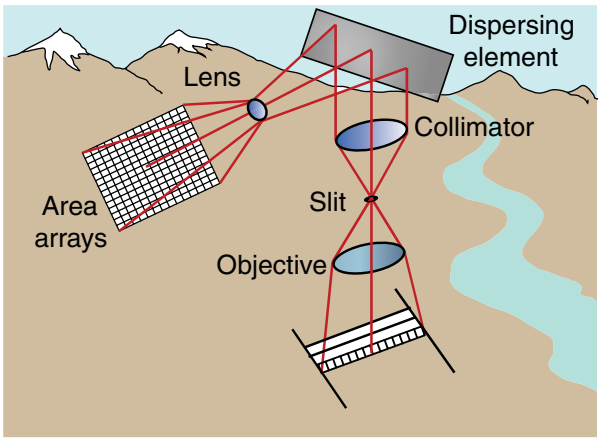


FIGURE 14.1 Imaging spectrometer. Based on a NASA diagram. Image by Susmita Sen.

Energy in each spectral band is detected by linear arrays of silicon (visible), indium gallium arsenide (NIR), and indium antimonide (SWIR). Because of the wide spectral range of AVIRIS, detectors are configured in four separate panels (0.4–0.7 μm, 0.7–1.3 μm, 1.3–1.9 μm, and 1.8–2.5 μm), each calibrated independently (Figure 14.2). AVIRIS operates over the spectral range of 400–2500 nm (0.4–2.45 μm), producing 224 spectral channels, each 10 nm wide (Figure 14.3). At its usual operating altitude, each image records a strip 11 km wide, processed to form scenes recording areas about 11 km × 512 lines (or about 10 km) each. Each line of data conveys about 677 pixels. If operated at low altitude, pixels might each represent ground areas about 4 m on a side; if operated at higher altitudes, as is common practice, each pixel represents ground areas as large as about 20 m on a side.

14.5 THE IMAGE CUBE

The *image cube* refers to the representation of hyperspectral data as a three-dimensional figure, with two dimensions formed by the x and y axes of the usual map or image display

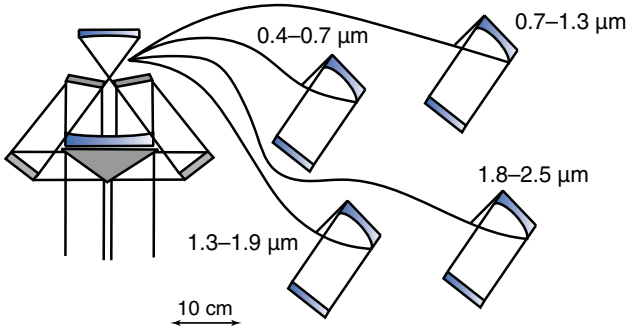


FIGURE 14.2 AVIRIS sensor. Based on a NASA diagram. Image by Susmita Sen.

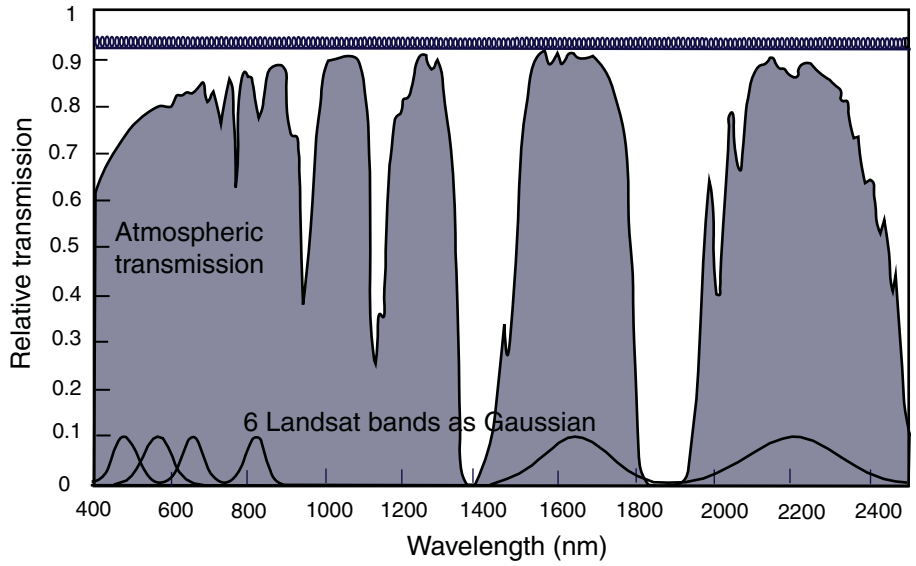


FIGURE 14.3 AVIRIS spectral channels compared to Landsat Thematic Mapper (TM) spectral channels. The 220 narrow bands at the top represent AVIRIS spectral channels. For comparison, Landsat TM channels are shown at the bottom. Based on Green and Simmonds (1993). Image by Susmita Sen.

and the third (z) formed by the accumulation of spectral data as additional bands are superimposed on each other. In Figure 14.4, the top of the cube is an image composed of data collected at the shortest wavelength (collected in the ultraviolet), and the bottom of the cube is an image composed of data collected at the longest wavelength (2.5 μm). Intermediate wavelengths are found as horizontal slices through the cube at intermediate positions. Values for a single pixel observed along the edge of the cube form a spectral trace describing the spectra of the surface represented by the pixel.

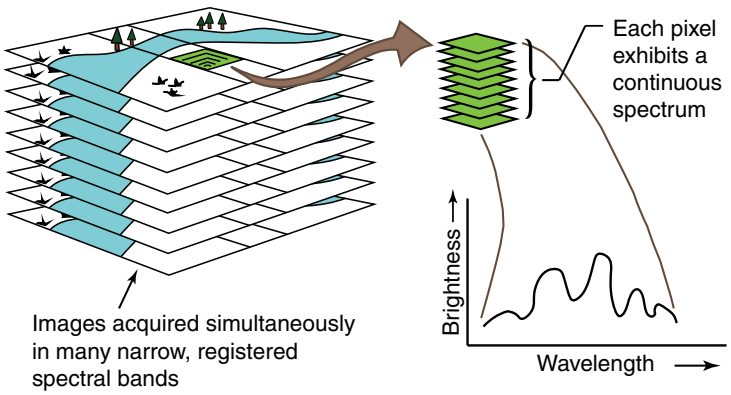


FIGURE 14.4 Image cube. Based on a NASA diagram. Image by Susmita Sen.

14.6 SPECTRAL LIBRARIES

The development of hyperspectral remote sensing has been accompanied by the accumulation of detailed spectral data acquired in the laboratory and in the field. These data are organized in *spectral libraries,* databases maintained primarily by governmental agencies but also by other organizations. These libraries assemble spectra that have been acquired at test sites representative of varied terrain and climate zones, observed in the field under natural conditions. Also included are other data describing, for example, construction materials, minerals, vegetation, and fabrics as observed in laboratories under standardized conditions.

Such data are publicly available to the remote sensing community and have been incorporated into software designed for use in hyperspectral remote sensing. Maintenance of a spectral library requires specialized effort to bring data into a common format that can be used by a diverse community of users. Spectral data are typically collected by diverse instruments under varied conditions of illumination. These many differences must be resolved to prepare data in a format that permits use by a diverse community.

Because of the fine spectral, spatial, and radiometric detail of hyperspectral analysis, identification and cataloging present special problems for design of spectral libraries. Therefore, each spectral record must be linked to detailed information specifying the instruments used, meteorological conditions, nature of the surface, and circumstances of the measurement. These kinds of ancillary data are required for successful interpretation and analysis of the image data.

14.7 OVERVIEW OF TYPICAL ABUNDANCE MAPPING PROCESSING STEPS

Figure 14.5, modified from Boardman and Kruse (2011), illustrates a sequence of analysis for hyperspectral data from conversion to surface reflectance to abundance mapping (additional details for many of these summarized steps are outlined below), assuming linear spectral mixing. Analysis begins with processing to apparent surface reflectance (e.g., level 2 from AVIRIS-Next Generation). Any of several methods discussed earlier can be used to correct for atmospheric effects if analysis-ready data are not available. Spectral data reduction reduces the number of bands to the inherent image dimensionality with respect to signal. Since the signal-to-noise ratio is often quite low in hyperspectral data, the maximum noise fraction (MNF; Green et al., 1988) is usually used at this step (a PCA variant that successively maximizes signal rather than variance, sometimes called erroneously but synonymously the minimum noise fraction). Using the retained MNF (Green et al., 1988) bands, pixels that protrude from the MNF feature space are selected as "pure" pixels and potential *endmembers* (spectrally pure constituent spectra; see Section 14.8). Two- and three-dimensional scatter plots are then used to locate the tips of the feature space protuberances that, as simplex vertices (see Section 14.8), represent the spectral endmembers, which are then identified using spectral libraries, *in situ* examination, and the like. Finally, endmember *abundance* (the *fraction* of each pixel composed of a given endmember) is mapped through one of several linear *spectral unmixing* methods.

Spectral unmixing is necessitated since most pixels are mixtures at the spatial resolution of typical hyperspectral sensors. The typical processing flow summarized here presumes that at least some pure pixels are present in the scene. That may not be the case,

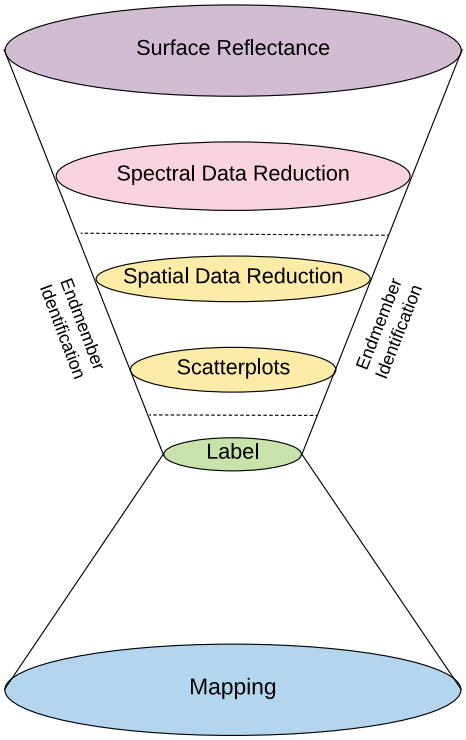


FIGURE 14.5 Spectral hourglass. From Boardman and Kruse (2011). Copyright © 2011 IEEE. Used by permission.

however, in which case endmembers have to be identified using an alternative strategy (see, for example, Plaza et al., 2011). In addition, even when pure pixels are present, the spectral hourglass processing flow is analyst-intensive and computationally expensive. There are more efficient mechanisms for finding simplex vertices (see, e.g., Nascimento and Bioucas-Dias, 2005) that combine the spatial reduction and scatter plot steps in Figure 14.5.

While more details are given in the following section, the overall linear mixing model (following the treatment and nomenclature of Settle and Drake, 1993) is straightforward mathematically. First, we have to assume that every endmember has been identified. Geometrically, as pointed out by Nascimento and Bioucas-Dias (2005), the endmembers are vertices of a *simplex* (the concept of a triangle extended to n-dimensions) containing all scene observations. Each endmember (simplex vertex) is a spectral signature, a column in endmember matrix $\mathbf{M}$. The rows in $\mathbf{M}$ are the spectral bands. The rows in pixel reflectance vector $\mathbf{x}$ are the spectral bands (like the rows in $\mathbf{M}$). Fraction vector $\mathbf{f}$ contains the proportion of each endmember present in each pixel, where the rows are the endmembers. Then,

$$\mathbf{x} = \mathbf{M}\mathbf{f} \qquad \text{(EQ. 14.1)}$$

(the error term excluded for clarity). In practice, of course, we know $\mathbf{x}$ and $\mathbf{M}$ and need to estimate $\mathbf{f}$. A least squares solution to this equation is feasible but requires (1) every fraction to be positive and (2) all fractions for a given pixel to add up to one. For additional

details on the linear algebra necessary for estimating **f** (including for less constrained solutions), see Settle and Drake (1993). An example of a mixture of just two endmembers is shown as Table 14.1.

14.8 SPECTRAL MIXING ANALYSIS

Fine spectral resolution, like fine spatial resolution, does not overcome the enduring obstacles to the practice of remote sensing. Surface materials recorded by the sensor are not always characteristic of subsurface conditions. Atmospheric effects, shadowing, and topographic variations contribute to observed spectra to confuse interpretations. Even when observed at fine spatial and spectral detail, surfaces are often composed of varied materials (Figure 14.6). Therefore, the sensor observes composite spectra that may not clearly match to the pure spectra of spectral libraries. *Linear mixing* refers to additive combinations (Figure 14.6b) of several diverse materials that occur in patterns

TABLE 14.1 Example of Mixture of Two Endmembers

	Water	Trees	Riparian
Blue	1014.0	941.0	962.9
Red	832.0	799.0	808.9
Green	538.0	579.0	566.7
NIR	338.0	1846.0	1393.6

Note: The riparian pixel is a linear mixture consisting (in this case) of 30% water and 70% trees. Endmember brightness values are from a Sentinel 2 Multispectral Imager acquisition over East Godavari district, Andhra Pradesh, India, acquired on December 22, 2017.

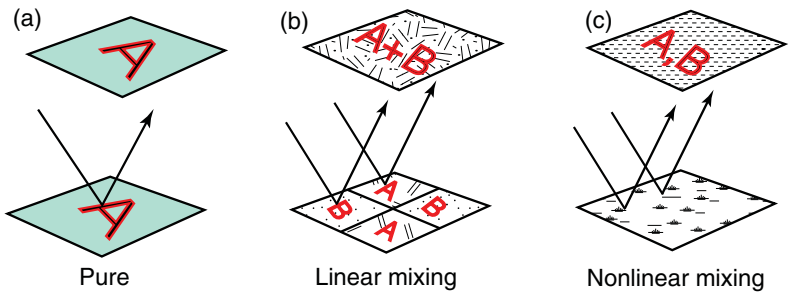


FIGURE 14.6 Linear and nonlinear spectral mixing. (a) If a pixel represents a uniform ground area at the resolution of the sensor, the pixel represents a pure spectrum. (b) If a pixel represents two or more surfaces that occur in patches that are large relative to the sensor's resolution, mixing occurs at the sensor. The pattern of the composite surfaces can never be resolved, but because mixing occurs in a linear manner, proportions of the components can be estimated. (c) If the composite occurs at a scale that is fine relative to the resolution of the sensor, mixing occurs before the radiation reaches the sensor, and components of the composite cannot be estimated using the methods described here. Image by Susmita Sen.

too fine to be resolved at the resolution of the sensors. This is the effect of mixed pixels. As long as the radiation from component patches remains separate until it reaches the sensor, it is possible to estimate proportions of component surfaces from the observed pixel brightness. Linear mixing might occur when components of a composite surface are found in a few compact areas. *Nonlinear mixing* occurs when radiation from several surfaces combines before it reaches the sensor. Nonlinear mixing occurs when component surfaces arise in highly dispersed patterns (Figure 14.6c). Nonlinear mixing cannot be described by the techniques addressed here, but they can be addressed using physics-based techniques, such as the isotropic multiple scattering approximation (Hapke, 2012) or statistical/machine learning techniques such as support vector machines (SVMs; see, for example, Brown, Gunn, and Lewis, 1999), though SVM can also be used to unmix linearly. The reader should refer to Hapke (2012) and Heylen, Parente, and Gader (2014) for further detail on nonlinear unmixing techniques.

Spectral mixing analysis (also known as *spectral unmixing*) is devoted to extracting pure spectra from the complex composites of spectra that by necessity form each image. It assumes that pixels are formed by linear mixing and further assumes that it is possible to identify the components contributing to the mixture. It permits analysts to define key components of a specific scene and forms an essential component in the process of spectral matching, discussed below. Analysts desire to match data from hyperspectral images to corresponding laboratory data in order to identify surfaces from their spectral data much more precisely than previously was possible. Whereas conventional image analysis (Chapter 12) matches pixels to broad classes of features, hyperspectral image matching attempts to make more precise identifications—to the specific mineral constituents of soils or rocks, for example.

Therefore, spectral matching requires the application of techniques that enable analysts to separate pure pixels from impure pixels. This problem is well matched to the capabilities of *convex geometry,* which examines multidimensional data envisioned in n dimensions. Individual points (pixels) within this data space can be examined as linear combinations of an unknown number of pure components. Convex geometry can solve such problems, provided that the components are linearly weighted, sum to unity, and are positive. We assume also that the data have greater dimensionality (more spectral bands) than the number of pure components.

The illustrations here, for convenience and legibility, show only two dimensions, although the power of the technique is evident only with much higher dimensionality. In Figure 14.7, the three points *A, B,* and *C* represent three spectral observations at the extreme limits of the swarm of data points, represented by the shaded pattern. That is, the shaded pattern represents the values of all the pixels within a specific image or subimage, which is simplified by the triangle shape. (Other shapes can be defined as appropriate to approximate the shape of data swarm, although an objective is to define the simplest shape, *simplex,* that can reasonably approximate the pattern of the data swarm.)

These three points (for the example in Figure 14.7a) form *endmembers,* defined as the pure pixels that contribute to the varied mixtures of pixels in the interior of the data swarm. Once the simple form is defined, the interior pixels can be defined as linear combinations of the pure endmembers. In general, interior points can be interpreted as positive unit-sum combinations of the pure variables represented by endmembers at the vertices. In general, a shape defined by $n + 1$ vertices is the simplest shape that encompasses interior points (i.e., for two dimensions [$n = 2$], the simplex is a triangle [$3 = n + 1$]). The faces of the shape are *facets,* and the exterior surface is a *convex* hull. In Figure

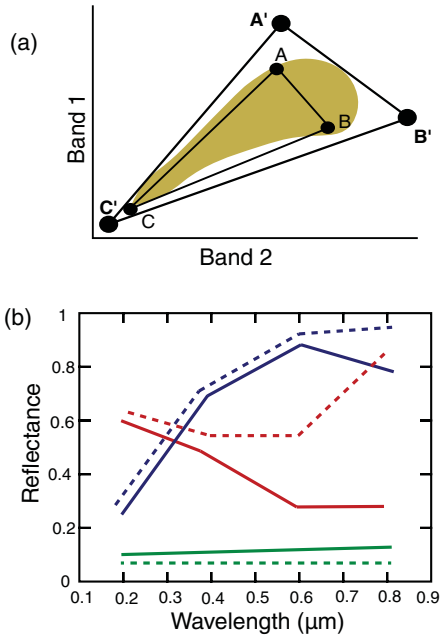


FIGURE 14.7 Spectral mixing analysis: (a) simplex; (b) endmembers. Based on Tompkins et al. (1993). Image by Susmita Sen.

14.7, *A*, *B*, and *C* are the observed approximations of the idealized spectra *A*, *B*, and *C* that may not be observed on any specific image.

In the application of convex geometry to hyperspectral data, it is first necessary to define the dimensionality of the data (see Chapter 11). The original hyperspectral data are converted to surface reflectances (from radiances) using atmospheric models. The data are condensed by applying a noise-segregated version of principal components analysis, the MNF (Green et al., 1988) as noted in Section 14.7. Although hyperspectral data may include many bands (224 for AVIRIS), duplication from one channel to another means that inherent dimensionality is much less (often somewhere between 10 and 20), depending on the specifics of each scene.

The analyst then examines the transformed data in a data space to define the smallest simplex that fits the data. This process defines the $n + 1$ facets that permit identification of the n endmembers (Figure 14.7a). These vertices, when projected back into the original spectral domain, estimate the spectra of the endmembers. These spectra are represented by Figure 14.7b. The objective is to match these endmembers to spectra from spectral libraries and then to prepare maps and images that reveal the varied mixtures of surfaces that contribute to the observed spectra in each image.

Typical endmembers in arid regions have included bare soil, water, partially vegetated surfaces, fully vegetated surfaces, and shadows. End members can be investigated in the field to confirm or revise identifications made by computer. Software for hyperspectral analysis often includes provisions for accessing spectral libraries (and for importing additional spectra as acquired in the field or laboratory), as well as for searching for matches with endmembers identified. Although it may not always be possible to uniquely identify matches in spectral libraries, such analyses can narrow the range of alternatives. In some instances, mathematical models can assist in defining poorly established endmembers.

14.9 SPECTRAL ANGLE MAPPING

Spectral angle mapping (SAM) is a classification approach that examines multispectral or hyperspectral data by evaluating the relationships between pixel values projected in feature space. Envision a pixel projected into feature space: its position can be described by a vector with an angle in relation to the measurement axes (Figure 14.8). Its position relative to another pixel (or perhaps a set of reference or training data) can be evaluated by assessing the difference between the angles of the two vectors. Small angles indicate a close similarity; large angles indicate lower similarity.

The effectiveness of this technique is achieved when applied to many more pixels and dimensions than can be represented in a single diagram. SAM is most useful when extended to tens or hundreds of dimensions (Kruse et al., 1993). SAMs differ from the usual classification approaches because they compare each pixel in the image with each spectral class and then assign a value between 0 (low resemblance) and 1 (high resemblance) to each pair. Although SAMs are valuable for hyperspectral analyses in many dimensions, their effectiveness is based on the implicit assumption that each pixel is being compared to pure spectra.

14.10 ANALYSES

Other investigations have attempted to understand relationships between spectral data and specific physical or biological processes. For example, Curran (1994) reviews efforts to use hyperspectral data to monitor botanical variables, such as the chlorophyll, lignin, cellulose, water, and nitrogen content of plant tissues. Others have examined, at high levels of spectral, radiometric, and spatial resolution, observed spectra of laboratory samples of plant tissues influenced by atomic and molecular structures of water and by specific concentrations of organic compounds (such as chlorophyll, lignin, and cellulose). These relationships have formed the basis for research devoted to the examination of hyperspectral data of vegetation canopies in order to derive estimates of foliar chemistry of plant tissues *in situ*. Such estimates support agricultural, forestry, and ecological studies by providing indications of nutrient availability, rates of productivity, and rates of decomposition. Green (1993) and Rivard and Arvidson (1992) report some of the applications of hyperspectral data to lithological and mineralogic analyses.

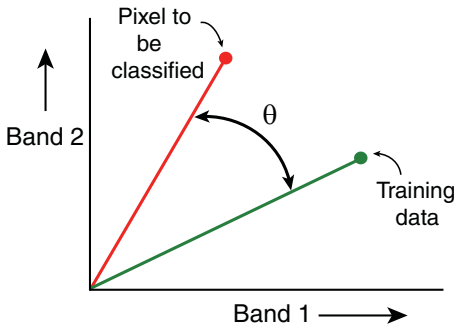


FIGURE 14.8 Spectral angle mapping. Image by Susmita Sen.

14.11 WAVELET ANALYSIS FOR HYPERSPECTRAL IMAGERY

A *wavelet* can be defined as a wave-like oscillation of fixed duration, characterized by an amplitude that has a value of zero, then increases, then declines again to zero—analogous perhaps to oscillations such as those that we encounter on seismographs or heart monitors. (See Figure 14.9 for an example of a Haar wavelet.) *Wavelet analysis* decomposes images, sounds, or spectra into wavelets of varying duration. Given the focus of this chapter, we will discuss wavelet analysis of spectra, but the basic techniques are the same for images and sounds. Wavelet analysis is inherently a *multiscale* (or *multiresolution*) analysis because the spectrum is decomposed into subspectra of ever-finer scales. In practice, we first determine the finest scale wavelets (from the first *detail*) present in a spectrum and then remove them, resulting in a less detailed *approximation* of the spectrum. The next finest-scale wavelets (second detail) are then removed, and so on, until we end up with a spectrum in which only the broadest absorption and reflection features are present. The details are illustrated by Figure 14.10, where on the right *d*1 (detail 1) through *d*6 (detail 6) represent ever-coarser scales. The resulting approximations of the spectrum, once each level of detail has been removed, are shown on the left. To use a musical analogy, for which we are indebted to Walker (2008), think of the process as successively stripping off "ornaments" to an essential melody. A thorough discussion of wavelets is well beyond the reach of our discussion here, so readers are referred to sources such as Walker (2008) for a proper introduction to this topic.

14.12 SUMMARY

The vast amounts of data collected by hyperspectral systems, and the problems they present for both collection and analysis, prevent routine use of hyperspectral data in the same way that we might collect Landsat or Sentinel-2 data on a regular basis. More likely in the near term, hyperspectral data will provide a means for discovering and refining the knowledge needed to develop improved sensors and analytical techniques that can be applied on a more routine basis. A second important role for hyperspectral data lies in the monitoring of long-term research sites, especially those devoted to study of biophysical processes and other phenomena that change over time (e.g., the National Ecological Observatory Network; Johnson et al., 2010). The fine detail of hyperspectral data will provide enhanced data for ecological monitoring and for understanding patterns recorded on lower resolution data of the same sites.

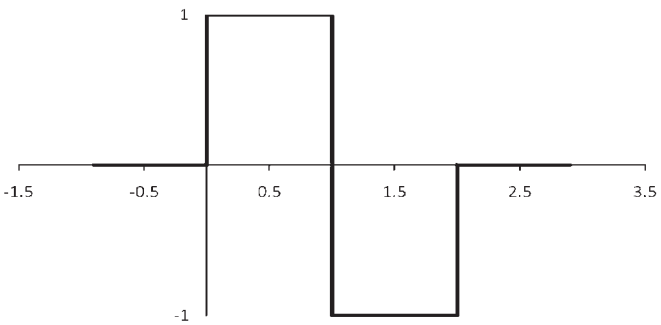


FIGURE 14.9 Illustration of a Haar wavelet.

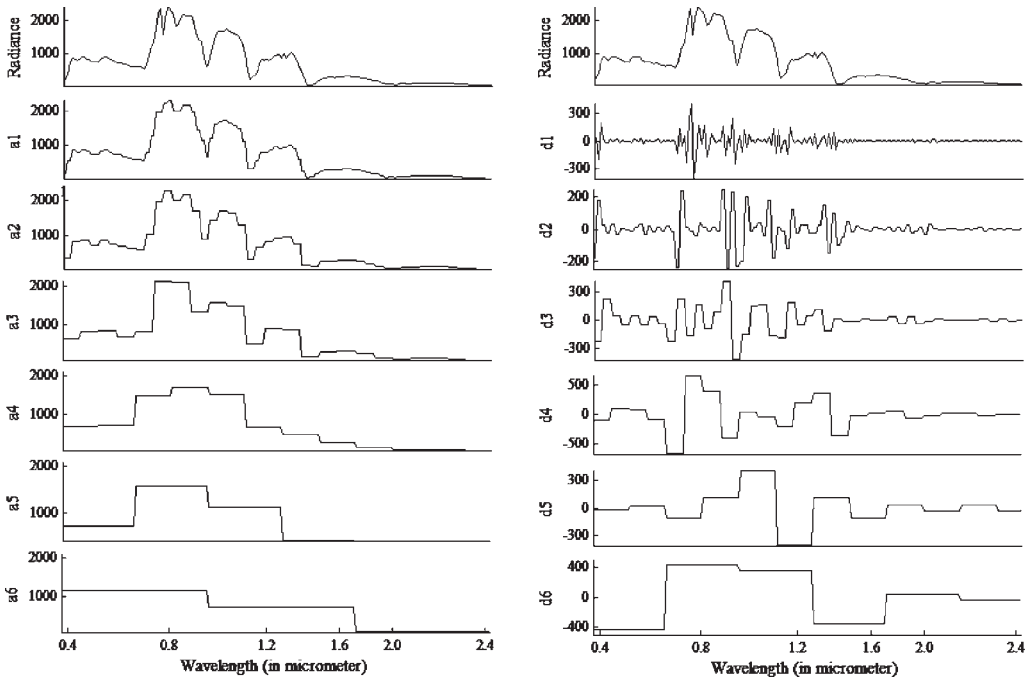


FIGURE 14.10 Wavelet analysis. Original hyperspectral radiance (in $W\,m^{-2}\,sr^{-1}$) curve of a randomly chosen shortleaf pine spectrum (acquired using the AVIRIS sensor) and its approximation (a_1, a_2, . . . a_6) and details (d_1, d_2, . . . d_6), up to six levels of decomposition. From Banskota et al. (2011, p. 3554). Copyright © 2011 Taylor and Francis Group, LLC.

Data from satellite-based hyperspectral sensors are available through, for example, the U.S. Geological Survey Earth Explorer (*https://earthexplorer.usgs.gov*). NASA's EO-1 satellite, launched in November 2000 and decommissioned in March 2017, was designed as a vehicle to test the feasibility of advanced imaging systems. One of EO-1's sensors was Hyperion, a high-resolution hyperspectral sensor capable of resolving 220 spectral bands from 0.4 to 2.5 m at a 30-m resolution (Figure 14.11). The instrument could image a 7.5 km × 100 km land area to provide detailed spectral mapping across all 220 channels with high radiometric accuracy. EO-1 was in an orbit permitting both Landsat 7 and EO-1 to image the same ground area at least once a day, permitting both systems to collect images under identical viewing conditions. The paired images were used for evaluation of the imaging technology.

Scientists continue to make compelling cases for specific, narrow wavelength bands (such as those needed for better articulation of the red edge or detection of crop residue). As the number of these new narrow bands increases, an engineering case can be made to convert the sensors on moderate resolution Earth resource satellite imaging from multispectral (Landsat, Sentinel-2) to hyperspectral. At the time of this writing, engineering challenges constraining signal-to-noise ratio at needed spatial resolutions were being addressed. As such, the future of moderate resolution Earth resource remote sensing is likely to include an increased use of imaging spectroscopy.

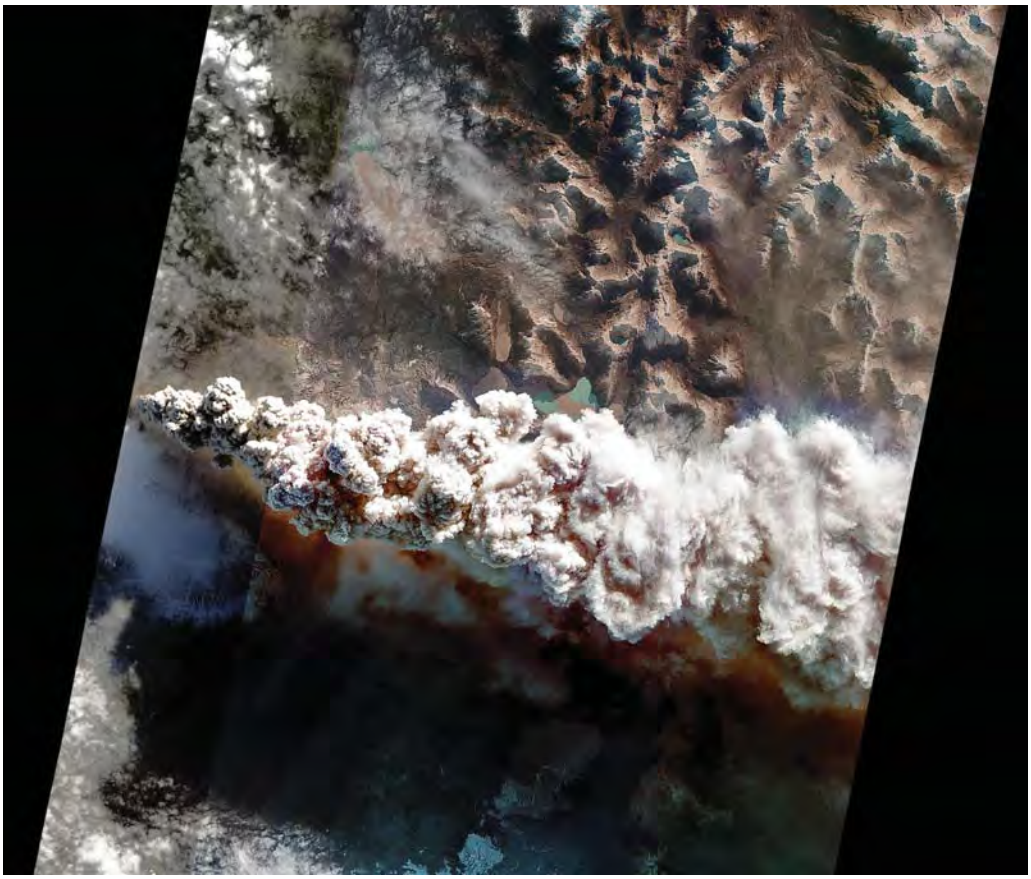


FIGURE 14.11 Hyperion images of the Puyehue-Cordón Caulle volcano eruption, Chile, obtained on June 14, 2011. In the main image, created from data at visible wavelengths, the ash-rich volcanic plume billows out of the vent, punching through a low cloud layer. The plume's gray color is a reflection of its ash content. Fine particles of ash are carried high into the atmosphere and get dispersed by the prevailing winds. The scene depicted here is about 19 mi (30 km) wide and has a resolution of 98 ft (30 m) per pixel. The vertical direction is north. The small Hyperion images are each 4.8 mi (7.7 km) across. From NASA.

REVIEW QUESTIONS

1. Discuss the *advantages* of hyperspectral remote sensing in relation to more conventional remote sensing instruments.
2. List some *disadvantages* of hyperspectral remote sensing relative to use of systems such as Sentinel-2 or Landsat.
3. Prepare a plan to monitor an agricultural landscape in northern France using *both* hyperspectral data *and* Sentinel-2 or Landsat data.
4. How would use of hyperspectral data influence collection of field data, compared to similar studies using Sentinel-2 or Landsat imagery?

5. Discuss how *preprocessing* and *image classification* differ with hyperspectral data, compared to more conventional multispectral data.
6. How would *equipment* needs differ for image processing of hyperspectral data, compared to more conventional multispectral data?
7. The question of choosing between broad-scale coverage at coarse detail and focused coverage at fine detail recurs frequently in many fields of study. How does the availability of hyperspectral data influence this discussion?
8. Can you think of ways that the availability of hyperspectral data will influence the concepts and theories of remote sensing?
9. Hyperspectral data have so much volume that for typical users it is not feasible to accumulate geographic coverage comparable to that of Sentinel-2 or Landsat, for example. What value, then, can hyperspectral data have?
10. Discuss the problems that arise in attempting to design and maintain a spectral library. Consider, for example, the multitude of different materials and surfaces that must be considered, each under conditions of varied illumination.

REFERENCES

Banskota, A., R. H. Wynne, and N. Kayastha. 2011. Improving Within-Genus Tree Species Discrimination Using the Discrete Wavelet Transform Applied to Airborne Hyperspectral Data. *International Journal of Remote Sensing,* Vol. 32, pp. 3551–3563.

Boardman, J. W., and F. A. Kruse. 2011. Analysis of Imaging Spectrometer Data Using N-Dimensional Geometry and a Mixture-Tuned Matched Filtering Approach. *IEEE Transactions on Geoscience and Remote Sensing,* Vol. 49, pp. 4138–4152.

Brown, M., S. R. Gunn, and H. G. Lewis. 1999. Support Vector Machines for Optimal Classification and Spectral Unmixing. *Ecological Modelling,* Vol. 120, pp. 167–179.

Curran, P. J. 1994. Imaging Spectrometry. *Progress in Physical Geography,* Vol. 18, pp. 247–266.

Green, A. A., M. Berman, P. Switzer, and M. D. Craig. 1988. A Transformation for Ordering Multispectral Data in Terms of Image Quality with Implications for Noise Removal. *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 26, pp. 65–74.

Green, R. O. (ed.). 1993. *Summaries of the Fourth Annual JPL Airborne Geoscience Workshop, October 25–29, 1993*. Pasadena, CA: NASA and the Jet Propulsion Laboratory, 209 pp.

Green, R. O., and J. J. Simmonds. 1993. A Role for AVIRIS in Landsat and Advanced Land Remote Sensing System Program. In *Summaries of the Fourth Annual JPL Airborne Geoscience Workshop, October 25–29, 1993* (R. O. Green, ed.). Pasadena, CA: NASA and the Jet Propulsion Laboratory, pp. 85–88.

Hapke, B. 2012. *Theory of Reflectance and Emittance Spectroscopy* (2nd ed.). Cambridge: Cambridge University Press, 528 pp.

Heylen, R., M. Parente, and P. Gader. 2014. A Review of Nonlinear Hyperspectral Unmixing Methods. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing.* Vol. 7, pp. 1844–1868.

Johnson, B. R., M. A. Kuester, T. U. Kampe, and M. Keller. 2010. National Ecological Observatory Network (NEON) Airborne Remote Measurements of Vegetation Canopy Biochemistry and Structure. *2010 IEEE International Geoscience and Remote Sensing Symposium,* pp. 2079–2082.

Kruse, F. A., A. B. Lefkoff, J. B. Boardman, K. B. Heidebrecht, A. T. Shapiro, P. J. Barloon, and A. F. H. Goetz. 1993. The Spectral Image Processing System (SIPS)—Interactive Visualiza-

tion and Analysis of Imaging Spectrometer Data. *Remote Sensing of Environment,* Vol. 44, pp. 145–163.

Nascimento, J. M. P., and Bioucas-Dias, J. M. 2005. Vertex Component Analysis: A Fast Algorithm to Unmix Hyperspectral Data. *IEEE Transactions on Geoscience and Remote Sensing,* Vol. 43, pp. 898–910.

Plaza, A., G. Martín, J. Plaza, M. Zortea, and S. Sánchez. 2011. Recent Developments in Endmember Extraction and Spectral Unmixing. In S. Prasad, L. Bruce, and J. Chanussot (eds.), *Optical Remote Sensing. Augmented Vision and Reality,* Vol 3. Berlin: Springer.

Plaza, J., E. M. T. Hendrix, I. Garcia, G. Martin, and A. Plaza. 2012. Endmember Identification in Hyperspectral Images Without Pure Pixels: A Comparison of Algorithms. *Journal of Mathematical Imaging and Vision,* Vol. 42, pp. 163–175.

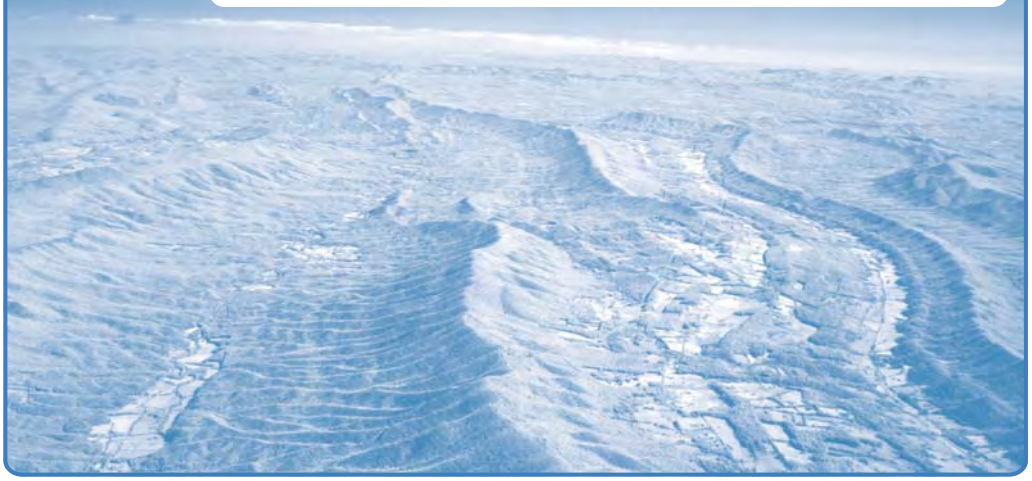
Rivard, B., and D. E. Arvidson. 1992. Utility of Imaging Spectrometry for Lithologic Mapping in Greenland. *Photogrammetric Engineering and Remote Sensing,* Vol. 58, pp. 945–949.

Settle, J. J., and N. A. Drake. 1993. Linear Mixing and the Estimation of Ground Cover Proportions. *International Journal of Remote Sensing,* Vol. 14, pp. 1159–1177.

Tompkins, S., J. F. Mustard, C. M. Peters, and D. W. Forsyth. 1993. Objective Determination of Image End-Members in Spectral Mixture Analysis of AVIRIS Data. In *Summaries of the Fourth Annual JPL Airborne Geoscience Workshop, October 25–29, 1993* (R. O. Green, ed.). Pasadena, CA: NASA Jet Propulsion Laboratory, pp. 177–180.

Walker, J. S. 2008. *A Primer on Wavelets and Their Scientific Applications* (2nd ed.). Boca Raton, FL: Chapman and Hall/CRC, 155 pp.

15 Change Detection



MAJOR TOPICS TO UNDERSTAND

- Bitemporal Spectral Change Detection Techniques
- Multitemporal Spectral Change Detection

15.1 INTRODUCTION

Repeated imaging enables assessment of changes in the type or condition of surface features. This is one of the most important of all analyses in remote sensing, typically called *change detection*. Many of these analyses use images acquired at two points in time, known as *bitemporal change detection,* the primary focus of this chapter. Increasingly, however, more widely available image time series have afforded the possibility of *multitemporal change detection,* addressed in Section 15.3.

Comparing images after classifying each is called *postclassification change detection,* which consists only of comparing, for a given location, the assigned class in one map to the assigned class in another. Even though this operation is a simple raster GIS analysis, it is often not the first choice of most analysts. This is because the resulting accuracy is typically low, as it incorporates any errors present in the original classifications,

approximated as the product of the overall accuracies of the individual classifications (Coppin et al., 2004).

In part because of the inherently low accuracy of postclassification change detection, analysis of spectral changes between two (or more) dates, called *spectral change detection*, is often preferred for analysis of traditional optical imagery (rather than lidar or radar data). Whether the unit of analysis is a pixel, a neighborhood, a multitemporal segment, or even (rarely) a spectral class, it is very important to follow the image preprocessing steps that minimize signal from variation that could be confused with the change detection signal of interest. Ideally, this means that the images being compared are:

1. Acquired from the same or well-intercalibrated sensors and acquired at the same time of day using the same instantaneous field of view and look angle.
2. For interannual analyses, acquired during the same season to minimize differences due to phenological changes.
3. Well co-registered, preferably to within two-tenths of a pixel or less (Dai and Khorram, 1999).
4. Free of clouds in the area of analysis.
5. Corrected to top-of-atmosphere or (preferably) surface (bottom-of-atmosphere) reflectance (see Chapter 11).
6. Free of other conditions not deemed part of the signal of interest. Examples of this last point could include the presence of soil moisture differences when assessing changes in forest wetland canopy cover or differences in harvest dates when assessing changes in vegetative cover using anniversary date images acquired in autumn.

It is often difficult in practice to meet all of these conditions, but maximizing signal from the change of interest requires minimizing signal from anything that could be confused with it.

15.2 BITEMPORAL SPECTRAL CHANGE DETECTION TECHNIQUES

Spectral change detection techniques can be classified (not mutually exclusively) as follows: (1) visual interpretation, (2) image algebra, (3) transformation/data reduction, (4) classification, and (5) statistical. The categories and techniques described here are common in both remote sensing research and applications, but they are by no means exhaustive. The reader should refer to edited compilations such as Lunetta and Elvidge (2000) or to reviews such as Coppin et al. (2004) and Lu et al. (2004) for additional information.

Visual Interpretation

Comparison of images from two dates is always the first place to start and is easily accomplished using one of three basic techniques. In the first, known as a multidate color image composite, two suitable radiometrically calibrated and co-registered images are first *composited* (in which all bands from both dates are stacked to form one image), and then bands from different dates are chosen for display to highlight the change of interest. In the second technique, the images are not composited but are displayed with one

on top of the other. Tools commonly available in most image processing packages can then be used to visualize change, including swiping one image over the other and flickering between images. The third technique is simply to view co-registered and geolinked images side by side, using the typical pan, zoom, and cursor inquiry tools for targeted analysis.

One primary goal of visual interpretation is often to help select the best change detection techniques based on the goals of the analysis. In some instances, however, it might be the mechanism by which the analysis is conducted. If the latter, heads-up digitizing is often used to both identify and classify change between the images. In some instances, a sample of points is interpreted. In either instance, however, the analyst is able to use elements of image interpretation besides hue (e.g., texture, shadows, pattern, shape, size, and association; Chapter 6), combined with knowledge of the area of application, with the potential for very accurate results (albeit through a time-consuming and subjective process).

Image Algebra

In image algebra, arithmetic operations are applied to corresponding pixels in each image, with the change image formed from the resulting values. While there are many possible algebraic operations, the most common are image differencing, image ratios, Euclidean distance, and change vector analysis.

Image differencing, particularly using vegetation indices, is one of the most widely used (and most effective) spectral change detection techniques. It consists of simply subtracting one or more bands from the same band(s) of the same area acquired at a different date. Values at or near zero identify pixels that have similar spectral values and therefore presumably have experienced no change between the two dates. The analyst must specify some threshold value, often heuristically, beyond which change is considered to have occurred. The results are very sensitive to the threshold chosen, and, as such, it is often desirable to use an objective process for its selection. A "salt-and-pepper" appearance also often results because of small spectral changes caused by other factors previously discussed, including slight misregistration (see **Figure 15.1**).

Taking the ratio of two bands, one from one image date, the other from the same band in the second image date, is another popular change detection technique. A simple numerical example using two Landsat images can illustrate the two approaches. On date one, we have a pixel in class *forest* with digital numbers of 100 and 24 in the near infrared and red, respectively. On date two, the digital numbers for the same pixel (now in class *urban or bare soil*) are 61 and 36 in the near infrared (NIR) and red, respectively. Let's take a look at the image difference and image ratio for the near-infrared band. The image difference would be 100 – 61 = 39. The image ratio would be 100 ÷ 61 = 1.64.

Now, what happens if we want to assess the magnitude of change using more than one band? One common means of doing this is the Euclidean distance introduced in Chapter 12, Equation 12.1. Stated in words, this calculation is nothing more than the square root of the sum of the squares of the distances. Returning to the previous example, we see that the difference in the red band for this pixel is 24– 36 = –12. The square root of the sum of the squares of the distances between these two brightness value vectors is then just

$$D = \sqrt{(39)^2 + (-12)^2} = \sqrt{1665} = 40.8 \qquad \text{(EQ. 15.1)}$$

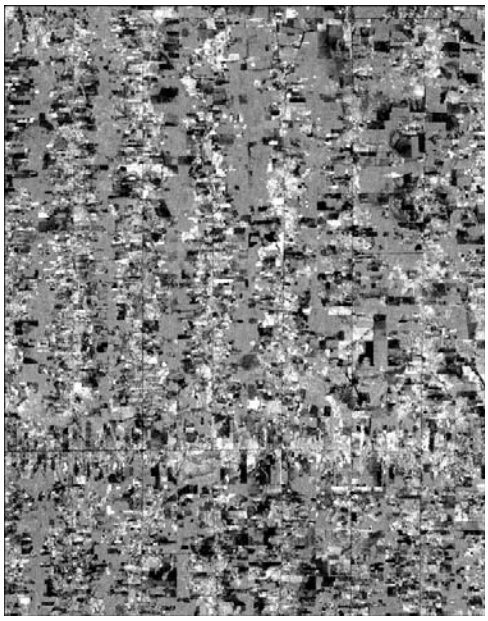


FIGURE 15.1 Normalized difference vegetation index (NDVI) image in which the 1992 Landsat Thematic Mapper (TM)–derived NDVI was subtracted from the 2002 Landsat TM–derived NDVI in Rondônia, Brazil. Areas that are brighter showed an increase in NDVI, and those that are darker showed a decrease.

Stated in the more easily extendable matrix form, the Euclidean distance is

$$D = \sqrt{(\boldsymbol{x}_1 - \boldsymbol{x}_2)'(\boldsymbol{x}_1 - \boldsymbol{x}_2)} \qquad \text{(EQ. 15.2)}$$

where $\boldsymbol{x}$ is the brightness value or reflectance vector for the given pixel, and the prime represents a matrix transposition. Working this through for our example gives us

$$D = \sqrt{(\boldsymbol{x}_1 - \boldsymbol{x}_2)'(\boldsymbol{x}_1 - \boldsymbol{x}_2)} = \sqrt{[39 - 12]\begin{bmatrix} 39 \\ -12 \end{bmatrix}} = \sqrt{39^2 + (-12)^2} = \sqrt{1665} = 40.8 \qquad \text{(EQ. 15.3)}$$

The Euclidean distance, however, gives only the *magnitude* of the multiband change. What if we also want to know the direction of the change? In this instance, *change vector analysis* (of which Euclidean distance can be considered a subset) is needed. This operation is best examined graphically. Consider Figure 15.2, in which the two pixels we have

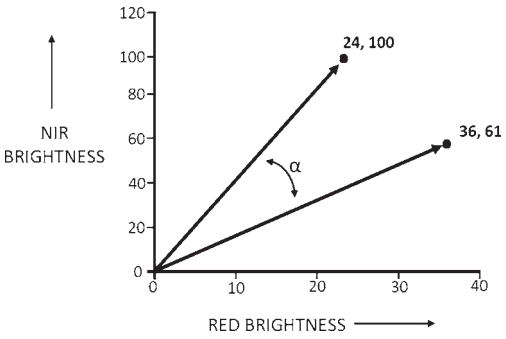


FIGURE 15.2 Change vectors. Two pixels are shown as vectors. The near-infrared digital number forms the *y* axis, and the red digital number is the *x* axis. The angle between them is α.

been examining are shown as vectors, with the near-infrared digital number as the *y* axis and the red digital number as the *x* axis. The angle between them is α, defined as

$$\alpha = \cos^{-1}\left(\frac{x_1 \cdot x_2}{\|x_1\| + \|x_2\|}\right) \quad \text{(EQ. 15.4)}$$

where

$$x_1 \cdot x_2 = \sum_{i=1}^{n} x_{1,n} x_{2,n} = x_{1,1}x_{2,1} + x_{1,2}x_{2,2} + \ldots + x_{1,n}x_{2,n} \quad \text{(EQ. 15.5)}$$

which is the *dot product* of the two vectors. For our two vectors, the dot product would be as follows:

$$x_1 \cdot x_2 = \begin{bmatrix} 100 \\ 24 \end{bmatrix}\begin{bmatrix} 61 \\ 36 \end{bmatrix} = (100)(61) + (24)(36) = 6100 + 864 = 6,964 \quad \text{(EQ. 15.6)}$$

The other term in Equation 15.4 that needs explanation is the vector *length* (also called the *norm*), defined as the sum of the squares of the vector elements, that is,

$$\|x\| = x_1^2 + x_2^2 + \ldots + x_n^2 \quad \text{(EQ. 15.7)}$$

For our two vectors, the lengths are then as follows:

$$\|x_1\| = \left\|\begin{matrix} 100 \\ 24 \end{matrix}\right\| = 100^2 + 24^2 = 10,000 + 576 = 10,576 \quad \text{(EQ. 15.8)}$$

$$\|x_2\| = \left\|\begin{matrix} 61 \\ 36 \end{matrix}\right\| = 61^2 + 36^2 = 3,721 + 1,296 = 5,017 \quad \text{(EQ. 15.9)}$$

Note that the vector length in two dimensions is just the hypotenuse calculated using the familiar Pythagorean theorem.

Substituting the dot product and vector lengths into Equation 15.4, we are able to compute the angle between the two vectors as follows:

$$\alpha = \cos^{-1}\left(\frac{x_1 \cdot x_2}{\|x_1\| + \|x_2\|}\right) = \cos^{-1}\left(\frac{6,964}{10,576 + 5,017}\right) = \cos^{-1}(0.45) = 63.5° \quad \text{(EQ. 15.10)}$$

Let's look at this now from the perspective of the land-use change we are trying to detect. The Euclidean distance between the *forest* and *urban or bare soil* is 40.8, giving the magnitude of the change vector. The angle between the two vectors is 63.5°. See **Figure 15.3** for an example using six bands from each of two Landsat TM images 11 years apart. Note that, for this example, only two bands and two pixels were used for simplicity, but there is nothing to preclude use of any number of spectral or derived bands or alternative units of analysis (e.g., image segments). Further analysis of the change between the vectors can proceed in many ways, such as thresholding, which is discussed earlier in the context of image differencing, or classification, which is discussed in the section "Classification," below.

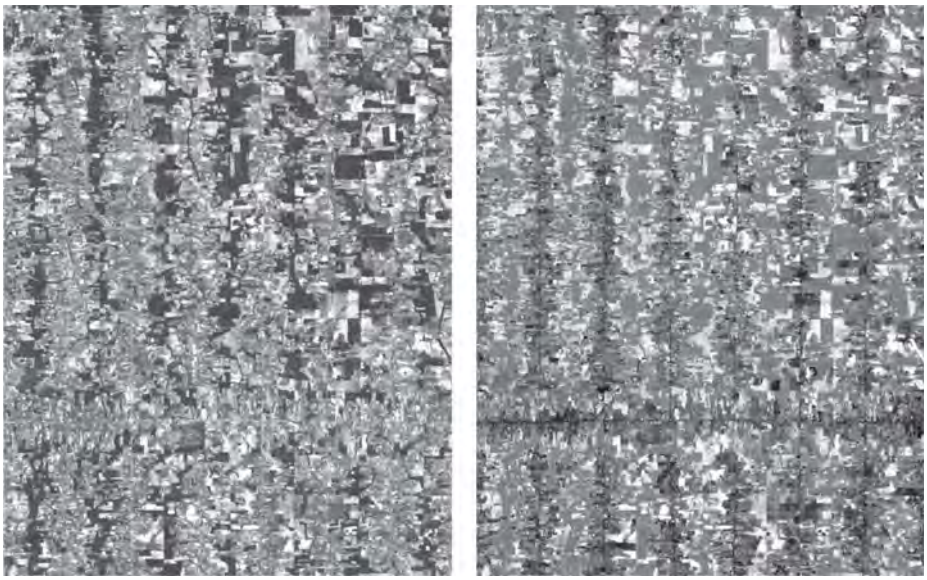


FIGURE 15.3 Change vector analysis. Example of change vector analysis output in which the distance (using all six nonthermal bands) between a 1992 Landsat Thematic Mapper (TM) image and a 2002 Landsat TM image acquired over Rondônia, Brazil, is shown at left. The right-hand image is the spectral angle image of the same area, in which the gray tones indicate the value of the spectral angle, with brighter tones indicating higher spectral angles and, therefore, higher spectral differences between the two dates. In this instance, at least, the spectral angle map seems to provide a more nuanced representation of the changes, with fewer areas represented as the dark gray or black tones indicating that no changes had occurred.

Transformation/Data Reduction

There are a variety of techniques by which the data in the original image can be transformed to new axes composed of linear combinations of the existing bands. The most widely used of these techniques are principal components analysis (PCA) and the tasseled cap (TC) transformation, both discussed in Chapter 11. Other possibilities, more widely utilized in hyperspectral data analysis, include the singular value decomposition and maximum (aka minimum) noise fraction.

Recall that PCA reorients the axes of multidimensional data space such that there is no longer any remaining covariance among the principal components (PC) bands, which are linear combinations of the original bands. This operation results in a de facto ability to provide the same information content using fewer bands. The result has value for change detection in two very particular ways. First, a PCA can be implemented on the composite image. If the change the analyst wishes to detect represents a large portion of the variance between the two dates, then one or more of the resulting PC bands may be oriented such that it highlights that change. When this effect occurs, it is very powerful. Unfortunately, there is no guarantee that PCA will produce this effect in every application. There are alternative transformations (such as canonical discriminant analysis) that can reorient the axes to maximize discrimination among classes, but these are beyond the scope of this discussion.

The second way in which PCA can be used in bitemporal spectral change detection is first to perform the transformation on each image and then to use one of the other techniques discussed in this section (e.g., image differencing, change vector analysis, classification, image cross-correlation) to assess whether there are important changes in the PC bands between dates. Again, however, this technique is not foolproof, as the axis orientations will be different for each date, producing results that are difficult to interpret. That being said, this technique can still be very valuable, particularly when the images are very similar (thus ensuring a similar axis rotation) except for the change being assessed.

The TC transformation uses the same coefficients (for a given sensor) to reorient the axes every time, and for this reason it circumvents many of the issues associated with the PCA's image-dependent nature. Most commonly, particularly when a mid-infrared band is available, the first three axes (*brightness*, *greenness*, and *wetness*), or some derivation therefrom (such as the greenness-brightness ratio), are used. Analysis can then proceed using one or more of the techniques discussed in this section.

Classification

Classification is often underutilized in spectral change detection, but it is extremely powerful, whether alone or in combination with other techniques discussed in this section. One reason for its use is that we often want to not only *detect* change but *attribute* it. As an example, we may want to know not only if there is a loss of forest cover, but also, if so, whether the lost forest became a pasture, strip mine, or suburban development. Classification can occur using the whole area of interest, or it can be used only to attribute change once the areas in which change has occurred have been identified employing some other technique.

If used alone on the whole area of interest, a common technique is to use a composite image of either the spectral or transformed bands (e.g., TC) as the source of data. The brightness value vectors from the composite, whether they be from pixels or multitemporal segments, are made up of bands from two dates. When the change of interest results in separable bitemporal spectral signatures, the classification protocol and decision rule (be it supervised or unsupervised, parametric or nonparametric) can then identify the change of interest. However, there may be other precursors to the classification other than the spectral or transformed bands. These might include, for example, the magnitude and direction from change vector analysis or the Pearson product-moment correlation coefficient from image cross-correlation (see the next section). As such, a variety of approaches or combinations of approaches can be used, depending on the specific needs of the analysis.

Statistical Techniques

There are numerous ways to compare images statistically, but one of the most widely used techniques is image cross-correlation using pixel neighborhoods or multitemporal segments. The mathematics are the same regardless, but this discussion will presume use of the latter. The equation forms used are those of Im and Jensen (2005) and Im, Jensen, and Tullis (2008), and they use only a single band for simplicity. Multiband correlations are feasible, but absent some method of combining the bands a priori (such as a vegetation index) will result in a correlation matrix, thus increasing the complexity of interpretation.

The Pearson product-moment correlation, r, between brightness values (or reflectances) from a single band across two dates of imagery is defined as

$$r = \frac{\text{cov}_{12}}{s_1 s_2} \quad \text{(EQ. 15.11)}$$

where s_1 and s_2 are the standard deviations of the brightness values found in each segment and cov_{12} is the covariance between brightness values of the two dates for the segment.

The standard deviation for a given band, date, and segment is as follows:

$$SD = \sqrt{\sum \frac{(y - \overline{y})^2}{n - 1}} \quad \text{(EQ. 15.12)}$$

where n is the number of pixels in the segment, y is the brightness value of each pixel, and $\overline{y}$ is the mean of all brightness values in the segment.

Standard tests for the significance of the resulting correlation(s) can be used as an additional variable, as can the slope and intercept derived from a least squares analysis of the brightness value pairs within the segment (Im and Jensen, 2005; Im et al., 2008).

Choosing a Technique

The first and most important choice is usually whether to first classify the data from each date and then proceed with a postclassification change detection. If the classification accuracy will be high, and the classification process objective, with both images, this is a clear choice. However, if as is often the case, the classification accuracy will neither be high nor the process objective, then spectral change detection presents a viable alternative as long as the change of interest is spectrally distinguishable. Visual interpretation can often help the analyst ascertain (1) whether the change of interest is distinguishable, and (2) if so, which of the several techniques presented here, alone or in combination, might be most useful. Knowledge of the spectral properties of the material type or conditions of interest—and how they change—is essential.

15.3 MULTITEMPORAL SPECTRAL CHANGE DETECTION

No-Cost Data Enabling New Class of Analyses

Data policies affording access to no-cost satellite data at a coarse resolution have been common within the meteorological community for decades. However, the land remote sensing community, typically desiring to address the scale at which land is used and managed (and thus changes) has required moderate- to high-resolution data. As discussed in other chapters, even satellites built, launched, and operated using public funds have had cost recovery mechanisms in place that effectively limited data availability (and thus algorithm development). However, analysis-ready Landsat data are now available for preview and download at no cost through the U.S. Geological Survey (USGS) Earth Explorer (for more details, see the discussion in Section 7.9). Data from the two Sentinel-2 satellites are

also available at no cost. These data are opening up a rich set of applications never before possible, as analysts can now use *multitemporal* stacks of images, be they of original or derived bands, to conduct their analyses.

Trade-Offs

Before getting into some of the details on how to conduct these analyses, it is worth thinking about some of the advantages and disadvantages of interannual multitemporal change detection. Advantages include the following:

1. Better separation of low-magnitude change from scene differences not arising from changes in land use or management (e.g., Sun or sensor angle, phenology, and atmospheric variability).
2. Slow rates of change (e.g., vegetation growth, ecological succession, and drought- or pathogen-induced mortality) can often be detected just as well as abrupt phenomena (e.g., vegetation clearing or fire).
3. Annual time steps are often better suited to many socioeconomic and natural processes than longer time steps.

There are also some disadvantages associated with multitemporal spectral change detection, including the following:

1. Persistent cloud cover over a given area that spans multiple years of images must be explicitly addressed for most algorithms to function (this is often the largest constraint to analysis).
2. Images must be geometrically and radiometrically the same as much as possible.
3. Phenology, drought, uncorrected atmospheric or sensor effects, and other factors can impact reflectance in a manner unrelated to the signal of interest (e.g., vegetation recovery).
4. Temporal signatures, while useful, require calibration to the timing of an event of interest (e.g., vegetation clearing) that can occur across a range of years.
5. Data volume can be very high.
6. Algorithm development is still maturing, with few commercial remote sensing image processing packages fully supporting this type of analysis.

Preprocessing

While approaches to analyze interannual multitemporal stacks of (primarily Landsat and Sentinel) images are rapidly evolving, a close look at some of the seminal early studies characterizing forest disturbance and recovery can help inform the preprocessing and algorithm selection for any given problem and data set. While steps can and do differ from application to application, the following preprocessing steps are widely utilized:

1. Acquire images at or near the same time of year and phenological condition.
2. Ensure the images are well co-registered.

3. Convert to top-of-atmosphere or surface reflectance (Chapter 11) if analysis-ready data are not available.
4. Remove clouds (usually a combination of semiautomated and manual techniques).
5. (Optionally) choose or derive a single band with which to work.
6. Normalize the image to remove remnant atmospheric, phenological, or sensor effects.
7. (Optionally) segment the image into objects in which spatially juxtaposed pixels with similar spectral and/or temporal trajectories are grouped (*sensu* Budreski et al., 2007; Lhermitte et al., 2008).

As noted earlier, cloud cover often forms a major constraint. For interannual multitemporal analyses in which only one image per year is used, the analyst will, of course, choose the best image for a given year that minimizes cloud cover, undesired phenological effects, and the like. However, for most areas the requirement of having one image per year inevitably leads to having to use some images that are not ideal, particularly with respect to cloud cover. Once the clouds are removed, the first question that arises is how to address the resulting gaps in the time series for a given area. This is an active area of research, but in general there are only three principal choices: (1) (for small areas) interpolate using adjacent pixel values in the same image, (2) interpolate using images acquired near the same date as the preferred image, or (3) use techniques by which missing values are explicitly handled in the subsequent analysis (e.g., exponentially weighted moving average change detection [EWMA-CD] or continuous change detection and classification algorithm [CCDC], see below).

A second constraint to multitemporal analysis is often the remnant variability in the surface reflectance not due to targeted changes in surface material type or condition. Often these effects are the result of subtle changes in phenology, but there are a host of other possible sources of this variability. The image normalization mentioned in step 6 is designed to address this remnant variability in reflectance at the preprocessing stage. This can be achieved through a variety of tools, but the analyst should be very careful in their selection and use, as all change the data from their original values. While the hope is that these changes help separate the signal of interest from the noise, in many instances use of these tools reduces the quality of the resulting analysis. While discussion of all the possibilities is beyond the scope of this book, area-based image matching techniques include histogram equalization, image cross-correlation, image-to-image regression, phase correlation, and maximization of mutual information (Zitová and Flusser, 2003). The reader should consult Zitová and Flusser (2003) plus standard remote sensing image processing texts (e.g., Richards, 2013) for additional information.

The particular demands of multitemporal image analysis have also given rise to their own class of image matching techniques. These are based on the idea of normalizing to a reference material type or condition. For example, for deforestation studies, it might be helpful to first identify forest that is undisturbed throughout the period of inquiry. In the next step, the difference between the mean brightness value or reflectance vector for the reference forest and each pixel's or object's brightness value or reflectance vector is calculated, as shown in the following equation (Browder, Wynne, and Pedlowski, 2005):

$$d = \sqrt{(\mathbf{f} - \mathbf{m}_k)(\mathbf{f} - \mathbf{m}_k)'} \qquad \text{(EQ. 15.13)}$$

where d is the Euclidean distance, $\mathbf{f}$ the mean brightness value vector for primary forest, and $\mathbf{m}$ the mean brightness value vector for each pixel, k. Browder et al. (2005) used this approach as a preprocessing step to assess whether the species planted in agroforestry plots in Rondônia made any difference to forest recovery. As shown in Figure 15.4, the onset of active management (1993) resulted in a notable increase in the distance from forest in all groups due to the clearing of vegetation in the plot establishment phase. The growth in the year following plot establishment resulted in a decrease in distance to forest as vegetation cover returned to the plot, but the degree to which the groups were temporally coherent after the establishment phase varied. The plot composition (and resulting management) differences in the three groups appear to be associated with different land-use strategies, with the *timber* group being most proximate to the spectral characteristics of primary forest toward the end of the project period.

Healey et al. (2005) took a different (but related) approach to both the image normalization and creation of a single derived band, the disturbance index (DI) based on the first three TC bands (Kauth and Thomas, 1976; Crist and Cicone, 1984a), brightness, B; greenness, G; and wetness, W. Each TC band is rescaled to its standard deviation above or below the scene's mean undisturbed forest value, as follows:

$$B_r = \frac{(B - \mu_B)}{\sigma_B} \quad \text{(EQ. 15.14)}$$

$$G_r = \frac{(G - \mu_G)}{\sigma_G} \quad \text{(EQ. 15.15)}$$

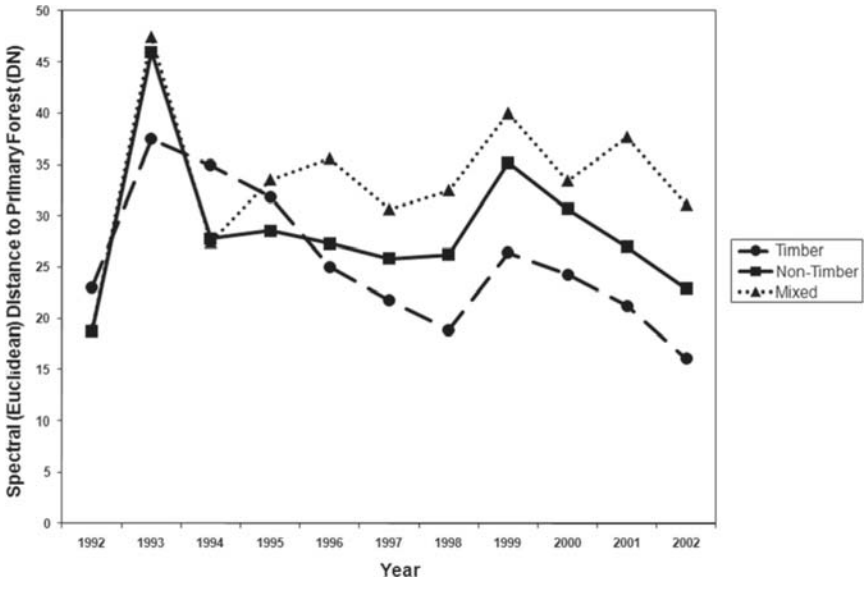


FIGURE 15.4 Mean spectral distance from primary forest by agroforestry group type. Euclidean distance between reference forest condition and agroforestry practices using Landsat TM/Enhanced Thematic Mapper Plus imagery from 1992 to 2002. From Browder et al. (2005). Used by permission of Springer.

$$W_r = \frac{(W - \mu_W)}{\sigma_W} \quad \text{(EQ. 15.16)}$$

where B_r, G_r, and W_r are the rescaled TC indices expressed as *z-scores*; μ_B, μ_G, and μ_W are the means of the TC indices in undisturbed forest; and σ_B, σ_G, and σ_W are the standard deviations of the TC indices in undisturbed forest. Note that this method explicitly addresses interband differences in variance. Healey et al. (2005) point out that recently cleared forest exhibits high brightness and low greenness and wetness in relation to undisturbed forest. *DI* is defined to take advantage of these disturbance related changes in the TC indices, as follows:

$$DI = B_r - (G_r + W_r) \quad \text{(EQ. 15.17)}$$

z-scores can, of course, be calculated using raw as well as derived bands, and other means can be used to combine them other than the one proposed by Healey et al. (2005). For example, Huang et al. (2010) show that the mean of the squared *z*-scores computed for the red and two shortwave infrared region (SWIR) bands can be used in a manner similar to the disturbance index.

Regardless of technique chosen, the analyst should realize that in many cases some sort of image normalization or matching procedure may be necessary to maximize signal and minimize noise across multiple scenes, even when using surface reflectance data.

Time Series Analysis

A *time series* is a sequence of data points evenly sampled through time. Classical analysis of time series can be used with image time series stacks. While specificity on these techniques is beyond the scope of this book, there is merit in a brief discussion (especially of key constraints to implementation) given the clear relevance to this chapter. For readers who wish to follow up, we recommend any of several good introductory texts on time series analysis (e.g., Montgomery, Jennings, and Kulahci, 2008).

Assuming an evenly sampled data set, classical time series analysis can separate trends in the data from periodic fluctuations (often due to seasonality) and measurement error (noise). However, two major constraints to time series analysis using imagery are (1) the presence of step changes in reflectance caused by land cover change such as forest clearing, and (2) the need to have the data points evenly sampled. The latter is typically quite unlikely using moderate resolution Earth resource satellite data, as the analyst typically has to choose the best possible image for a given period under other constraints (such as cloud cover or being within a certain part of the growing season). This has two major deleterious effects on time series analysis. First, the time series is then inherently unevenly sampled. Second, and perhaps more perniciously, subtle differences in reflectance due to phenology are often introduced. New methods to predict daily surface reflectance at moderate (e.g., Landsat) resolution have been introduced (Gao et al., 2006; Brooks et al., 2012). These show promise for addressing both of these issues when appropriate imagery is available, and when coarse resolution data are also used (e.g., Gao et al., 2006), landscape fragmentation is not overly high.

There are three categories of time series algorithms (Saxena et al., 2018, liberally quoted given the shared authorship), as follows:

- *Kernel regression methods*. These methods represent the time series as a linear combination of *basis functions*, "a set of functions that can be linearly combined to form a more general set of functions" (American Meteorological Society, 2015). Typically, a linear system of equations is solved to determine the coefficients. Analysis and predictions are done based on this representation.

- *Top-down approaches*. In these algorithms, an approximation is first made to the whole time series. Then, typically using error estimate criteria, finer partitions of the time series are sought so that each new partition is a refinement of the previous partition. This is repeated until either a maximum number of iterations is reached or each segment of the partition satisfies a convergence criterion.

- *Bottom-up approaches*. Bottom-up approaches represent the most elementary units of the data first; on each iteration, increasingly larger structures are composed from the simpler structures and their evaluations in the previous iteration. Dynamic programming and all recursive algorithms classify as bottom-up algorithms, which are frequently implemented using backtracking, "a method of searching . . . in stages, beginning with a depth-first search" (American Meteorological Society, 2015).

Table 15.1 labels the approach taken by each of seven multitemporal change detection algorithms.

Analysis of Vegetation Disturbance and Recovery

Time series analysis will form an ever larger portion of the suite of tools used for multitemporal change analysis. Unfortunately, many methods are quite complicated mathematically and thus are beyond the scope of an introductory course. However, we highlight some examples, which can give the beginning remote sensing student an entry point into this very active area of research and applications.

Control-chart change detection (e.g., Brooks et al., 2012, 2014) is an example of a *kernel regression* method that can be used with well-calibrated dense time series stacks available from Landsat, Sentinel-2, and the like. Polynomial or, more usually, *harmonic regression* (Figure 15.5, Brooks et al., 2012; Equation 15.18, *sensu* Brooks et al., 2012, but after Zhu and Woodcock, 2014) is first used to establish the typical intra-annual pat-

TABLE 15.1 Classification of Recent Multitemporal Change Detection Algorithms

Classification	Algorithms
Kernel regression	CCDC[a], EWMA-CD[b], ShapeSelectForest[c]
Top-down	LandTrendR[d], VeRDET[e]
Bottom-up	BFAST[f], VCT[g]

Note: BFAST = Breaks For Additive Seasonal and Trend; CCDC = continuous change detection and classification; EWMA-CD = exponentially weighted moving average change detection; LandTrendR = Landsat-based Detection of Trends in Disturbance and Recovery; VCT = vegetation change tracker; VeRDET = Vegetation Regeneration and Disturbance Estimates through Time. From Saxena et al. (2018). Copyright © 2018. Reprinted by permission of Elsevier.

[a]Zhu and Woodcock (2014). [b]Brooks et al. (2014). [c]Moisen et al. (2016). [d]Kennedy et al. (2010). [e]Hughes et al. (2017). [f]Verbesselt et al. (2010). [g]Huang et al. (2010).

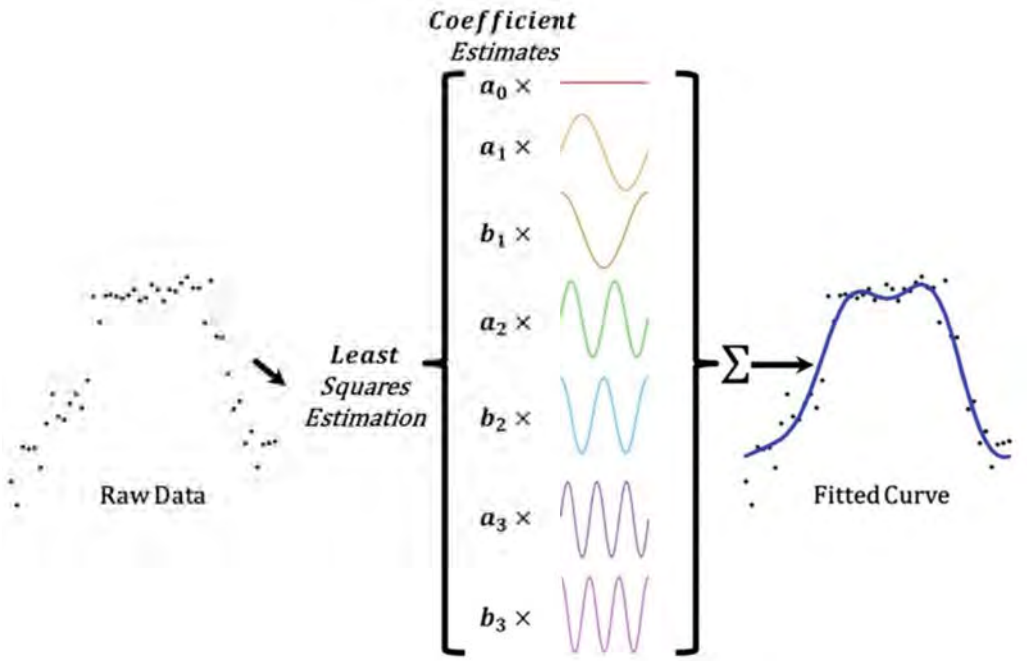


FIGURE 15.5 Harmonic regression. From Brooks et al. (2012). Copyright © 2012 IEEE. Used by permission.

tern (the *process*) in a given pixel for a given spectral band or vegetation index using a least squares fit on all available cloud-free images for a given pixel/year:

$$\hat{\rho}_x = a_0 + a_1 \cos\left(\frac{2\pi}{T}x\right) + a_2 \sin\left(\frac{2\pi}{T}x\right) \quad \text{(EQ. 15.18)}$$

where

a_{0-2} = harmonic regression coefficients (0, overall value / constant; 1–2, cosine and sine for intra-annual change)

T = days / year (365)

x = date (Julian)

$\hat{\rho}_x$ = predicted value at Julian date x

Multiple years of data can be used to establish this pattern, but care must be taken that the land use or management practices for a given use did not change during the multiyear period. An additional bonus is that the coefficients for the resulting regression can themselves be used in subsequent analyses and serve as a de facto compression mechanism, as they describe the temporal trajectory of a given band or ratio throughout the year for a given pixel (Figure 15.6).

Once that pattern (including variation inherent to the modeled process) has been established, departures from the pattern can be assessed using *statistical control charts*,

"graphs used to study how a process changes over time" (ASQ, 2018). Small shifts (such as forest growth or forest thinning) in the modeled process are best modeled using an EWMA control chart (Figure 15.7; Brooks et al., 2014). Larger process shifts (such as remnant clouds, stand-replacing disturbances, and land cover change) are best detected using Shewhart X-bar charts, as also noted by Brooks et al. (2014).

One issue with control chart change detection is that the "process" being "controlled" is, in general, an intra-annual phenological pattern that is assumed to be stable across time. However, for areas experiencing constant gradual change (e.g., forest growth), this assumption is not met. One option is to signal the departure from the extant process and then (perhaps nearly continually) retrain (*sensu* Brooks et al., 2017).

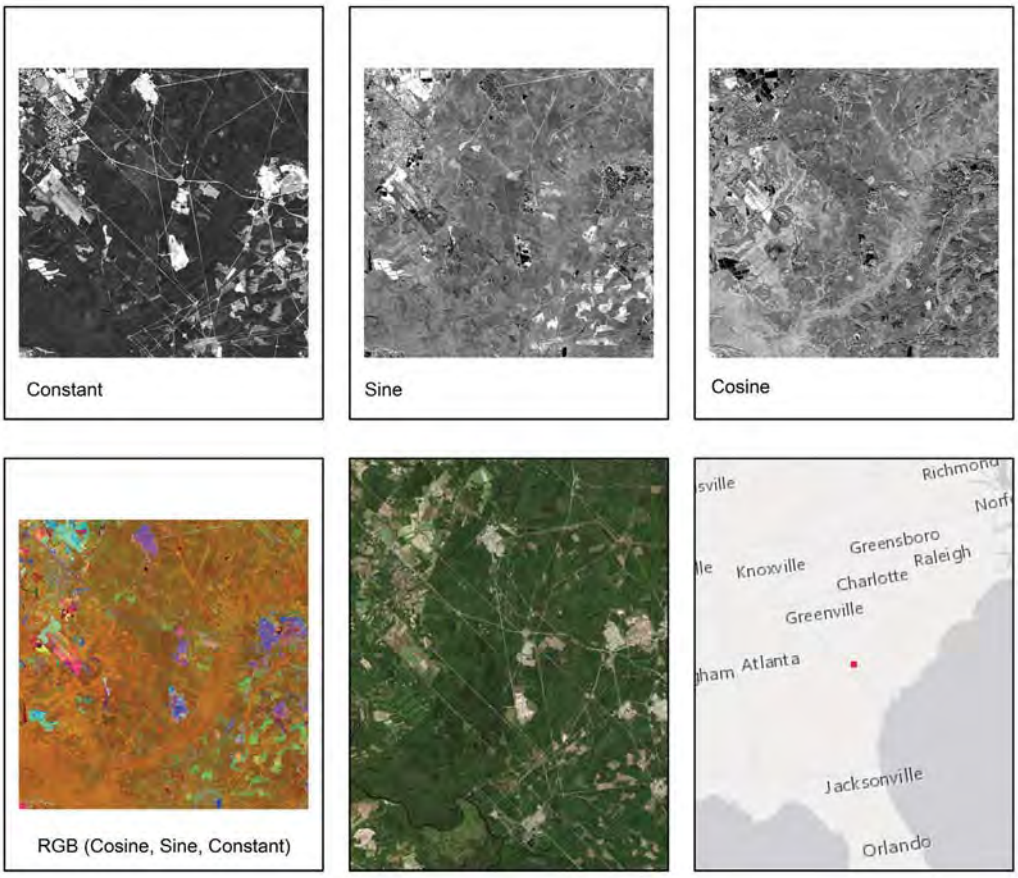


FIGURE 15.6 Harmonic regression coefficients for the Landsat 8 panchromatic band (one harmonic) using all available Landsat data from 2014 to 2016 in South Carolina. Note that only three bands of data are needed to fully represent scores of input images, also enabling simulation of the value of a given band or ratio for any day of the year. The first four panels are from Jill Marie Datsko. Used by permission. The last two panels are from ESRI's ArcMap. Copyright © 2021 ESRI and its licensors. Used by permission.

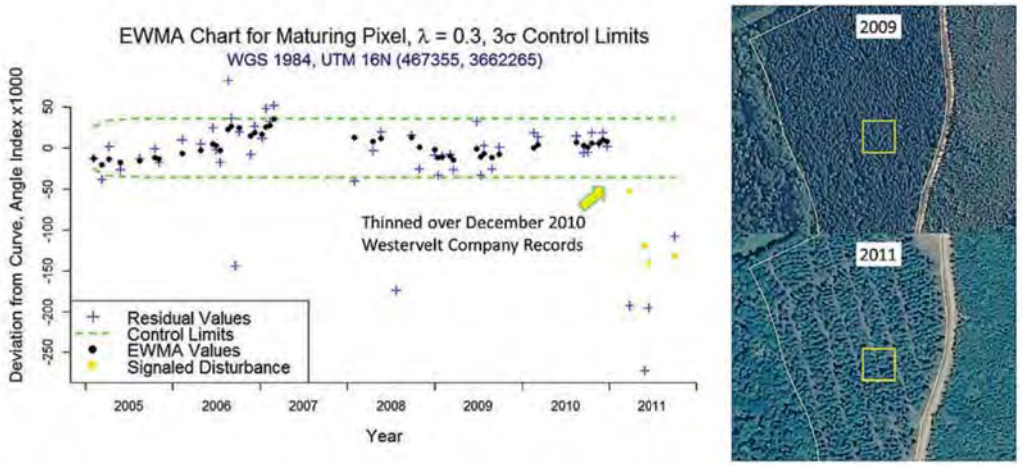


FIGURE 15.7 Exponentially weighted moving average chart for residual values after removing seasonality. The forested Landsat pixel (yellow boxes on 2009 and 2011 National Agriculture Imagery Program images at right) was thinned in December 2010. From Brooks et al. (2014). Copyright © 2014 IEEE. Used by permission.

Another option is to concomitantly model both the intra-annual and interannual trends, resulting in, for one harmonic, four coefficients rather than three (constant, sine, cosine, and interannual trend):

$$\hat{\rho}_x = a_0 + a_1 \cos\left(\frac{2\pi}{T}x\right) + a_2 \sin\left(\frac{2\pi}{T}x\right) + a_3 x \qquad \text{(EQ. 15.19)}$$

where

a_3 = coefficient for linear interannual change trend

This is the approach used by the CCDC algorithm (Zhu and Woodcock, 2014), which, as added benefits, can (1) explicitly model abrupt change as time series breaks (not shown in Equation 15.20) and (2) models all available wavelength bands for each pixel, retaining the root mean square error (RMSE). Change is flagged in the original version of CCDC if the following condition is true for three consecutive new observations (Zhu and Woodcock, 2014):

$$\frac{1}{k}\sum_{i=1}^{k}\frac{\left|\rho_{i,x} - \hat{\rho}_{i,x}\right|}{3RMSE_i} > 1 \qquad \text{(EQ. 15.20)}$$

where

x = Julian date
i = the ith spectral band/index

k = the number of spectral bands/indices
$\rho_{i,x}$ = observed value for the ith spectral band at Julian date x
$\hat{\rho}_{i,x}$ = predicted value for the ith spectral band at Julian date x

Notice that this approach is explicitly multispectral, presenting an advantage over most commonly implemented control chart change detection methods. In addition, since interannual trends, such as forest growth, are explicitly incorporated, they do not themselves constitute change ipso facto, which is an advantage for most protocols. Like control charts, the CCDC algorithm uses the inherent variability (through the RMSE), has control limits (i.e., 3 × RMSE), and requires multiple observations (three) to flag a true change.

One potential issue with both control chart-based change detection and CCDC is that there is a de facto assumption of a stable intra-annual process, but this is not always the case. Take, for example, an extremely dry area that only receives enough rainfall for seeds to sprout on a quite irregular basis, or an inherently mixed pixel with rapidly changing land cover composition (a situation common in some developing peri-urban areas and in smallholder agriculture). Additional research is under way to address these issues, but, in general, the practitioner can feel comfortable with these algorithms for most typical applications.

Correctly flagging change, however, is only part of the information need. As noted earlier in the discussion of bitemporal change detection, *attribution* of a flagged change is important for many applications. By attribution we mean knowing what caused the change or, more subtly, characterizing the material type or condition prior to and subsequent to the flagged change. The pattern predisturbance, the magnitude of the disturbance, and the pattern postdisturbance are often combined algorithmically to both flag and attribute change and its aftermath. While changes in vegetation are clearly just a subset of possible changes of interest, many tools for analysis of vegetation *disturbance* and *recovery* have been developed and have merit by themselves or with simple extensions to other multitemporal change analysis problems.

There are many ways to analyze time series stacks that explicitly utilize the pattern predisturbance, the magnitude of the disturbance, and the pattern postdisturbance. Identifying surface coal mines in the southern Appalachians and analyzing their postreclamation vegetation recovery using interannual, multitemporal Landsat data provide a good illustration of typical steps. It should be noted that this is but an illustrative example; a wide range of both applications and analytical approaches is feasible. **Figure 15.8** shows some of the key features of such an analysis through analysis of vegetation index values at or near the annual maximum for the life cycle of a surface coal mine in the southern Appalachians.

Most important are the features that are diagnostic of the disturbance and its type. Notice the substantial drop (in 1997) as the clearing of the native hardwoods precedes mining. The three years of the mining period have only minor fluctuations in the vegetation index value. However, the area is reclaimed after the coal has been extracted, leading to a rapid recovery (starting in 2000) of the vegetation. Note, however, that even after 8 years the vegetation index has not yet recovered its value prior to the onset of mining.

Both the magnitude of the drop and the slope of the vegetation recovery are diagnostic of the type of disturbance (**Figure 15.9**). The mined areas are more completely

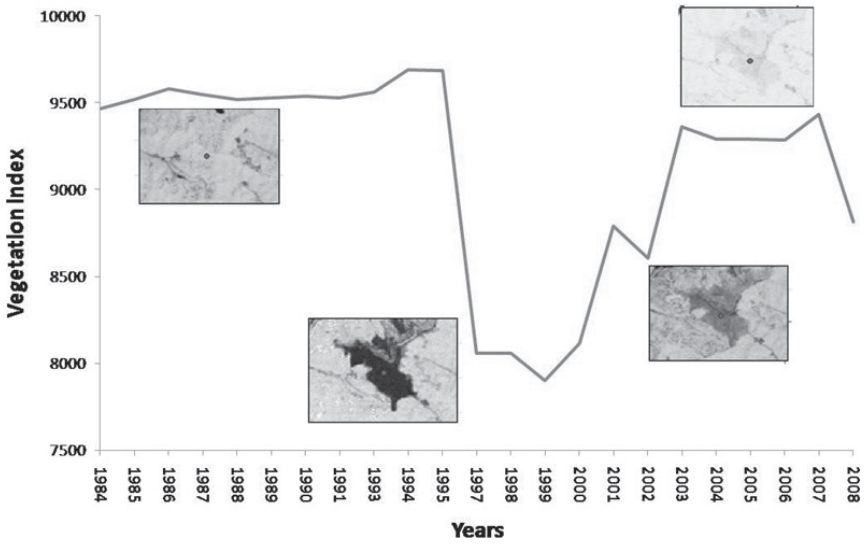


FIGURE 15.8 Vegetation index trajectory (1984–2008) for a mining disturbance that is reclaimed promptly compared with other trajectories common in the area. (Vegetation index images are shown at key points in the trajectory.) Vegetation index trajectories contrasted to the background undisturbed forest of surrounding areas (left) and typical recovery trajectories of other common disturbances (right). The horizontal time dimension constitutes several decades. Image by Susmita Sen.

cleared of vegetation and thus have the most precipitous drop, more than either forest fires or clearing for urban development (in this area). Furthermore, recovery does not begin immediately in mined areas, but only after the cessation of mining and the onset of reclamation. The recovery is steep for mines but more gradual (with a lower asymptote) for urban development.

As such, the following components of a typical mining trajectory can be used to discriminate mining from other common disturbances:

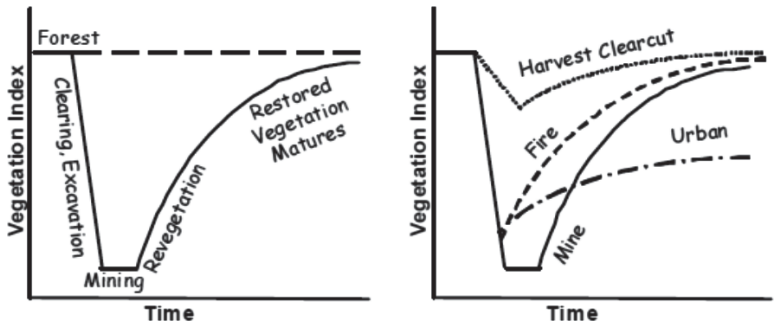


FIGURE 15.9 Interannual multitemporal trajectory of a vegetation index for a surface coal mine in the southern Appalachians. Notice the significant drop in vegetation as the area to be mined is cleared of trees, and then observe its gradual recovery after reclamation. Adapted from Sen et al. (2012). Used by permission of the American Society for Photogrammetry and Remote Sensing.

1. The *minimum* value that equals or is less than a threshold (the mining delineation threshold): the sharp drop of the vegetation index, caused by the rapid clearing of vegetation and surface soil to expose underlying geologic material, can discriminate mining from less drastic landscape disturbances such as forest harvests and forest fires.
2. The *slope* of the recovery trajectory: following the drop, the trajectory takes an upward trend due to vegetation development. The recovery is steepest during its first 3 to 4 years, and the slope of the trajectory in these years can be used to discriminate mining from disturbances such as urban development and road construction.
3. The recovery *maximum*: the maximum value in the recovery phase can be used to help assess the quality of the reclamation process or separate mining from other disturbances.

House and Wynne (2018) automated the process of calculating these and related variables (Figure 15.10) through release of the Disturbance Detection and Diagnostics (D^3)

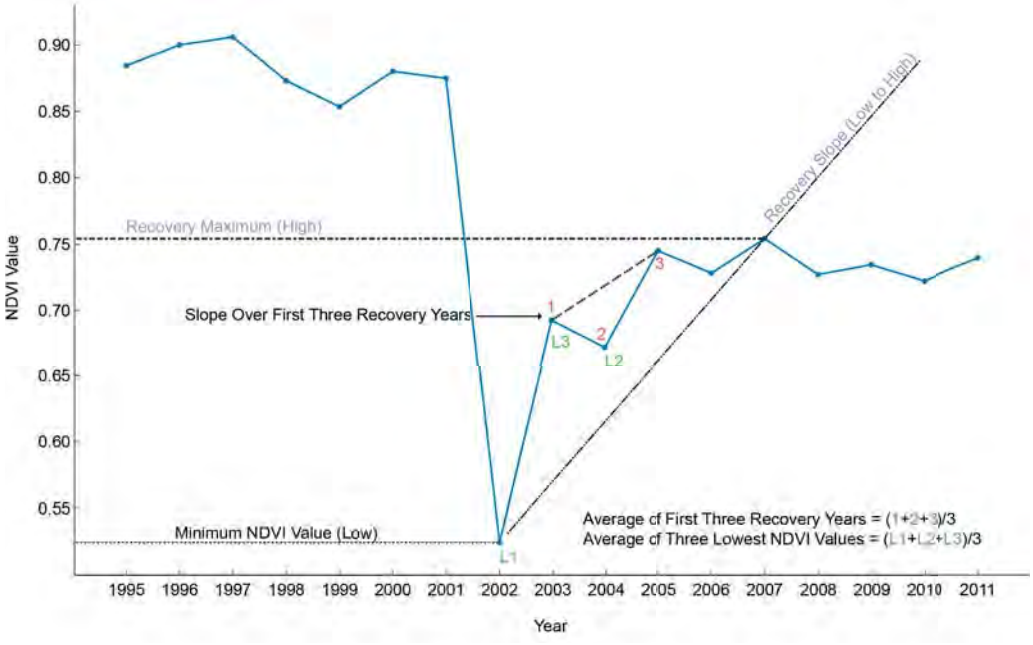


FIGURE 15.10 The six predictor variables calculated by D^3 (House and Wynne, 2018). The same type of disturbance and recovery reflectance or vegetation index trajectory can lend itself to many other parameters describing its shape. For example, Kennedy et al. (2007) defined four disturbance classes (disturbance, disturbance and recovery, recovery, and recovery to stable state). For hypothesized trajectories describing each of these classes they defined three shape parameters, as follows: (1) the end year of the disturbance, (2) the pre-disturbance mean, and (3) the post-disturbance mean. Each pixel was classified based on the best nonlinear least-squares fit between the observed spectral trajectory (from Landsat TM/ETM+ band 5) and each of the four hypothesized trajectories.

algorithm (which can be found at *https://doi.org/10.7294/W4FB513G*). The refined algorithm has been shown to be effective in separating low-density development from actively managed forests (House and Wynne, 2018). Kennedy et al. (2007) defined four disturbance classes (disturbance, disturbance and recovery, recovery, and recovery to stable state). See **Figure 15.11**.

Focusing only on recovery trajectories, Schroeder, Cohen, and Yang (2007) first mapped clearcuts and then used ISODATA clustering to develop and map four forest regrowth classes (*little to no, slow, moderate,* and *fast*) developed using Landsat TM/ETM+ data. Third-order polynomial curves fit to each of three periods' forest regrowth classes are shown as **Figure 15.12**.

Moisen et al. (2016) extend this concept further by fitting nonparametric shape-restricted regression splines to time series. The seven predefined temporal patterns (*shapes*) are shown in **Figure 15.13** for SWIR1. While the specifics of the algorithm are beyond the scope of this book, this is a very important advance, as change detection and attribution (causal agent of the disturbance, *sensu* Schroeder et al., 2017) can be addressed concomitantly.

The above examples are helpful in understanding the important role that multitemporal change detection can play in analyzing disturbance and recovery. The important points to remember are as follows:

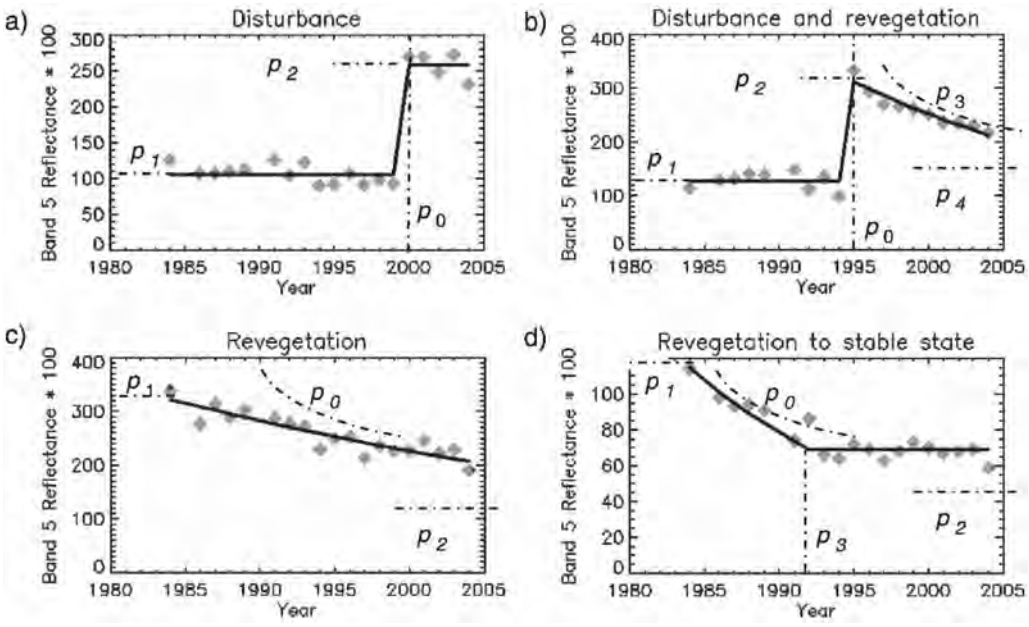


FIGURE 15.11 A detailed view of fitted values from four pixels representing four hypothesized models of disturbance or recovery. (a) Simple disturbance, (b) disturbance followed by exponential revegetation, (c) revegetation from disturbance prior to observation record, (d) revegetation from prior disturbance, reaching a stable point during the observation period. From Kennedy et al. (2007). Copyright © 2007. Used by permission of Elsevier.

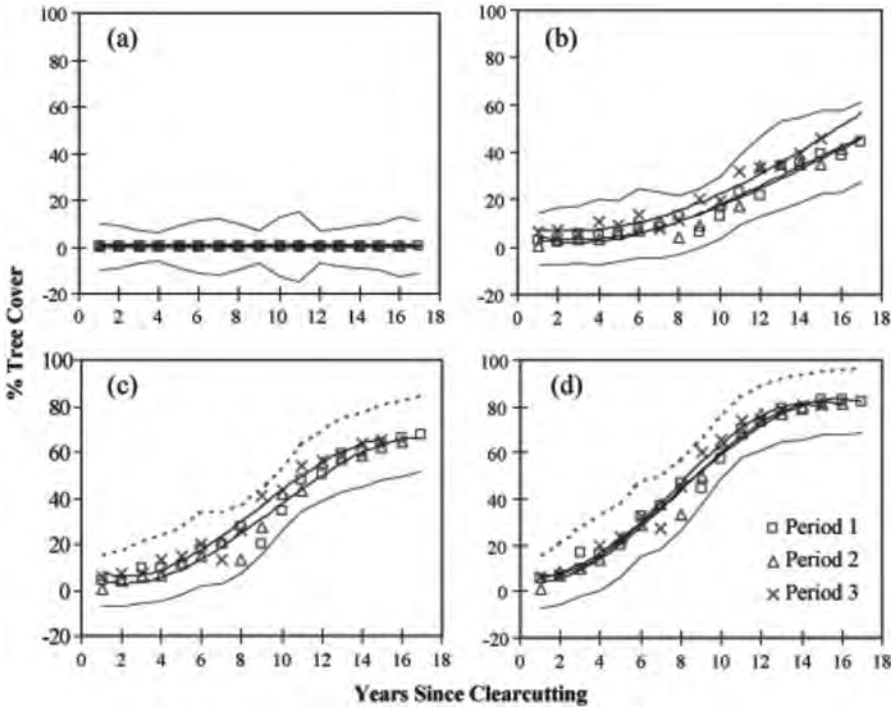


FIGURE 15.12 Mean forest regrowth trajectories for (a) little to no, (b) slow, (c) moderate, and (d) fast regrowth classes. Solid lines are fitted third-order polynomial curves; dashed lines are the average across period standard deviations. From Schroeder et al. (2007). Copyright © 2007. Used by permission of Elsevier.

1. Local knowledge of the land use and land cover transitions that are typical in the area is essential.
2. Analysis of the resulting time series requires a close understanding of the spectral dynamics associated with land use and land cover change.
3. Spectral trajectories require a new way of thinking about and analyzing remotely sensed data.

It is important to recognize that while these examples focus primarily on vegetation disturbance and recovery, there is nothing to preclude using similar approaches for more traditional classification problems, including vegetation mapping. In Figure 15.14, for example, invasive shrub species are separable from native hardwoods and pines because of their growth trajectories. Figure 15.15 shows how the minimum vegetation index in a time series, along with the slope to the maximum value postrecovery, enables separation of low-density development from actively managed forests.

For the analyst or student interested in analyzing spectral trajectories, we must reiterate the importance of appropriate preprocessing. After that, it is extremely important to know your area and problem well to choose the best technique—as is true with all remote sensing. Increasingly, practitioners are realizing that there is very little agree-

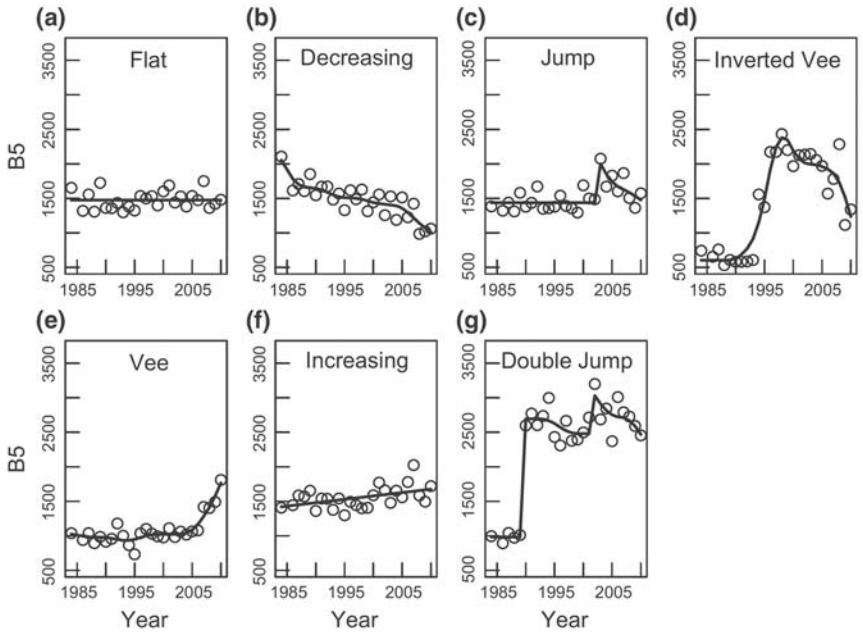


FIGURE 15.13 Seven possible shapes describing temporal patterns in SWIR1 (here denoted as B5) as described by Moisen et al. (2016). Note that the patterns are inverted from what would be expected if a vegetation index had been used.

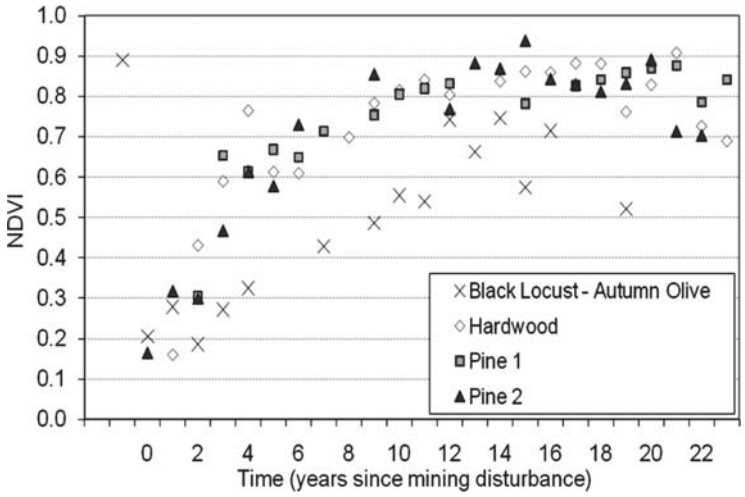


FIGURE 15.14 Invasive species detected using interannual, multitemporal image chronosequences. NDVI trajectories are from disturbed sites in Virginia and West Virginia that have developed full canopies of differing woody species types. The trajectories were not coincident in time but have been plotted with "Year 0" as the estimated date of the original disturbance. Missing annual values, which are evident in the Black Locust—Autumn Olive data series, occur because of occasional obscuration by clouds.

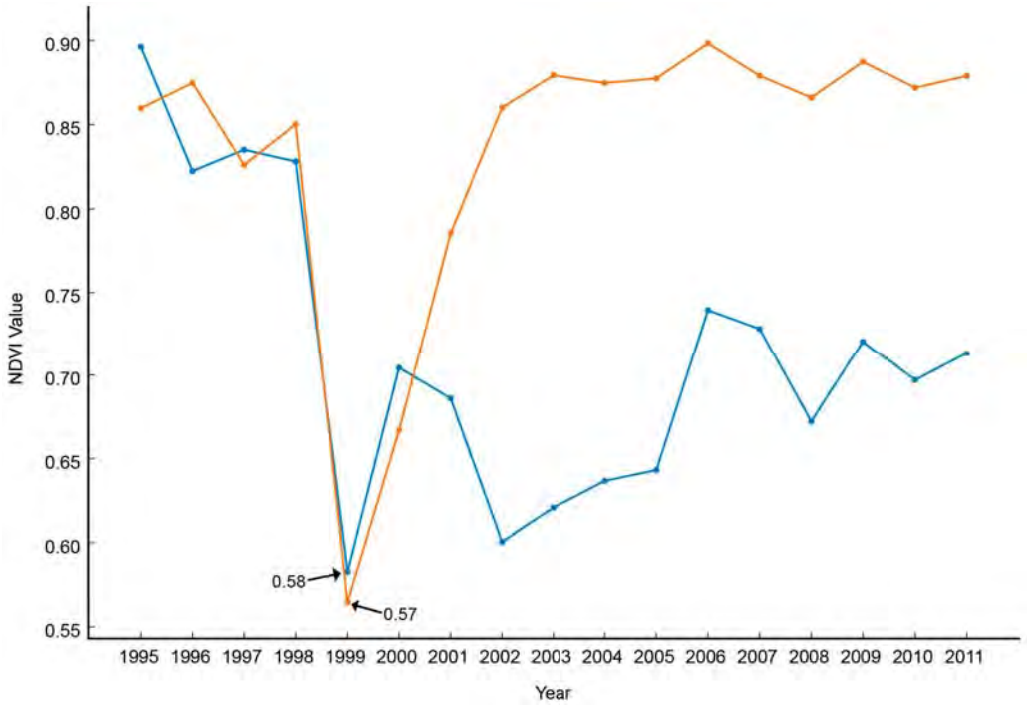


FIGURE 15.15 Example of exurban development (blue) compared to active forest management (orange) over time. Both pixels were flagged as having been disturbed in 2003. Exurban development remains lower than active forest management through the remainder of the time series and is much more constant (House and Wynne, 2018).

ment among multitemporal change detection techniques (Cohen et al., 2017). As such, an ensemble of techniques (Healey et al., 2018) or a polyalgorithm combining techniques (Saxena et al., 2018) might be an improvement over any given algorithm. Multitemporal change detection and attribution using time series stacks will burgeon with both innovation and increasingly diverse applications in the coming years as both data and appropriate analytical tools become increasingly available.

15.4 SUMMARY

Change detection, whether it be bitemporal or multitemporal, is a mainstay of remote sensing analyses. Its importance is magnified by the extensive changes in land use occurring with swelling—and increasingly well-off—human populations. No-cost well-preprocessed data, increasingly available from programs such as Landsat (NASA and USGS) and Sentinel (European Space Agency), are essential to ensuring the data quantity and quality necessary for these analyses. Remotely sensed data are essential to our ability to monitor and model the changes we are making to the Earth's ecosystems—and to ensure their sustainability.

REVIEW QUESTIONS

1. Prepare a list of reasons why development of effective change detection capability is an important capability to support public policy at international, national, and local scales.
2. In your next job, you are called on to testify in court to defend the rigor of a change detection analysis that has been challenged as unscientific. Prepare notes that you would use to defend the validity of the analysis.
3. You do such a good job in preparing your responses for Question 2 that the opposition has now hired you to attack the validity of a change analysis in dispute in their next case. Prepare notes that you will use to question the validity of this new change analysis.
4. Can you identify landscapes that might be inherently difficult to analyze using spectral change detection?
5. Outline ideal circumstances for conducting change analysis using remotely sensed data.
6. Review the text for this chapter that describes preprocessing necessary for execution of change detection. Prepare a list of the most important considerations.
7. Even if images have been selected carefully and processed, change detection can be subject to errors. Review the text for this chapter to identify some of the sources of error that are inherent to the process of change detection.
8. Why are established techniques for analyzing time series seldom used in analysis of remotely sensed data?
9. For a given spatial resolution, there is often a level of landscape fragmentation that precludes accurate spectral change detection. Why might this be the case?
10. List three advantages and three disadvantages of multitemporal change detection as compared with bitemporal change detection.

REFERENCES

American Meteorological Society. 2015. Glossary of Meteorology. Available online via *https://definedterm.com*.

ASQ. 2018. The 7 Basic Quality Tools for Process Improvement. *http://asq.org/learn-about-quality/seven-basic-quality-tools/overview/overview.html*. Accessed June 29, 2018.

Brooks, E. B., V. A. Thomas, R. H. Wynne, and J. W. Coulston. 2012. Fitting the Multitemporal Curve: A Fourier Series Approach to the Missing Data Problem in Remote Sensing Analysis. *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 50, pp. 3340–3353.

Brooks, E. B., R. H. Wynne, V. A. Thomas, C. E. Blinn, and J. W. Coulston. 2014. On-the-Fly Massively Multitemporal Change Detection Using Statistical Quality Control Charts and Landsat Data. *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 52, pp. 3316–3332.

Brooks, E. B., Z. Yang, V. A. Thomas, and R. H. Wynne. 2017. Edyn: Dynamic Signaling of Changes to Forests Using Exponentially Weighted Moving Average Charts. *Forests*, Vol. 8, Article No. 304.

Browder, J., R. H. Wynne, and M. Pedlowski. 2005. Agroforestry Diffusion and Secondary Forest

Regeneration in the Brazilian Amazon: Further Findings from the Rondônia Agroforestry Pilot Project (1992–2002). *Agroforestry Systems,* Vol. 65, pp. 99–111.

Budreski, K. A., R. H. Wynne, J. O. Browder, and J. B. Campbell. 2007. Comparison of Segment and Pixel-based Non-parametric Land Cover Classification in the Brazilian Amazon Using Multitemporal Landsat TM/ETM Imagery. *Photogrammetric Engineering and Remote Sensing,* Vol. 73, pp. 813–827.

Cohen, W. B., S. P. Healey, Z. Yang, S. V. Stehman, C. K. Brewer, E. B. Brooks, N., . . . Z. Zhu. 2017. How Similar Are Forest Disturbance Maps Derived from Different Landsat Time Series Algorithms? *Forests,* Vol. 8, Article No. 98.

Coppin, P., I. Jonckheere, K. Nackaerts, B. Muys, and E. Lambin. 2004. Digital Change Detection Methods in Ecosystem Monitoring: A Review. *International Journal of Remote Sensing,* Vol. 25, pp. 1565–1596.

Crist, E. P., and R. C. Cicone. 1984a. Application of the Tasseled Cap Concept to Simulated Thematic Mapper Data. *Photogrammetric Engineering and Remote Sensing,* Vol. 50, pp. 343–352.

Crist, E. P., and R. C. Cicone. 1984b. A Physically-Based Transformation of Thematic Mapper Data: The TM Tasseled Cap. *IEEE Transactions on Geoscience and Remote Sensing,* Vol. GE-22, pp. 256–263.

Dai, X., and S. Khorram. 1999. Data Fusion Using Artificial Neural Networks: A Case Study on Multitemporal Image Analysis. *Journal of Computers, Environment, and Urban Systems,* Vol. 23, pp. 19–31.

Donoghue, D. N. M. 2002. Remote Sensing: Environmental Change. *Progress in Physical Geography,* Vol. 26, pp. 144–151.

Franklin, S. E., M. B. Lavigne, M. A. Wulder, and G. B. Stenhouse. 2002. Change Detection and Landscape Structure Mapping Using Remote Sensing. *Forestry Chronicle,* Vol. 78, pp. 618–625.

Gao, F., J. Masek, M. Schwaller, and F. Hall. 2006. On the Blending of the Landsat and MODIS Surface Reflectance: Predicting Daily Landsat Surface Reflectance. *IEEE Transactions on Geoscience and Remote Sensing,* Vol. 44, pp. 2207–2218.

Healey, S. P., W. B. Cohen, Z. Yang, C. K. Brewer, E. B. Brooks, N. Gorelick, . . . and Z. Zhu. 2018. Mapping Forest Change Using Stacked Generalization: An Ensemble Approach. *Remote Sensing of Environment,* Vol. 204, pp. 717–728.

Healey, S. P., W. B. Cohen, Y. Zhiqiang, and O. N. Krankina. 2005. Comparison of Tasseled Cap-Based Landsat Data Structures for Use in Forest Disturbance Detection. *Remote Sensing of Environment,* Vol. 97, pp. 301–310.

Hilker, T., M. A. Wulder, N. C. Coops, N. Seitz, J. C. White, F. Gao, and J. C. White. 2009. A New Data Fusion Model for High Spatial- and Temporal-Resolution Mapping of Forest Disturbance Based on Landsat and MODIS. *Remote Sensing of Environment,* Vol. 113, No. 8, pp. 1613–1627.

House, M., and R. Wynne. 2018. Identifying Forest Impacted by Development in the Commonwealth of Virginia through the Use of Landsat and Known Change Indicators. *Remote Sensing,* Vol. 10, No. 1, p. 135.

Huang, C., S. N. Goward, J. G. Masek, N. Thomas, Z. Zhu, and J. E. Vogelmann. 2010. An Automated Approach for Reconstructing Recent Forest Disturbance History Using Dense Landsat Time Series Stacks. *Remote Sensing of Environment,* Vol. 114, pp. 183–198.

Hughes, M. J., S. D. Kaylor, and D. J. Hayes. 2017. Patch-Based Forest Change Detection from Landsat Time Series. *Forests,* Vol. 8, Article No. 166.

Im, J., and J. R. Jensen. 2005. A Change Detection Model Based on Neighborhood Correlation Image Analysis and Decision Tree Classification. *Remote Sensing of Environment,* Vol. 99, pp. 326–340.

Im, J., J. R. Jensen, and J. A. Tullis. 2008. Object-Based Change Detection Using Segmentation Correlation Image Analysis and Image Segmentation. *International Journal of Remote Sensing,* Vol. 29, pp. 399–423.

Jackson, R. D., P. N. Slater, and P. J. Pinter. 1983. Discrimination of Growth and Water Stress in

Wheat by Various Vegetation Indices through Clear and Turbid Atmospheres. *Remote Sensing of Environment,* Vol. 13, No. 3, pp. 187–208.

Kauth, R. J., and G. S. Thomas. 1976. The Tasseled Cap: A Graphic Description of the Spectral–Temporal Development of Agricultural Crops as Seen by Landsat. In LARS: Proceedings of the Symposium on Machine Processing of Remotely Sensed Data. West Lafayette, IN: Purdue University Press, pp. 4B-41–4B-51.

Kennedy, R. E, W. B. Cohen, and T. A. Schroeder. 2007. Trajectory-Based Change Detection for Automated Characterization of Forest Disturbance Dynamics. *Remote Sensing of Environment,* Vol. 10, pp. 370–386.

Kennedy, R. E., Z. Yang, and W. B. Cohen. 2010. Detecting Trends in Forest Disturbance and Recovery Using Yearly Landsat Time Series: 1. LandTrendr—Temporal Segmentation Algorithms. *Remote Sensing of Environment,* Vol. 114, pp. 2897–2910.

Lhermitte, S., J. Verbesselt, I. Jonckheere, K. Nackaerts, J. A. N. van Aardt, W. W. Verstraeten, & P. Coppin. 2008. Hierarchical Image Segmentation Based on Similarity of NDVI Time Series. *Remote Sensing of Environment,* Vol. 112, No. 2, pp. 506–521.

Lu, D. P., Mausel, E. Brondizio, and E. Moran. 2004. Change Detection Techniques. *International Journal of Remote Sensing,* Vol. 26, pp. 2365–2401.

Lunetta, R., and C. Elvidge. 2000. *Remote Sensing Change Detection: Environmental Monitoring Methods and Applications*. London: Taylor and Francis, 320 pp.

Moisen, G. G., M. C. Meyer, T. A. Schroeder, X. Liao, K. G. Schleeweis, E. A. Freeman, and C. Toney. 2016. Shape Selection in Landsat Time Series: A Tool for Monitoring Forest Dynamics. *Global Change Biology,* Vol. 22, pp. 3518–3528.

Montgomery, D. C., C. L. Jennings, and M. Kulahci. 2008. *Introduction to Time Series Analysis and Forecasting.* Hoboken, NJ: Wiley-Interscience, 392 pp.

Richards, J. A., 2013. *Remote Sensing Digital Image Analysis: An Introduction.* Heidelberg: Springer, 474 pp.

Saxena, R. L. T., Watson, R. H. Wynne, E. B. Brooks, V. A. Thomas, Y. Zhiqiang, and R. E. Kennedy. 2018. Towards a Polyalgorithm for Land Use Change Detection. *ISPRS Journal of Photogrammetry and Remote Sensing,* Vol. 144, pp. 217–234.

Schroeder, T. A., W. B. Cohen, and Z. Q. Yang. 2007. Patterns of Forest Regrowth Following Clearcutting in Western Oregon as Determined by a Landsat Time-Series. *Forest Ecology and Management,* Vol. 243, pp. 259–273.

Schroeder, T. A., K. G. Schleeweis, G. G. Moisen, C. Toney, W. B. Cohen, E. A. Freeman, . . . and C. Huang. 2017. Testing a Landsat-Based Approach for Mapping Disturbance Causality in U.S. Forests. *Remote Sensing of Environment,* Vol. 195, pp. 230–243.

Sen, S., C. E. Zipper, R. H. Wynne, and P. F. Donovan. 2012. Identifying Revegetated Mines as Disturbance/Recovery Trajectories Using an Interannual Landsat Chronosequence. *Photogrammetric Engineering and Remote Sensing,* Vol. 78, No. 3, pp. 223–235.

Todd A., T. A. Schroeder, W. B. Cohen, and Z. Yang. 2007. Patterns of Forest Regrowth Following Clearcutting in Western Oregon as Determined from a Landsat Time Series. *Forest Ecology and Management,* Vol. 243, pp. 259–273.

Verbesselt, J., R. Hyndman, A. Zeileis, and D. Culvenor. 2010. Remote Sensing of Environment Phenological Change Detection While Accounting for Abrupt and Gradual Trends in Satellite Image Time Series. *Remote Sensing of Environment,* Vol. 114, pp. 2970–2980.

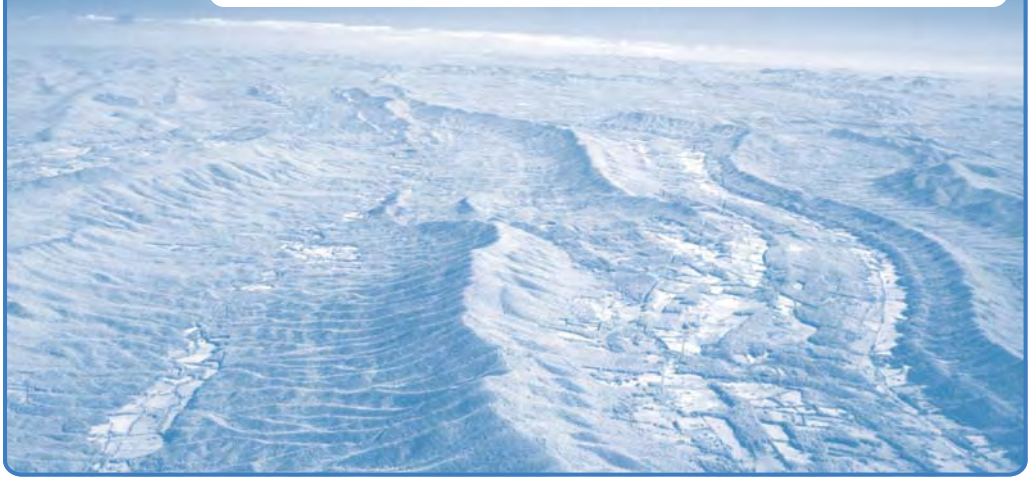
Vogelmann, J. E. 1988. Detection of Forest Change in the Green Mountains of Vermont Using Multispectral Scanner Data. *International Journal of Remote Sensing,* Vol. 9, pp. 1187–1200.

Zhu, Z., and C. E. Woodcock. 2014. Continuous Change Detection and Classification of Land Cover Using All Available Landsat Data. *Remote Sensing of Environment,* Vol. 144, pp. 152–171.

Zitová, B., and J. Flusser. 2003. Image Registration Methods: A Survey. *Image and Vision Computing,* Vol. 21, pp. 977–1000.

PART IV
APPLICATIONS

16 Plant Science Fundamentals



MAJOR TOPICS TO UNDERSTAND

- Structure of the Leaf
- Spectral Behavior of the Living Leaf
- Vegetation Indices
- Applications of Vegetation Indices
- Phenology
- Land Surface Phenology
- Foliar Chemistry

16.1 INTRODUCTION

The Earth's vegetative cover is often the first surface encountered by the energy we use for remote sensing. So, for much of the Earth's land area, remote sensing imagery records chiefly the character of the vegetation at the surface. Therefore, our ability to interpret the Earth's vegetation canopy forms the key to knowledge of other distributions, such as geologic and pedologic patterns that are not directly visible but do manifest themselves indirectly through variations in the character and distribution of the vegetation cover.

In other situations, we have a direct interest in the vegetation itself. Remote sensing can be useful for monitoring areas planted to specific crops, for detecting plant diseases and insect infestations, and for contributing to accurate crop production forecasts. In addition, remote sensing has been used to map forests, including assessments of timber volume, insect infestation, and site quality.

Furthermore, remote sensing provides the only practical means of mapping and monitoring changes in major ecological regions that, although not directly used for production of food or fiber, have great long-term significance for humankind. For example, the tropical forests that cover significant areas of the Earth's surface have never been mapped or studied except in local regions that are unlikely to be representative of the unstudied regions. Yet these regions are of critical importance to humankind due to their role in maintaining the Earth's climate (Rouse et al., 1979) and as a source of genetic diversity. Humans are rapidly destroying large areas of tropical forests; it is only by means of remote sensing that we are ever likely to understand the nature and locations of these changes. Similar issues exist with respect to other ecological zones; remote sensing provides a means to observe such regions at global scales and to better understand the interrelationships among the many factors that influence such patterns.

Vegetation Classification and Mapping

Vegetation classification can proceed along any of several alternative avenues. The most fundamental approach is simply to separate vegetated from nonvegetated regions or forested from open lands. Such distinctions, though ostensibly very simple, can have great significance in some contexts, especially when data are aggregated over large areas or are observed over long intervals of time. Thus, national or state governments, for example, may have an interest in knowing how much of their territories are covered by forest or may want to make changes in forested land from one 10-year period to the next, even though there may be no data available regarding the different kinds of forest.

However, it is usually important to acquire information at finer levels of detail. Although individual plants can be identified on aerial photographs, seldom if ever is it practical to use the individual plant as the unit for mapping vegetation. Instead, it is more useful to define mapping units that represent groups of plants. A *plant community* is an aggregation of plants with mutual interrelationships among each other and with the environment. Thus, an *oak-hickory forest* is a useful designation because we know that communities are not formed by random collections of plants, but by consistent associations of the same groups of plants—plants that tend to prefer the same environmental conditions and to create the environments that permit certain other plants to exist nearby. A community consists of many stands—specific, individual occurrences of a given community (**Figure 16.1**).

Plants within communities do not occur in equal proportions. Certain species tend to dominate; these species are often used to name communities (e.g., hickory forest), although others may be present. Dominant species may also dominate physically, forming the largest plants in a sequence of layers, or *strata,* that are present in virtually all communities. *Stratification* is the tendency of communities to be organized vertically, with some species forming an upper canopy, others a middle stratum, and then shrubs, mosses, lichens, and so on, forming other layers nearer to the ground. Even ostensibly simple vegetation communities in grasslands or arctic tundra, for example, can be shown

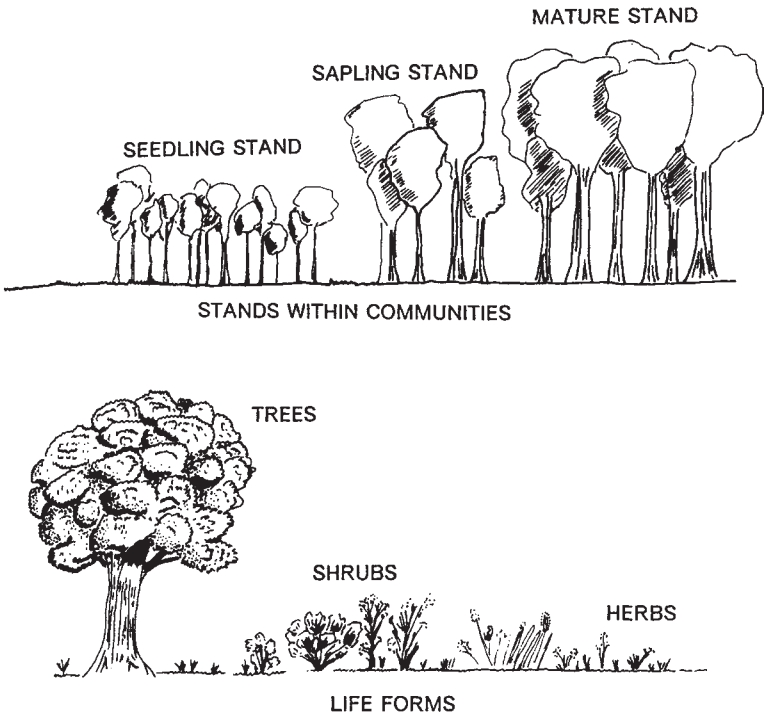


FIGURE 16.1 Vegetation communities, stands, vertical stratification, and different life forms.

to consist of distinct strata. Specific plants within each community tend to favor distinctive positions within each stratum because they have evolved to grow best under the conditions of light, temperature, wind, and humidity that prevail in their strata.

Floristics refers to the botanical classification of plants, which is usually based on the character of the reproductive organs and uses the system founded by Carolus Linnaeus (1707–1778). Linnaeus, a Swedish botanist, created the basis for the binomial system of designating plants by Latin or Latinized names that specify a hierarchical nomenclature, of which the *genus* and *species* (Table 16.1) are most frequently used. The *Linnaean sys-*

TABLE 16.1 Floristic Classification

Level	Example[a]
Class	Angiospermae (broad/general)
Order	Sapindales
Family	Aceraccae
Genus	*Acer*
Species	*saccharum* (narrow/specific)

[a]Example = sugar maple (*Acer saccharum*).

tem provides a distinctive name that places each plant in relationship with others in the taxonomy. Floristic classification reveals the genetic character and evolutionary origin of individual plants.

In contrast, the *life form* or *physiognomy* describes the physical form of plants (Table 16.2). For example, common physiognomies might include "tree," "shrub," "herbaceous vegetation," and so on. Physiognomy is important because it reveals the *ecological role* of the plant—the nature of its relationship with the environment and other plants. Floristics and physiognomy often have little direct relationship to one another. Plants that are quite close floristically may have little similarity in their growth form, and plants that are quite similar in their ecological roles may be very different floristically. For example, the rose family (Rosaceae) includes a wide variety of trees, shrubs, and herbaceous plants that occupy diverse environments and ecological settings. Conversely, a single environment can be home to hundreds of different species. For example, the alpine meadows of New England are composed of some 250 species, including such diverse families as Primulaceae (primrose), Labiatae (mint), Araliaceae (ginseng), and Umbelliferae (parsley). Thus, a purely floristic description of plants (e.g., "the rose family") seldom conveys the kind of information we would like to know about plants, their relationships with one another, and their environment.

Although it is often possible to identify specific plants and to assign taxonomic designations from large-scale imagery, vegetation studies founded on remotely sensed images typically employ the structure and physiognomy of vegetation for classification purposes. That is, usually we wish to separate forest from grassland, for example, or distinguish between various classes of forest. Although it is important to identify the dominant species for each class, our focus is usually on separation of vegetation communities based on their overall form and structure rather than on floristics alone.

No single approach to vegetation classification can be said to be universally superior to others. At given levels of detail, and for specific purposes, each of the approaches mentioned above serves important functions. Floristic classification is useful when scale is large and mapping is possible in fine detail that permits identification of specific plants. For example, analyses for forest management often require a large scale, both for mea-

TABLE 16.2 Classification by Physiognomy and Structure

	Woody plants	Broadleaf evergreen Broadleaf deciduous Needleleaf evergreen Leaves absent Mixed Semideciduous
	Herbaceous plants	Graminoids Forbs Lichens and mosses
	Special life forms	Climbers Stem succulents Tuft plants Bamboos Epiphytes

Note: Data from Küchler (1967). Küchler's complete classification specifies plant height, leaf characteristics, and plant coverage.

surement of timber volume and for identification of individual trees. Physiognomy or structure is important whenever image scale is smaller, detail is coarser, and the analyst focuses more clearly on the relationships of plants to environment. Ecological classification may be used at several scales for analyses that require consideration of broader aspects of planning for resource policy, wildlife management, or inventory of biological resources.

Another approach to classification of vegetation considers vegetation as the most easily observed component of an environmental complex, including vegetation, soil, climate, and topography. This approach classifies regions as ecological zones, usually in a hierarchical system comparable to that shown in Table 16.3. Bailey (1998, p. 145) defines his units as *ecoregions*: "major ecosystem[s], resulting from large-scale predictable patterns of solar irradiation and moisture, which in turn affect the kinds of local ecosystems and animals and plants found there." At the very broadest scales, ecological classification is based on long-term climate and very broad-scale vegetation patterns, traditionally derived from information other than remotely sensed data. However, later sections of this chapter will show how it is now possible to use remotely sensed data to derive these classifications with much more precision and accuracy than was previously possible.

At finer levels of detail, remotely sensed imagery is essential for delineating ecoregions. The interpreter considers not only vegetation cover, but also elevation, slope, aspect, and other topographic factors in defining units on the map.

Biophysical Measures of Vegetation Cover

A contrasting approach applies multispectral remote sensing to estimate basic biophysical properties of the Earth's vegetation cover. That is, instead of labeling a region to a particular class or category as explained above, this approach attempts to provide specific measures of the biophysical function of that region. Three common biophysical measures follow.

Vegetation fraction (VF) is defined as the percentage of vegetation occupying a pixel as viewed in vertical projection. It is a comprehensive quantitative index in forest man-

TABLE 16.3 Bailey's Ecosystem Classification

Level	Example
Domain	Humid temperate
Division	Hot continental
Province	Eastern deciduous forest
Section	Mixed mesophytic forest
District	
Land-type association	
Land type	
Land-type phase	
Site	

Data from Bailey (1978).

agement and vegetation community cover conditions, and it is also an important parameter in many remote sensing ecological models.

Leaf area index (LAI) is the ratio between the total area of the upper leaf surface of vegetation and the surface area of the pixel in question. LAI is a dimensionless value, typically ranging between 0 for a pixel composed of bare soil and values as high as 6 for a dense forest. See **Figure 16.2** for an application that applies remote sensing imagery to estimate LAI for the continental United States.

Net primary production (NPP) is a measure of the inherent productivity of a region or ecological system—mainly the Earth's production of organic matter, principally through the process of photosynthesis in plants. The organisms responsible for primary production, *primary producers,* form the base of the food chain. Whereas *gross primary productivity* (GPP) estimates the total productivity of a region, NPP is the value that remains after allowing for the amount of organic material used to sustain the producers. Because of the basic significance of NPP to understanding agricultural systems and ecosystems in general, the use of remote sensing is of interest as a tool to measure place-to-place and seasonal variations in NPP (Ruimy, Saugier, and Dedieu, 1994).

Kinds of Aerial Imagery for Vegetation Studies

Aerial imagery enables the analyst to conduct quick and accurate delineation of major vegetation units, providing at least preliminary identification of their nature and composition. Note, however, that interpretation from aerial images cannot replace ground

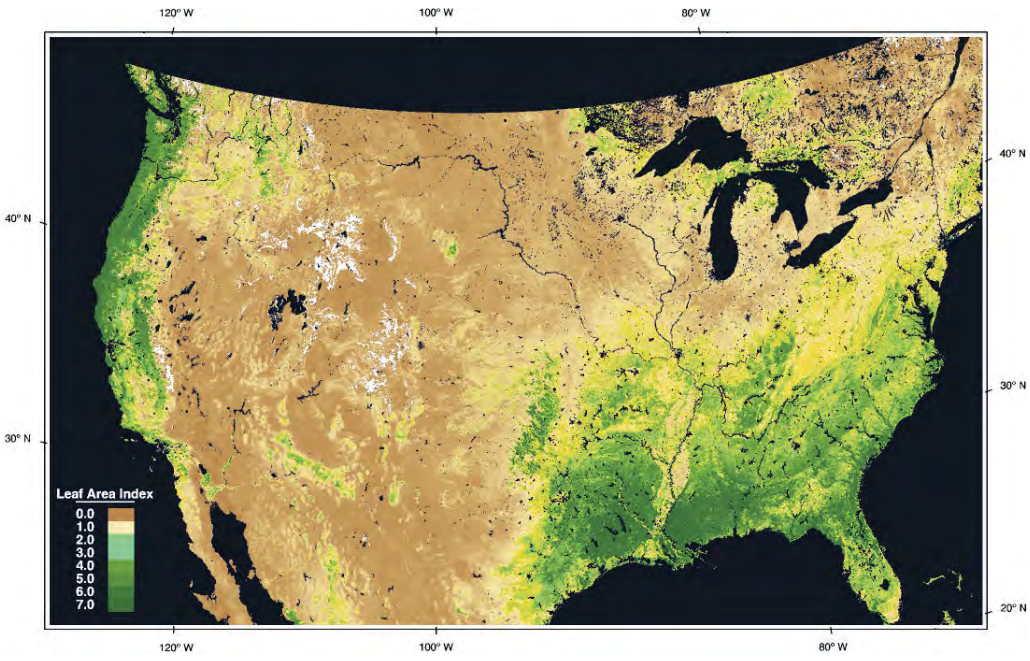


FIGURE 16.2 Moderate Resolution Imaging Spectroradiometer (MODIS) leaf area index (LAI), March 24–April 8, 2000.

observations. An accurate interpretation assumes that the analyst has field experience and knowledge of the area to be examined and will be able to evaluate her or his interpretation in the field.

Stereoscopic coverage provides advantages for interpretation of vegetation coverage and for study of vertical *canopy structure,* the three-dimensional spatial arrangement of a plant canopy. Optimum choice of image scale, if the analyst has control over such variables, depends on the nature of the map and the complexity of the vegetation pattern. Detailed studies have been conducted using photography at scales as large as 1:5,000, but scales from about 1:15,000 to 1:24,000 are probably more typical for general-purpose vegetation studies. Of course, if photographs at several different scales, dates, or seasons are available, multiple coverage can be used to good advantage to study changes. Small-scale images can be used as the basis for delineating the extents of major regions, whereas the greater detail of large-scale images can be used to identify specific plants and plant associations.

Lidar data, especially in conjunction with optical data, are often preferred to stereoscopic photographs alone when information on the canopy structure is particularly important to a scientific or management objective. Modern lidar systems, whether they be waveform, discrete return, or photon-counting, typically enable better penetration—and thus representation—of dense canopies.

Panoramic photographs have been successfully used for a variety of purposes pertaining to vegetation mapping and forest management. Modern panoramic cameras permit acquisition of high-resolution imagery over very wide regions, such that large areas can be surveyed quickly. Such images have extreme variations in scale and perspective, so they cannot be used for measurements, but their wide areal coverage permits a rapid, inexpensive inventory of wide areas, and identification of areas that might require examination using more detailed imagery.

The timing of flights, which may not always be under the analyst's control, can be a critical factor in some projects. For example, mapping the understory in forested areas can be attempted only in the early spring when shrubs and herbaceous vegetation have already bloomed, but the forest canopy has not fully emerged to obscure the smaller plants from overhead views. Because not all plants bloom at the same time, a succession of carefully timed photographic missions in the spring can sometimes record the sequence in which specific species bloom, thereby permitting mapping of detail that could not be reliably determined by a single image showing all trees in full bloom.

Infrared sensors have obvious advantages for vegetation studies of all kinds, as emphasized previously. Given the availability of visible near-infrared sensors in increasingly less expensive, commercially available, remotely piloted aircraft systems, custom-flown, stereoscopic, color-infrared (CIR) photography is once again becoming commonplace.

16.2 STRUCTURE OF THE LEAF

Many applications of remote sensing to vegetation patterns depend on knowledge of the spectral properties of individual leaves and plants. These properties are best understood by examining leaf structure at a rather fine level of detail.

The cross section of a typical leaf reveals its essential elements (Figure 16.3). The uppermost layer, the *upper epidermis,* consists of specialized cells that fit closely together

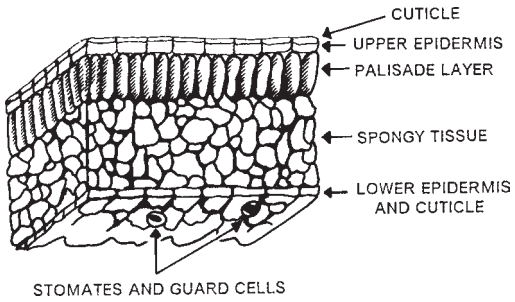


FIGURE 16.3 Diagram of a cross-section of a typical leaf.

without openings or gaps between cells. This upper epidermis is covered by the *cuticle*, a translucent waxy layer that prevents moisture loss from the interior of the leaf. The underside of the leaf is protected by the *lower epidermis*, similar to the upper epidermis except that it includes openings called *stomates* (or *stomata*), which permit movement of air into the interior of the leaf. Each stomate is protected by a pair of *guard cells* that can open and close as necessary to facilitate or prevent movement of air to the interior of the leaf. The primary function of stomata is apparently to allow atmospheric CO_2 to enter the leaf for photosynthesis. Although the guard cells and the epidermis appear to be small and inefficient, they are in fact very effective in transmitting gases from one side of the epidermis to the other. Their role in permitting CO_2 to enter the leaf is obviously essential for the growth of the plant, but they also play a critical role in maintaining the thermal balance of the leaf by permitting movement of moisture to and from the interior of the leaf. The guard cells can close to prevent excessive movement of moisture and thereby economize moisture use by the plant. The positions of stomata on the lower side of the leaf favor maximum transmission of light through the upper epidermis and minimize moisture loss when stomata are open.

On the upper side of the leaf, just below the epidermis, is the *palisade tissue* consisting of vertically elongated cells arranged in parallel, at right angles to the epidermis. Palisade cells include *chloroplasts*—cells composed of chlorophyll and other ("accessory") pigments active in photosynthesis, as described below. Below the palisade tissue is the spongy *mesophyll tissue*, which consists of irregularly shaped cells separated by interconnected openings. The surface of the mesophyll has a very large surface area; it is the site for the O_2 and CO_2 exchange necessary for photosynthesis and respiration. Although leaf structure is not identical for all plants, this description provides a general outline of the major elements common to most plants, especially those that are likely to be important in agricultural and forestry studies.

In the visible portion of the spectrum, chlorophyll controls much of the spectral response of the living leaf (Figure 16.4). *Chlorophyll* is the green pigment that is chiefly responsible for the green color of living vegetation. Chlorophyll enables the plant to absorb sunlight, thereby making photosynthesis possible; it is located in specialized lens-shaped structures, known as *chloroplasts*, found in the palisade layer. Light that passes through the upper tissues of the leaf is received by chlorophyll molecules in the palisade layer, which is apparently specialized for photosynthesis, as it contains the largest chloroplasts, in greater abundance than other portions of the plant. CO_2 (as a component of the natural atmosphere) enters the leaf through stomata on the underside of the leaf and then diffuses throughout cavities within the leaf. Thus, photosynthesis creates car-

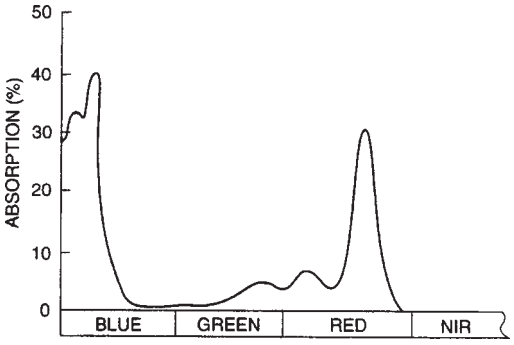


FIGURE 16.4 Chlorophyll *a* absorption spectrum of a typical leaf. This pigment is used in oxygenic photosynthesis and absorbs blue and red radiation.

bohydrates from CO_2 and H_2O, using the ability of chloroplasts to absorb sunlight as a source of energy. Chloroplasts include a variety of pigments, some known as *accessory pigments,* that can absorb light and then pass its energy to chlorophyll. Chlorophyll occurs in two forms. The most common is chlorophyll *a,* the most important photosynthetic agent in most green plants. A second form, known as chlorophyll *b,* has a slightly different molecular structure; it is found in most green leaves, as well as in some algae and bacteria.

16.3 SPECTRAL BEHAVIOR OF THE LIVING LEAF

Chlorophyll does not absorb all sunlight equally. Chlorophyll molecules preferentially absorb blue and red light for use in photosynthesis (Figure 16.5). They may absorb as much as 70 to 90% of incident light in these regions. Less of the green light is absorbed and more is reflected, so the human observer, who can see only the visible spectrum, sees the dominant reflection of green light as the color of living vegetation (Figures 16.5 and 16.6).

Chlorophyll emits light (in addition to heat) as photons are absorbed, called *fluorescence*. This subtle emitted signal, with peaks at approximately 690 and 740 nm (Khosravi et al., 2015), is difficult to observe because it is mixed in with a large amount of reflected light (comprising less than 3% of this total in the near infrared [NIR]; Moya et al., 2004). However, there are dark features in the observed solar spectrum (Chapter

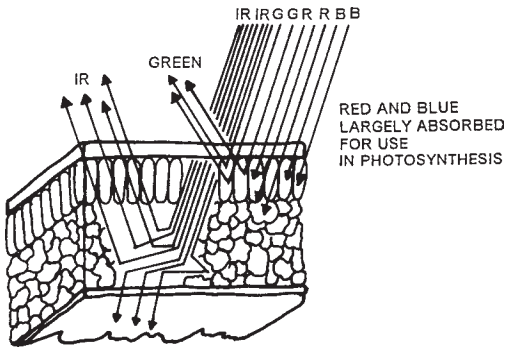


FIGURE 16.5 Interaction of leaf structure with visible and infrared radiation.

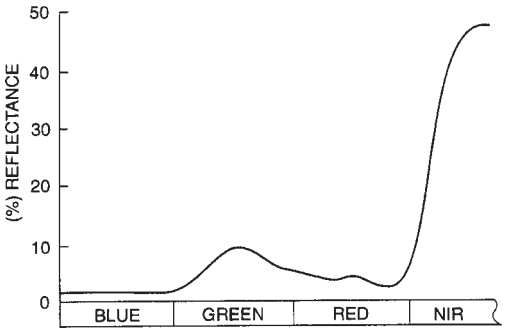


FIGURE 16.6 Typical spectral reflectance from a living leaf. The leaf is brightest in the near-infrared region, although in the visible region (blue through red) the maximum is in the green region. Note that most of even the green light is absorbed.

14), *Fraunhofer lines*, caused by absorption by atmospheric constituents. Two of these lines, labeled *A* (760 nm), and *B* (687 nm), are among the four principal Fraunhofer lines caused by molecular oxygen absorption. These two lines overlap the chlorophyll fluorescence emission spectrum. In these two Fraunhofer lines, the additive signal from fluorescence is relatively larger, and methods have been developed (Plascyk and Gabriel, 1975, and successive refinements in other studies) enabling fluorescence quantification.

In the near-infrared spectrum, reflection of the leaf is controlled not by plant pigments but by the structure of the spongy mesophyll tissue. The cuticle and epidermis are almost completely transparent to infrared radiation, so very little infrared radiation is reflected from the outer portion of the leaf. Radiation passing through the upper epidermis is strongly scattered by mesophyll tissue and cavities within the leaf. Little of this infrared energy is absorbed internally—most (up to 60%) is scattered upward (which we call "reflected energy") or downward ("transmitted energy"). Some studies suggest that palisade tissue may also be important in infrared reflectance. Thus, the internal structure of the leaf is responsible for the bright infrared reflectance of living vegetation.

At the edge of the visible spectrum, as the absorption of red light by chlorophyll pigments begins to decline, reflectance rises sharply (Figure 16.7). Thus, if reflectance

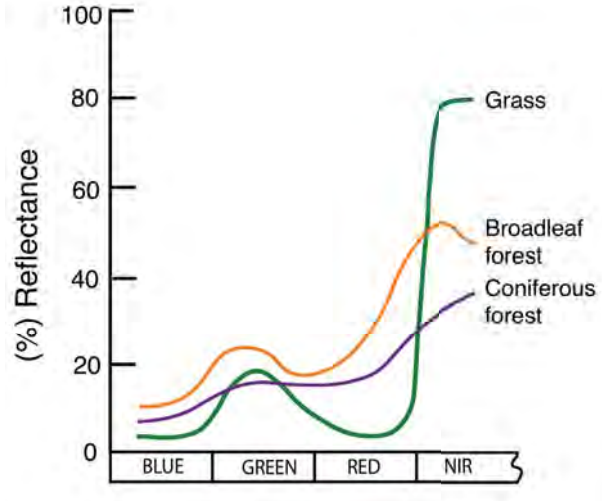


FIGURE 16.7 Vegetation spectral signatures in the visible and near infrared wavelengths. Differences between vegetation classes are often more distinct in the near-infrared than in the visible. Image by Susmita Sen.

is considered not only in the visible but across the visible and the near infrared, peak reflectance of living vegetation is not in the green but in the near infrared. This behavior explains the great utility of the near-infrared spectrum for vegetation studies and facilitates separation of vegetated from nonvegetated surfaces, which are usually much darker in the near infrared. Furthermore, differences in the spectral signatures of plant species often are more pronounced in the near infrared than they are in the visible, meaning that discrimination of vegetation classes is sometimes made possible by use of the near-infrared reflectance (Figure 16.7).

As a plant matures or is subjected to stress by disease, insect attack, or moisture shortage, the spectral characteristics of the leaf may change. In general, these changes apparently occur more or less simultaneously in both the visible and the near-infrared regions, but changes in near-infrared reflectance are often more noticeable. Reflectance in the near-infrared region is apparently controlled by the nature of the complex cavities within the leaf and internal reflection of infrared radiation within these cavities. Although some scientists suggest that moisture stress or natural maturity of a leaf causes these cavities to "collapse" as a plant wilts, others maintain that it is more likely that decreases in near-infrared reflection are caused by deterioration of cell walls rather than physical changes in the cavities themselves. Thus, changes in infrared reflectance can reveal changes in vegetative vigor; infrared images have been valuable in detecting and mapping the presence, distribution, and spread of crop diseases and insect infestations. Furthermore, changes in leaf structure that accompany natural maturing of crops are subject to detection with infrared imagery, so that it is often possible to monitor the ripening of crops as harvest time approaches. CIR sensors are valuable for observing such changes because of their ability to show spectral changes in both the visible and near-infrared regions and to provide clear images that show subtle tonal differences.

In the shortwave (or "mid-") infrared wavelengths (beyond 1.3 μm), leaf water content appears to control the spectral properties of the leaf (Figure 16.8). The term *equivalent water thickness* (EWT) has been proposed to designate the thickness of a film of water that can account for the absorption spectrum of a leaf within the 1.4–2.5 μm interval (Allen et al., 1971). Within the shortwave infrared region (SWIR), Tucker (1980) found the 1.55–1.75 μm (SWIR1) best suited to monitor canopy water status from space platforms. However, since both internal structure and dry matter influence SWIR reflectance across species (Ceccato et al., 2001), interspecific EWT-reflectance relationships require both the NIR (structure, dry matter) and the SWIR.

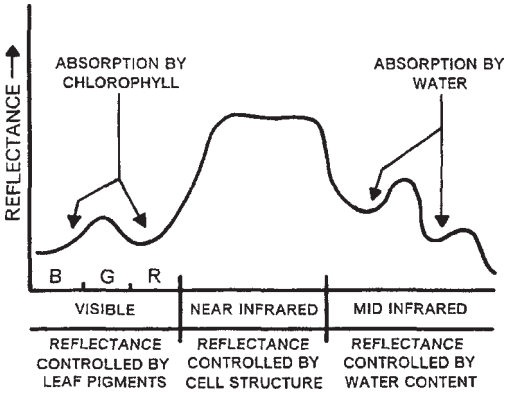


FIGURE 16.8 Reflectance differences between equivalent water thicknesses of 0.018 cm and 0.014 cm. Changes in leaf water content may be pronounced in the mid-infrared region. Diagram based on simulated data reported by Tucker (1979, p. 10).

As such, indices have been developed using both the NIR and SWIR that have been shown to be related to leaf or canopy water content using wavelength bands available on well-calibrated (Landsat, Sentinel, SPOT) moderate-resolution Earth resource satellites. The *moisture stress index* (MSI; Hunt and Rock, 1989), shown as Equation 16.1, is linearly related to leaf water content in a species-specific fashion. More important, however, MSI is linearly related to $log_{10}EWT$. Since EWT and LAI are correlated, MSI is also related to LAI.

$$\text{MSI} = \frac{\text{SWIR1}}{\text{NIR}} \qquad \text{(EQ. 16.1)}$$

The *normalized difference moisture index* (NDMI; Hardisky, Klemas, and Smart, 1983; Jin and Sader, 2005), shown as Equation 16.2, is highly correlated with canopy water content and, by extension, to changes in the canopy that affect the water content.

$$\text{NDMI} = \frac{\text{NIR} - \text{SWIR1}}{\text{NIR} + \text{SWIR1}} \qquad \text{(EQ. 16.2)}$$

Reflection from Canopies

Knowledge of the spectral behavior of individual leaves is important for understanding the spectral characteristics of vegetation canopies but cannot itself completely explain reflectance from areas of complete vegetative cover. Vegetation canopies are composed of many separate leaves that may vary in their size, orientation, shape, and coverage of the ground surface. In the field, a vegetation canopy (e.g., in a forest or a cornfield) is composed of many layers of leaves; the upper leaves form shadows that mask the lower leaves, creating an overall reflectance created by a combination of leaf reflectance and shadow.

Shadowing tends to decrease canopy reflectance below the values normally observed in the laboratory for individual leaves. Knipling (1970) cited percentages reported in several previously published studies (shown in Table 16.4).

Thus, the reflectance of a canopy is considerably lower than reflectances measured for individual leaves. But the relative decrease in the near-infrared region is much lower than that in the visible. The brighter canopy reflection for the near infrared is attributed to the fact that plant leaves transmit NIR radiation, perhaps as much as 50–60%. Therefore, infrared radiation is transmitted through the upper layers of the canopy, reflected in part from lower leaves, and retransmitted back through the upper leaves, resulting in bright infrared reflectance (Figure 16.9). Physicists and botanists have been able to develop mathematical models for canopy reflectances by estimating the optical properties of the leaves and the canopy as a whole. Because some of the transmitted infrared radia-

TABLE 16.4 Reflectance in Canopies versus Individual Leaves

	Visible	Near infrared
Single-leaf reflectance (%)	10	50
Canopy reflectance (%)	3–5	35

Source: Based on Knipling (1970).

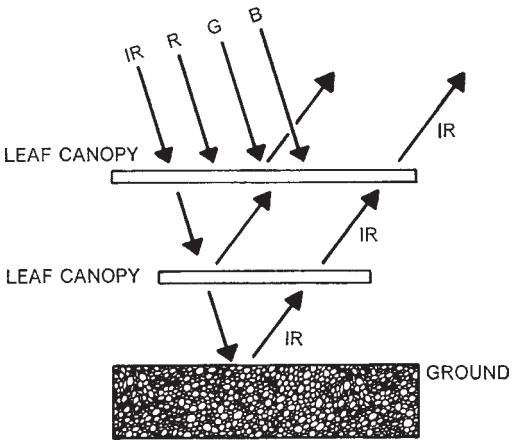


FIGURE 16.9 Simplified cross-sectional view of behavior of energy interacting with a vegetation canopy. (See **Figures 16.3 and 16.4.**) In some portions of the spectrum, energy transmitted through the upper layer is available for reflection from lower layers (or from the soil surface).

tion may in fact be reflected from the soil surface below the canopy, such models have formed an important part of the research in agricultural remote sensing.

The Red Shift

Collins (1978) reports the results of studies that show changes in the spectral responses of crops as they approach maturity. His research used high-resolution multispectral scanner (MSS) data of numerous crops at various stages of the growth cycle. Collins's research focused on examination of the far-red region of the spectrum, where chlorophyll absorption decreases and infrared reflection increases (Figure 16.9). In this zone, the spectral response of living vegetation increases sharply as wavelength increases: in the region from just below 0.7 μm to just above 0.7 μm, brightness increases by about 10 times (Figure 16.10).

Collins observed that as crops approach maturity, the position of the chlorophyll absorption edge shifts toward longer wavelengths, a change he refers to as the "red shift" (Figure 16.10; also known as the *red edge shift*). The red edge shift is observed not only in crops but also in other plants. The magnitude of the red edge shift varies with crop type (it is a pronounced and persistent feature in wheat).

Collins observed the red edge shift along the entire length of the chlorophyll absorption edge, although it was most pronounced near 0.74 m, in the infrared region, near the shoulder of the absorption edge. He suggested that very narrow bands at about 0.745 μm and 0.780 μm would permit observation of the red edge shift over time and thereby provide a means of assessing differences between crops and the onset of maturity of a specific crop. Many modern multispectral sensors (RapidEye, Sentinel-2) have added narrow wavelength bands enabling detection of subtle changes in the red edge.

Causes of the red edge shift appear to be very complex. Chlorophyll *a* appears to increase in abundance as the plant matures; increased concentrations change the molecular form in a manner that adds absorption bands to the edge of the chlorophyll *a* absorption region, thereby producing the red shift. Biochemical concentrations, particularly canopy nitrogen (Mutanga and Skidmore, 2007) also shift the red edge. (Certain factors can alter the spectral effect of chlorophyll, thereby shifting the edge of the absorption band toward shorter wavelengths; this is the "blue shift" observed in geobotanical studies.)

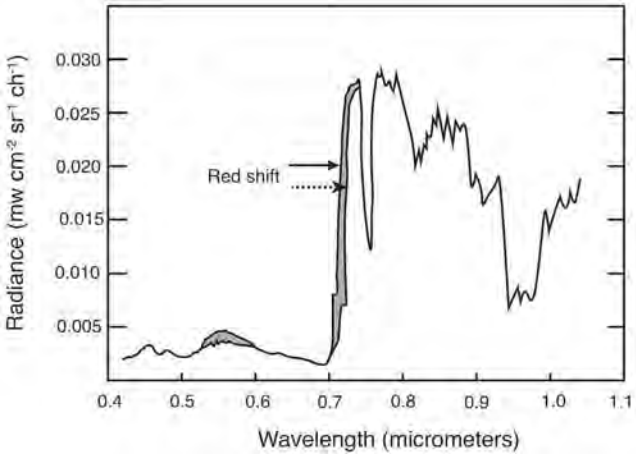


FIGURE 16.10 Red shift. The absorption edge of chlorophyll shifts toward longer wavelengths as plants mature. The shaded area represents the magnitude of this shift as the difference between the spectral response of headed wheat and mature alfalfa. Similar but smaller shifts have been observed in other plants. The red shift is important for distinguishing the headed stage of a crop from earlier stages of the same crop and for distinguishing headed grain crops from other green, nongrain crops. From Collins (1978, p. 47). Copyright © 1978 by the American Society of Photogrammetry and Remote Sensing. Used by permission.

16.4 VEGETATION INDICES

Vegetation indices (VIs), based on digital brightness values, attempt to measure biomass or vegetative vigor. A VI is formed from combinations of several spectral values that are added, divided, and/or multiplied in a manner designed to yield a single value that indicates the amount or vigor of vegetation within a pixel. High values of the VI identify pixels covered by substantial proportions of healthy vegetation. The simplest form of VI is a ratio between two digital values from separate spectral bands. Some band ratios have been defined by applying knowledge of the spectral behavior of living vegetation.

Band ratios are quotients between measurements of reflectance in separate portions of the spectrum. Ratios are effective in enhancing or revealing latent information when there is an inverse relationship between two spectral responses to the same biophysical phenomenon. If two features have the same spectral behavior, ratios provide little additional information. If they have quite different spectral responses, however, the ratio between the two values provides a single value that concisely expresses the contrast between the two reflectances.

For living vegetation, the ratioing strategy can be especially effective because of the inverse relationship between vegetation brightness in the red and infrared regions. That is, absorption of red light (R) by chlorophyll and strong reflection of NIR radiation by mesophyll tissue ensure that the red and near-infrared values will be quite different and that the simple ratio (SR; Equation 16.3) of actively growing plants will be high. Nonvegetated surfaces, including open water, human-made features, bare soil, and dead or stressed vegetation will not display this specific spectral response, and the ratios will decrease in magnitude. Thus, the SR can provide a measure of photosynthetic activity and biomass within a pixel.

$$SR = \frac{NIR}{R} \quad \text{(EQ. 16.3)}$$

The SR is only one of many measures of vegetation vigor and abundance. One of the most widely used VIs is known as the normalized difference vegetation index (NDVI; Equation 16.4):

$$NDVI = \frac{NIR - R}{NIR + R} \quad \text{(EQ. 16.4)}$$

This index in principle conveys the same kind of information as the SR, but it is constrained to vary within limits that preserve desirable statistical properties in the resulting distributions. Tucker (1980) and Perry and Lautenschlager (1984) suggest that in practice few differences among the many VIs have been proposed.

16.5 APPLICATIONS OF VEGETATION INDICES

Many of the first studies examining applications of VIs attempted to "validate" their usefulness by establishing that values of the VIs are closely related to the biological characteristics of plants. For example, such studies might examine test plots during an entire growing season and then compare values of the VIs, measured throughout the growing season, to samples of vegetation taken at the same times. The objective of such studies is ultimately to establish use of VIs as a means of remote monitoring of the growth and productivity of specific crops, or of seasonal and yearly fluctuations in productivity. Often values of the VIs have been compared to *in situ* measurements of LAI or to above-ground biomass (the weight of vegetative tissue, excluding roots). A number of VIs appear to be closely related to LAI (at least for specific crops), but no single VI seems to be equally effective for all plants and all agricultural conditions. The results of such studies have in general confirmed the utility of the quantitative uses of VIs, but details vary with the specific crop considered, atmospheric conditions, and local agricultural practices. A second category of applications uses VIs as a mapping device—that is, much more as a qualitative rather than a quantitative tool. Such applications use VIs to assist in image classification, to separate vegetated from nonvegetated areas, to distinguish between different types and densities of vegetation, and to monitor seasonal variations in vegetative vigor, abundance, and distribution (Figure 16.11).

Although such studies have established the utility of VIs for practical applications, there are many precautions for uses of VIs that are to be interpreted quantitatively rather than qualitatively. Values of VIs can be influenced by many factors external to the plant leaf, including viewing angle, soil background, and differences in row direction and spacing in the case of agricultural crops. Ratios may be sensitive to atmospheric degradation because atmospheric effects have greater impact on bands at shorter wavelengths than those at longer wavelengths. Atmospheric effects can alter the ratios from their true values (Figure 16.12). Because atmospheric path length varies with viewing angle, values calculated using off-nadir satellite data (Chapter 5) vary according to position within the image. Clevers and Verhoef (1993) found that the relationship between LAI and the VI they studied was very sensitive to differences in leaf orientation. Although preprocessing can sometimes address such problems, it may still be difficult to compare values of VIs

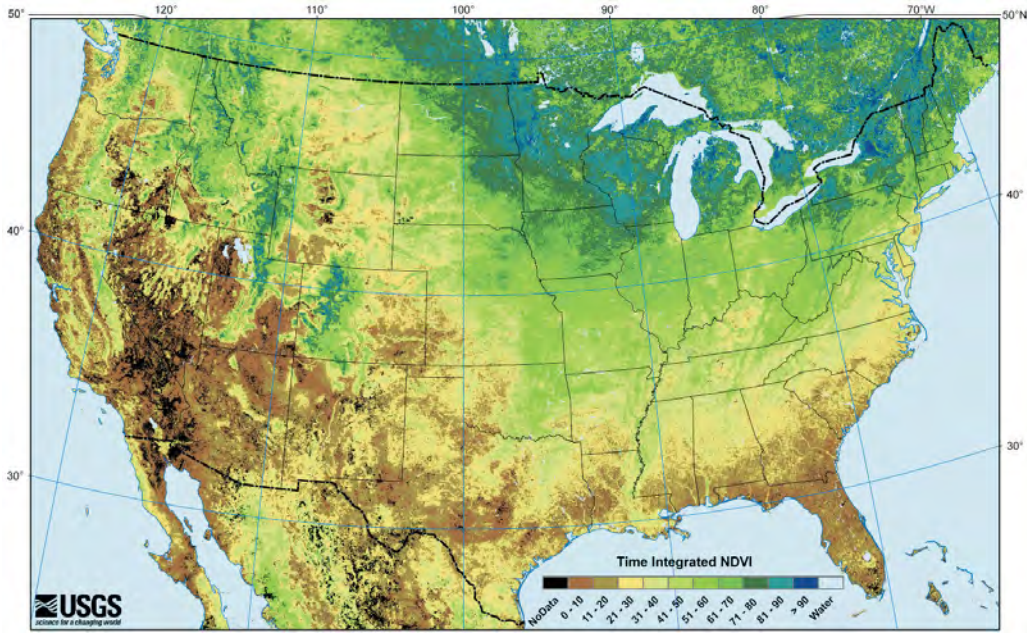


FIGURE 16.11 Time-integrated NDVI (TIN) for 2008. TIN estimates a daily (interpolated) NDVI above a baseline summed for the entire duration of the growing season and interpreted as growing season canopy photosynthetic activity (related to primary productivity). Darker blues and greens represent the higher values, whereas beiges and browns represent the lowest values. This figure shows a pattern related to vegetation cover, moisture, and insolation. From U.S. Geological Survey (USGS), Earth Resources Observation and Science (EROS) Center.

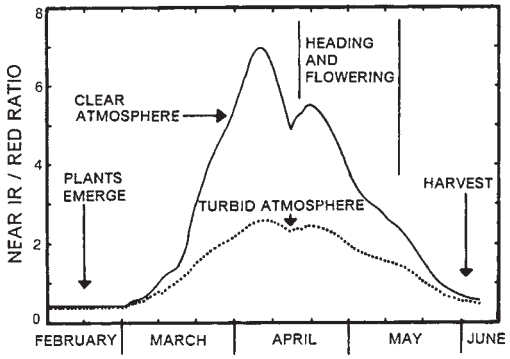


FIGURE 16.12 Influence of atmospheric turbidity on near infrared/red ratio, or simple ratio, as estimated from simulated data. From Jackson et al. (1983, p. 195). Copyright © 1983 Elsevier Science Publishing Co., Inc. Used by permission.

over a period of time because of variation in external factors. Price (1987) and others have noted that efforts to compare ratios or indices over time, or from one sensor to another, should reduce digital values to radiances before calculating ratios, thereby accounting for differences in calibration of sensors. Turner et al. (1999) showed the importance of correcting for atmospheric effects as well (converting to surface reflectance), which, when feasible, should always be done prior to computing a VI.

NDVI has been validated as a measure effective in separating pixels with active vegetation growth from those that are unvegetated, or characterized by senescent vegetation. However, it may not be appropriate for some applications in which the NDVI values are to be interpreted as continuous measures of biophysical parameters. Myneni et al. (1995, 2002) and Huete et al. (2002), for example, have shown that NDVI exhibits nonlinear relationships with biophysical measures such as VF or LAI, for example. Values of NDVI approach saturation (their maximum) before the biomass of the observed target reaches a maximum; that is, NDVI values reach a maximum and do not increase despite continued increases in LAI, biomass, or similar measures. Researchers have proposed modified indices or alternative analytical approaches to compensate for such effects.

16.6 PHENOLOGY

Phenology is "the study of the ways in which the timing and other aspects of periodic events . . . are affected by climate and other environmental factors" (Hine and Martin, 2016). Often phenology refers specifically to seasonal changes in vegetative growth and decline, but other seasonal phenomena, such as lake ice, are also extensively studied (Wynne et al., 1998). Many phenological changes can be monitored by means of remote sensing because plants change in appearance and structure during their growth cycle. Of special significance are spectral and physiological changes that occur as a plant matures. Each season, plants experience chemical, physical, and biological changes, known as senescence, that result in progressive deterioration of leaves, stems, fruit, and flowers. In the midlatitudes, annual plants typically lose most or all of their roots, leaves, and stems each year. Woody perennial plants typically retain some or all of their roots, woody stems, and branches but shed leaves. Evergreens, including tropical plants, experience much more elaborate phenological cycles. Individual leaves may experience senescence separately (i.e., trees do not necessarily shed all leaves simultaneously), and individual trees or branches of trees may shed leaves on cycles quite distinct from others in the same forest (Koriba, 1958).

During the onset of senescence, deterioration of cell walls in the mesophyll tissue produces a distinctive decline in infrared reflectance; an accompanying increase in visible brightness may be the result of decline in the abundance and effectiveness of chlorophyll as an absorber of visible radiation. Changes in chlorophyll produce the red shift mentioned above. Such changes can be observed spectrally, so remote sensing imagery can be an effective means of monitoring seasonal changes in vegetation.

The phenology of a specific plant defines its seasonal pattern of growth, flowering, senescence, and dormancy. Remotely sensed images can expand the scope of study to include overviews of vegetation communities or even of entire biomes. (Biomes are broad-scale vegetation regions that correspond roughly to the Earth's major climate regions.) Dethier et al. (1973), for example, used several forms of imagery and data to observe the geographic spread of the emergence of new growth in spring, as it progressed from

south to north in North America. This phenomenon has been referred to as the green wave. Then in late summer, the brown wave sweeps across the continent as plant tissues mature, dry, and are harvested.

Figure 16.13 illustrates local phenological differences in the spread of the green wave in the early spring. The image shows Landsat MSS band 4 (NIR) radiation, which is sensitive to variations in the density, type, and vigor of living vegetation. The region represented by Figure 16.13 has uneven topography covered by a mixture of forested and open land, including large regions of cropland and pasture. When this image was acquired in mid-April, vegetation in the open land was just starting to emerge. Grasses and shrubs in these regions have bright green leaves, but leaves on the larger shrubs and trees have not yet emerged. Therefore, the pattern of white depicts the regions occupied primarily by early blooming grasses and shrubs in lower elevations. Within a week or so after this image was acquired, leaves on trees began to emerge, first at the lower elevations and later at higher elevations. Thus, a second image acquired only 2 weeks or so after the first would appear almost completely white, due to the infrared brightness of the almost complete vegetation cover. If it were possible to observe this region on a daily basis, under cloud-free conditions, we could monitor the movement of the green wave upward from lower to higher elevations and from south to north, as springtime temperatures prevail over more and more of the region. In reality, of course, we can see only occasional snapshots of the movement of the green wave, as the infrequent passes of Landsat and cloud cover prevent close observation during the short period when we must watch its prog-

FIGURE 16.13 Landsat MSS band 4 image of southwestern Virginia. Data from USGS, EROS Center.

ress. *Virtual constellations* (in which data from two or more sensors are combined into a single data set, *sensu* Claverie et al., 2018) improve temporal resolution and therefore enable more subtle characterizations of phenology even at the spatial resolution of moderate resolution Earth resource satellite data.

16.7 LAND SURFACE PHENOLOGY

Traditionally, the study of phenology focused largely on individual plants. However, as the field of remote sensing developed the tools to assess seasonal changes over rather large areas, it became possible to evaluate the phenological behavior at regional, or even continental, scales using metrics to be described below. This broad-scale focus is known as *land surface phenology* (de Beurs and Henebry, 2004; Morisette et al., 2009).

NDVI, as observed by satellite sensors, permits tracking the seasonal rise and decline of photosynthetic activity within large regions, over several growing seasons. Reed and Yang (1997) were among the first to systematically apply broad-scale remote sensing instruments to derive phenology for very large regions. They accumulated cloud-free Advanced Very High Resolution Radiometer (AVHRR) composites of the United States over a 4-year period, calculating NDVI for each pixel. Examination of NDVI over such long periods and such large areas permitted inspection of phenological patterns of varied land cover classes, as well as responses to climatic and meteorological events, including drought, floods, and freezes. They used AVHRR 14-day composites during the March–October growing season and once a month during the winter season. Such products are useful for monitoring agricultural regions, assessing climatic impact on ecosystems, production of food, and cover for wildlife. Agricultural systems can be observed at a broad scale to assess the response of crops to meteorological and climatic variations throughout the growing season, enabling yield forecasts. Subsequent applications of the same approach have used MODIS data (Stöckli and Vidale, 2004) and Landsat data (Fisher, Mustard, and Vadeboncoeur, 2006; Walker, de Beurs, and Wynne, 2012) to provide additional spatial detail.

Some precautions are necessary in making detailed interpretations of such data. NDVI is, of course, subject to atmospheric effects, and the coarse resolution of AVHRR, MODIS, and Visible Infrared Imaging Radiometer Suite data often results in a given pixel encompassing multiple land cover classes with varied phenological responses, especially in regions where landscapes are locally complex and diverse. Atmospheric effects can create noise that contaminates the brightness values, so it is necessary to smooth the observations using a moving average to filter the effects of the atmosphere, snow cover, and the like upon NDVI values (Figure 16.14).

Observed over time, values of the NDVI depict a seasonal pattern of increase and decrease (Figure 16.14). During the winter season, NDVI reaches minimum values as vegetation enters dormancy, although in some regions, residual islands of green vegetation may remain. As springtime approaches, NDVI begins to rise. The onset of longer days, warmer temperatures, and spring rains promotes the budding of vegetation and the emergence of leaves, first with grasses and shrubs, then with trees. The seasonal response in NDVI is known as *green-up*. The rise in NDVI marks the *start of season* (SOS). Eventually, green-up peaks at maximum NDVI. Mature plants invest energy in growing the seeds and fruits necessary to prepare the species for success in the next year's growing season. As the season progresses, green-up slows, stops, and then declines, marking

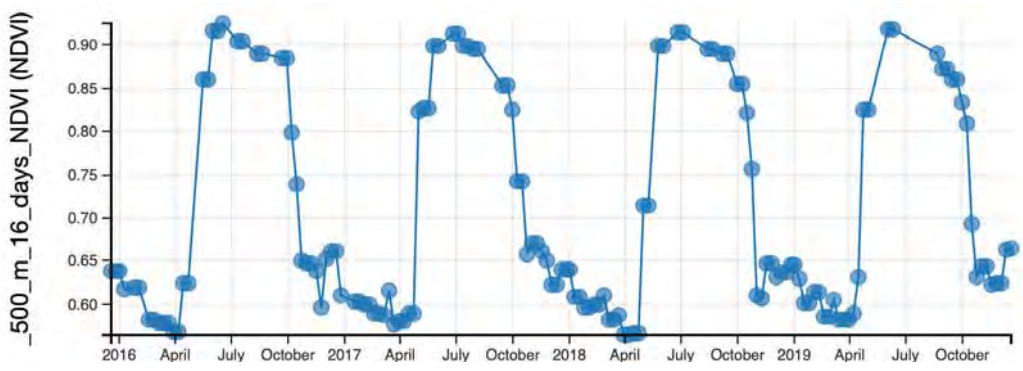


FIGURE 16.14 Seasonal phenological variation in NDVI for a single pixel tracked over the interval 2016–2019. These data are from the VIIRS (Visible Infrared Imaging Radiometer Suite) 500 m 16-day composite NDVI product. The latitude and longitude of the pixel location are 37.27° N, 80.48° W.

the *end of season* (EOS). The decline from mid- to late summer records the decline in vigor as *senescence,* and harvest time (for agricultural crops) approaches. The regional decline in NDVI in the fall is sometimes known as *brown-down*. The interval between SOS and EOS constitutes a region's growing season, which is sometimes measured as the *duration of season* or by integrating the area under the phenological curve as the TIN. Thus, these data can provide a means of assessing place-to-place variation in the length of growing seasons. The result produces a typical seasonal pattern as depicted in Figure 16.15.

Land surface phenology metrics, such as SOS, are important indicators of regional and global climate change, with broad implications for plant productivity. For example, Figure 16.16 shows that for a 20-year period (1982–2011), the SOS in much of Fennoscandia (Finland, Sweden, Norway, and northwest Russia west of the White Sea) is now much earlier.

16.8 FOLIAR CHEMISTRY

The development of operational hyperspectral sensors (Chapter 14) has opened possibilities for application of remote sensing to observation of characteristics of vegetated surfaces, which was not feasible by using the instruments that were previously at hand. Hyperspectral data have created the capability to investigate foliar chemistry, the chemical composition of living leaves. If such a capability could be applied in an operational context, it could help identify plants at levels of detail not previously feasible and monitor the growth and health of crops and forests.

Although a considerable body of research is devoted to remote sensing of foliar chemistry, much of this work is specific to particular sensors, individual test sites, or specific processing algorithms, and has been acquired by varied means (laboratory analyses, field observations, and airborne data). Therefore, much work remains to be completed to develop the suggestions offered by existing studies. The following paragraphs outline some of the efforts to investigate the hyperspectral sensing of foliar chemistry.

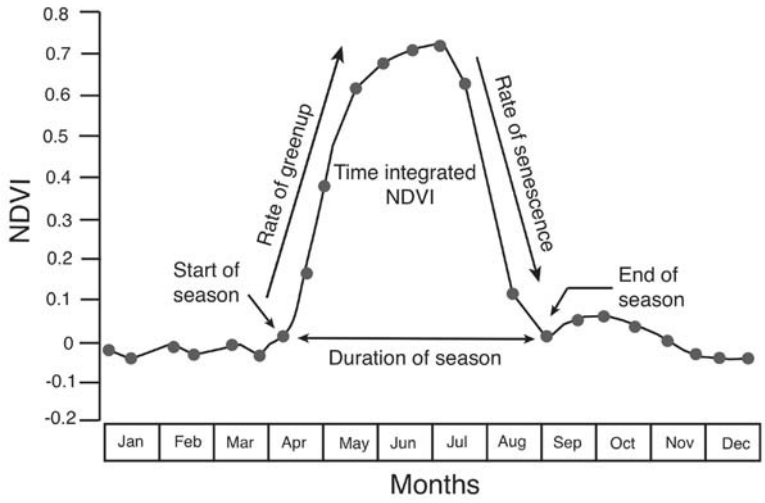


FIGURE 16.15 Idealized phenological diagram for a single season. The vertical axis represents NDVI, whereas the horizontal axis represents time, beginning and ending in the northern hemisphere winter season. Annotations identify key metrics, including start of season (SOS) and end of season (EOS), defined by quantitative assessment of the sequence of individual NDVI measurements throughout the year. From Kirsten de Beurs. Used by permission.

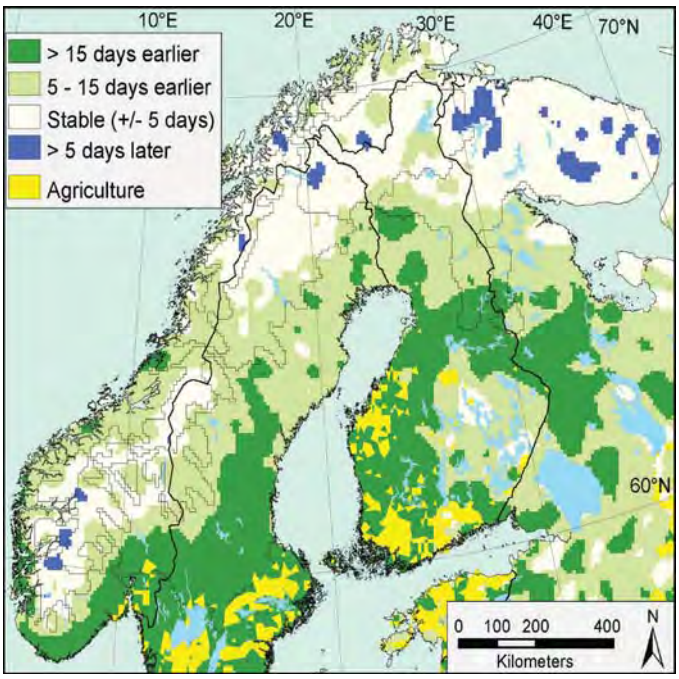


FIGURE 16.16 Number of days of the earlier start of the growing season in Fennoscandia in 2011 compared to 1982, assuming a linear trend. Black line, national borders; gray line, vegetation region borders. From Høgda et al. (2013).

Lignin and Starch Content

Kokaly and Clark (1999) applied a normalization procedure that permitted detection of lignin and cellulose in dried ground leaves in a laboratory setting. Peterson et al. (1988) found that the spectral region in the 1500–1750 nm range in airborne imaging spectrometer (AIS) data was linked to levels of lignin and starch content. Martin and Aber (1997) reported positive results for predicting lignin using four airborne visible infrared imaging spectrometer (AVIRIS) bands in the 1660–2280 nm range, at test sites in the northeastern United States. More recent work, focused on the cellulose-to-lignin ratio important to nutrient cycling rates after litterfall, has highlighted the importance of the 2.1- and 2.3-μm absorption features (Kokaly et al., 2009).

Nitrogen

Card et al. (1988) found promising relationships between spectroradiometer data and the nitrogen content of ground, dried laboratory samples. Wessman, Aber, and Peterson (1989) investigated forest canopy chemistry using AIS imagery; they found strong relationships between foliar nitrogen and reflectance in bands between 1,265 nm and 1,555 nm. Martin and Aber (1997) examined AVIRIS data (20-m resolution) on a mixed broadleaf forest of oak (*Quercus rubra*), maple (*Acer rubrum*), and needle-leaved species. They found a relationship between nitrogen content and brightness in the intervals 750–2140 nm and 950–2290 nm. Pellissier et al. (2015) were able to estimate percent foliar nitrogen in lawn and other cultivated grasslands using airborne imaging spectroscopy data. Root mean square error, expressed as a percentage of the mean, was 16.9%, with the partial least squares variable importances focused in the 750–1300 nm and, to a lesser extent, 1550–1750 nm wavelength ranges.

Chlorophyll Concentration and Fluorescence

Kupiec and Curran (1993) examined AVIRIS data and found that brightness at 723 nm was strongly correlated with chlorophyll content in the needles of slash pine (*Pinus elliotti*). Sims and Gamon (2002), in a comprehensive study of 53 species, found the modified SR at 705 nm (mSR_{705}; Equation 16.5) best correlated with total leaf chlorophyll (mmol m^{-2}). However, variability was induced from both leaf anthocyanin content and leaf surface reflectance.

$$mSR_{705} = \frac{R_{750} - R_{445}}{R_{705} - R_{445}} \quad \text{(EQ. 16.5)}$$

Solar-induced fluorescence (SIF; emitted light from sunlit leaves) has been mapped globally (Figure 16.17), showing photosynthetic activity by both annual totals and growing season maxima. Of particular interest in Figure 16.17 is the time series, reminding us of the continued importance of the humid tropics with respect to global photosynthetic activity. SIF is linearly related to GPP as measured by eddy flux towers in a largely biome-independent fashion (Sun et al., 2017; Figure 16.18), at least when skies are clear (Zhang et al., 2018). As such, SIF is giving the scientific community a new window on global GPP dynamics at a critical time with respect to anthropogenic influences.

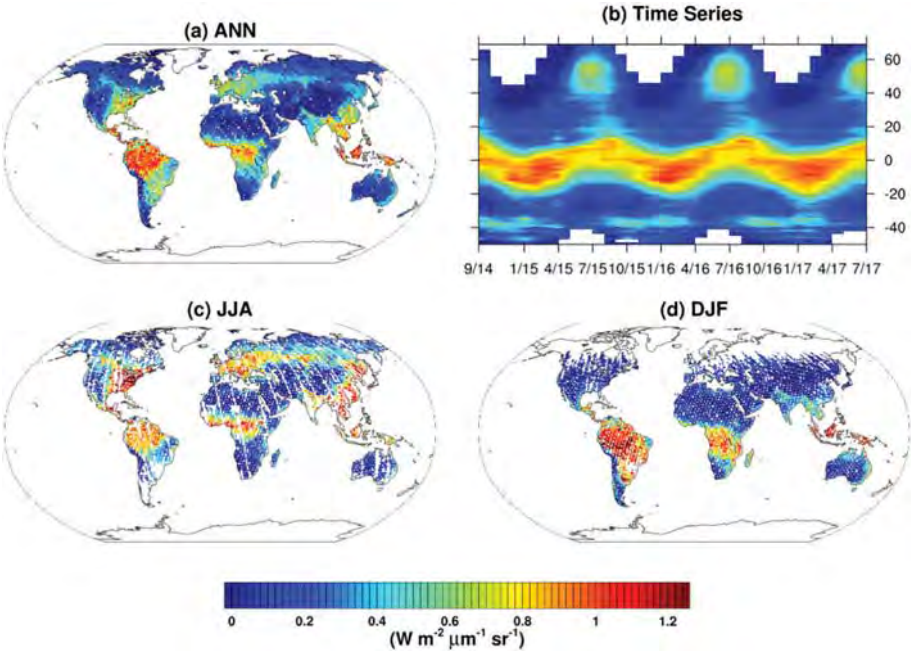


FIGURE 16.17 The Orbiting Carbon Observatory–2 (OCO-2) solar-induced chlorophyll fluorescence (SIF) retrieval (the composite of 757 nm and 771 nm) on a 1° × 1° grid for (a) the annual average, (c) the summer average (June–July–August), and (d) the winter average (December–January–February) of 2015–2016. (b) Time series of the latitudinal mean SIF from September 2014 to July 2017. Only grid-cells with more than five soundings are shown. From Sun et al. (2017). Used by permission of the American Association for the Advancement of Science.

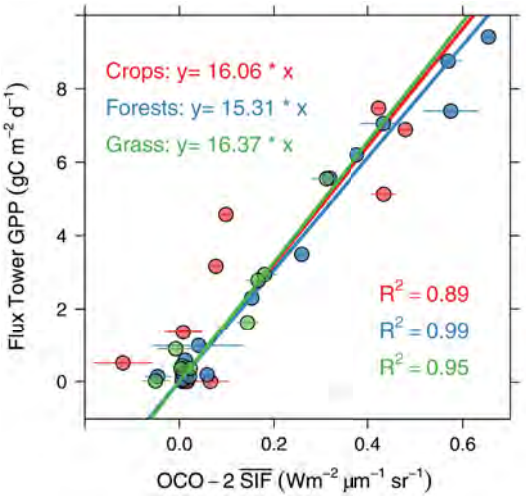


FIGURE 16.18 The relationship between gross primary productivity and OCO-2 SIF (daily mean value, denoted as SIF, converted from instantaneous measurements) at three flux tower sites representative of three different biomes. Error bars represent the standard error of the OCO-2 SIF retrieval. From Sun et al. (2017). Used by permission of the American Association for the Advancement of Science.

16.9 SUMMARY

For a large proportion of the Earth, vegetation cover forms the surface observed by remote sensing instruments. In some instances, we have a direct interest in examining this vegetative blanket in order to map the patterns of different forests, rangelands, and agricultural production. In other instances, we must use vegetation as a means of understanding those patterns that may lie hidden beneath the plant cover. In either instance, it is essential that we be able to observe and understand the information conveyed by the vegetated surface.

Given the obvious significance of plants to global ecosystem function, and sustainable food and fiber systems, the techniques outlined in this chapter are likely to form one of the most significant contributions of remote sensing to the well-being of humankind. The ability to examine vegetation patterns using the techniques described here, combined with the synoptic view and repetitive coverage of satellite sensors, provides an opportunity to survey vegetation patterns and productivity in a manner that was not possible even a few years ago. The timing of our new capabilities is critical, since increases in human population and living standards are changing the structure and function of ecosystems globally.

REVIEW QUESTIONS

1. Summarize differences between classification of vegetation from ground observations and classification from aerial images. Consider such factors as (a) the basis of the classification and (b) the units classified. Identify distinctions for which remote sensing is especially well suited and those for which it is not likely to be useful.
2. To apply some of the knowledge presented in this chapter (e.g., the red shift), it is necessary to have data with very fine spectral, radiometric, and spatial detail. From your knowledge of remote sensing, discuss how this requirement presents difficulties for operational applications.
3. List some of the reasons why an understanding of the spectral behavior of an individual plant leaf is not itself sufficient to conduct remote sensing of vegetation patterns.
4. List some reasons why multispectral satellite images might be especially well suited for observation of vegetation patterns. Or, if you prefer, list some reasons why they are not quite so useful for vegetation studies. Briefly compare the Sentinel-2 MultiSpectral Instrument and Landsat 8 Operational Land Imager with respect to their utility for vegetation studies.
5. Write a short description of a design for a multispectral sensor tailored specifically for recording information about living vegetation and vegetation patterns, disregarding all other applications. Suggest optimum timing for a satellite to carry the sensor in a sun-synchronous orbit.
6. Describe some of the ways that image classification (Chapter 12) might be useful in the study of vegetation patterns. Also identify some of the limitations of such methods in the study of vegetation.
7. How do changes in Sun angle and Sun azimuth (due to differences in season and time of day) influence the way in which vegetative patterns are recorded on remotely sensed images?

8. What is solar induced fluorescence, how is it measured from space, and what is its utility in vegetation studies?
9. Note (with an equation) an index related to leaf water content. Why would such an index also be useful to estimate LAI? Why would it be useful to study vegetation disturbance and recovery?

REFERENCES

Allen, W. A., H. W. Gausman, A. J. Richardson, and R. Cardenas. 1971. Water and Air Changes in Grapefruit, Corn, and Cotton Leaves with Maturation. *Agronomy Journal,* Vol. 63, pp. 392–394.

Bailey, R. G. 1978. *Description of Ecoregions of the United States.* Ogden, UT: U.S. Forest Service, 77 pp.

Bailey, R. G. 1998. *Ecoregions: The Ecosystem Geography of the Oceans and the Continents.* New York: Springer-Verlag, 176 pp.

Card, D. H., D. L. Peterson, P. A. Matson, and J. D. Aber. 1988. Prediction of Leaf Chemistry by the Use of Visible and Near Infrared Reflectance Spectroscopy. *Remote Sensing of Environment,* Vol. 26, pp. 123–147.

Ceccato, P., S. Flasse, S. Tarantola, S. Jacquemoud, and J. M. Grégoire. 2001. Detecting Vegetation Leaf Water Content Using Reflectance in the Optical Domain. *Remote Sensing of Environment,* Vol. 77, pp. 22–33.

Claverie, M., J. Ju, J. G. Masek, J. L. Dungan, E. F. Vermote, J.-C. Roger, . . . and C. Justice. 2018. The Harmonized Landsat and Sentinel-2 Surface Reflectance Data Set. *Remote Sensing of Environment,* Vol. 219, pp. 145–161.

Clevers, J. G. P. W., and W. Verhoef. 1993. LAI Estimation by Means of the WDVI: A Sensitivity Analysis with a Combined PROSPECT SAIL Model. *Remote Sensing Reviews,* Vol. 7, pp. 43–64.

Collins, W. 1978. Remote Sensing of Crop Type and Maturity. *Photogrammetric Engineering and Remote Sensing,* Vol. 44, pp. 43–55.

de Beurs, K. M., and G. M. Henebry. 2004. Land Surface Phenology, Climatic Variation, and Institutional Change: Analyzing Agricultural Land Cover Change in Kazakhstan. *Remote Sensing of Environment,* Vol. 89, pp. 497–509.

Dethier, B. E., M. D. Ashley, B. O. Blair, J. M. Caprio, R. J. Hopp, and J. Rouse. 1973. Phenology Satellite Experiment: Detection of Brown Wave and Green Wave in North-South Corridors of the United States. NASA Technical Report E74–10282, CR-136668, 789 pp.

Fisher, J. I., J. F. Mustard, and M. A. Vadeboncoeur. 2006. Green Leaf Phenology at Landsat Resolution: Scaling from the Field to the Satellite. *Remote Sensing of Environment,* Vol. 100, pp. 265–279.

Hardisky, M. A., V. Klemas, and R. M. Smart. 1983. The Influence of Soil-Salinity, Growth Form, and Leaf Moisture on the Spectral Radiance of Spartina-Alterniflora Canopies. *Photogrammetric Engineering and Remote Sensing,* Vol. 49, pp. 77–83.

Hine, R., and E. Martin (eds.). 2016. *A Dictionary of Biology* (7th ed.). New York: Oxford University Press.

Høgda, K., H. Tømmervik, and S. Karlsen. 2013. Trends in the Start of the Growing Season in Fennoscandia 1982–2011. *Remote Sensing,* Vol. 5, pp. 4304–4318.

Huete, A., K. Didan, T. Miura, E. P. Rodriguez, X. Gao, and L. G. Ferreira. 2002. Overview of the Radiometric and Biophysical Performance of the MODIS Vegetation Indices. *Remote Sensing of Environment,* Vol. 83, pp. 195–213.

Hunt, Jr., E., and B. Rock. 1989. Detection of Changes in Leaf Water Content Using Near- and Middle-Infrared Reflectances. *Remote Sensing of Environment,* Vol. 30, pp. 43–54.

Jackson, R. D., P. N. Slater, and P. J. Pinter. 1983. Discrimination of Growth and Water Stress in Wheat by Various Vegetation Indices through Clear and Turbid Atmospheres. *Remote Sensing of Environment*, Vol. 13, pp. 187–208.

Jin, S., and S. A. Sader. 2005. Comparison of Time Series Tasseled Cap Wetness and the Normalized Difference Moisture Index in Detecting Forest Disturbances. *Remote Sensing of Environment,* Vol. 94, pp. 364–372.

Khosravi, N., M. Vountas, V. V. Rozanov, A. Bracher, A. Wolanin, and J. P. Burrows. 2015. Retrieval of Terrestrial Plant Fluorescence Based on the In-Filling of Far-Red Fraunhofer Lines Using SCIAMACHY Observations. *Frontiers in Environmental Science,* Vol. 3, Article No. 78.

Knipling, E. B. 1970. Physical and Physiological Basis for the Reflectance of Visible and Near-Infrared Radiation from Vegetation. *Remote Sensing of Environment,* Vol. 1, pp. 155–159.

Kokaly, R. F., G. P. Asner, S. V. Ollinger, M. E. Martin, and C. A. Wessman. 2009. Characterizing Canopy Biochemistry from Imaging Spectroscopy and Its Application to Ecosystem Studies. *Remote Sensing of Environment,* Vol. 113, pp. S78–S91.

Kokaly, R. F., and R. N. Clark. 1999. Spectroscopic Determination of Leaf Biochemistry Using Band-Depth Analysis of Absorption Features and Stepwise Multiple Linear Regression. *Remote Sensing of Environment,* Vol. 67, pp. 267–287.

Koriba, K. 1958. On the Periodicity of Tree-Growth in the Tropics, with Special Reference to the Mode of Branching, the Leaf Fall, and the Formation of the Resting Bud. *Gardens' Bulletin* (Singapore), Series 3, Vol. 17, pp. 11–81.

Küchler, A. W. 1967. *Vegetation Mapping.* New York: Ronald Press. 472 pp.

Kupiec, J., and P. J. Curran. 1993. AVIRIS Spectra Correlated with the Chlorophyll Concentration of a Forest Canopy. In *Summaries of the Fourth Annual JPL Airborne Geoscience Workshop*. JPL Publication 93–26. Pasadena CA: Jet Propulsion Laboratory, Vol. 1, pp. 105–108.

Martin, M. E., and J. D. Aber. 1997. High Spectral Resolution Remote Sensing of Forest Canopy Lignin, Nitrogen, and Ecosystem Processes. *Ecological Applications,* Vol. 7, pp. 431–443.

Morisette, J. T., A. D. Richardson, A. K. Knapp, J. I. Fisher, E. A. Graham, J. Abatzoglou, . . . and L. Liang. 2009. Tracking the Rhythm of the Seasons in the Face of Global Change: Phenological Research in the 21st Century. *Frontiers in Ecology and the Environment,* Vol. 7, pp. 253–260.

Moya, I., L. Camenen, S. Evain, Y. Goulas, Z. G. Cerovic, G. Latouche, . . . and A. Ounis. 2004. A New Instrument for Passive Remote Sensing: 1. Measurements of Sunlight-Induced Chlorophyll Fluorescence. *Remote Sensing of Environment,* Vol. 91, pp. 186–197.

Mutanga, O., and A. K. Skidmore. 2007. Red Edge Shift and Biochemical Content in Grass Canopies. *ISPRS Journal of Photogrammetry and Remote Sensing,* Vol. 62, pp. 34–42.

Myneni, R. B., F. G. Hall, P. J. Sellers, and A. L. Marshak. 1995. The Interpretation of Spectral Vegetation Indexes. *IEEE Transactions on Geoscience and Remote Sensing,* Vol. 33, pp. 481–486.

Myneni, R. B., S. Hoffman, Y. Knyazikhin, J. L. Privette, J. Glassy, Y. Tian, . . . and S. W. Running. 2002. Global Products of Vegetation Leaf Area and Fraction Absorbed PAR from Year One of MODIS Data. *Remote Sensing of Environment,* Vol. 83, pp. 214–231.

Pellissier, P. A., S. V. Ollinger, L. C. Lepine, M. W. Palace, and W. H. McDowell. 2015. Remote Sensing of Foliar Nitrogen in Cultivated Grasslands of Human Dominated Landscapes. *Remote Sensing of Environment,* Vol. 167, pp. 88–97.

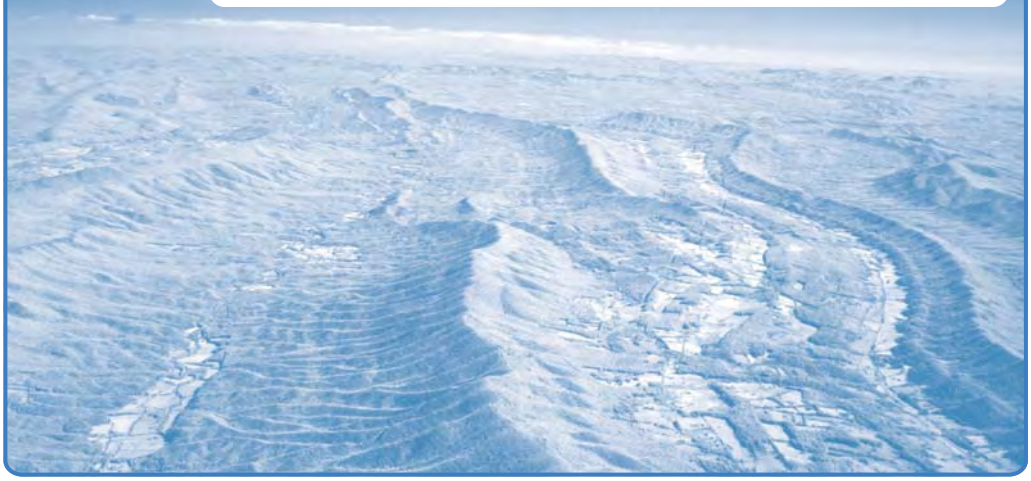
Perry, C. R., and L. F. Lautenschlager. 1984. Functional Equivalence of Spectral Vegetation Indices. *Remote Sensing of Environment,* Vol. 14, pp. 169–182.

Peterson, D. L., J. D. Aber, P. A. Matson, D. H. Card, N. Swanberg, C. Wessman, and M. Spanner. 1988. Remote Sensing of Forest Canopy and Leaf Biochemistry Contents. *Remote Sensing of Environment,* Vol. 24, pp. 85–108.

Plascyk, J. A., and F. C. Gabriel. 1975. The Fraunhofer Line Discriminator MKII-An Airborne

Instrument for Precise and Standardized Ecological Luminescence Measurement. *IEEE Transactions on Instrumentation and Measurement,* Vol. 24, pp. 306–313.
Price, J. C. 1987. Calibration of Satellite Radiometers and Comparison of Vegetation Indices. *Remote Sensing of Environment,* Vol. 18, pp. 35–48.
Reed, B. C., and L. Yang. 1997. Seasonal Vegetation Characteristics of the United States. *Geo-Carto International,* Vol. 12, pp. 65–71.
Rouse, J. W., R. H. Haas, J. A. Schell, D. W. Deering, J. C. Sagan, C., O. B. Toon, and J. B. Pollock. 1979. Anthropogenic Changes and the Earth's Climate. *Science*, Vol. 206, pp. 1363–1368.
Ruimy, A., B. Saugier, and G. Dedieu. 1994. Methodology for the Estimation of Terrestrial Net Primary Production from Remotely Sensed Data. *Journal of Geophysical Research,* Vol. 99, pp. 5263–5283.
Sims, D. A., and J. A. Gamon. 2002. Relationships between Leaf Pigment Content and Spectral Reflectance across a Wide Range of Species, Leaf Structures and Developmental Stages. *Remote Sensing of Environment,* Vol. 81, pp. 337–354.
Stöckli, R., and P. L. Vidale. 2004. European Plant Phenology and Climate as Seen in a 20-Year AVHRR Land-Surface Parameter Dataset. *International Journal of Remote Sensing,* Vol. 25, pp. 3303–3330.
Sun, Y., C. Frankenberg, M. Jung, J. Joiner, L. Guanter, P. Köhler, and T. Magney. 2018. Overview of Solar-Induced Chlorophyll Fluorescence (SIF) from the Orbiting Carbon Observatory-2: Retrieval, Cross-Mission Comparison, and Global Monitoring for GPP. *Remote Sensing of Environment,* Vol. 209, pp. 808–823.
Sun, Y., C. Frankenberg, J. D. Wood, D. S. Schimel, M. Jung, L. Guanter, . . . and K. Yuen. 2017. OCO-2 Advances Photosynthesis Observation from Space via Solar-Induced Chlorophyll Fluorescence. *Science,* Vol. 358, Article No. 6360, 6 pp.
Tucker, C. J. 1979. *Remote Sensing of Leaf Water Content in the Near Infrared* (NASA Technical Memorandum 80291). Greenbelt, MD: Goddard Space Flight Center, 17 pp.
Tucker, C. J. 1980. Remote Sensing of Leaf Water Content in the Near-Infrared. *Remote Sensing of Environment,* Vol. 10, pp. 23–32.
Turner, D. P., W. B. Cohen, R. E. Kennedy, K. S. Fassnacht, and J. M. Briggs. 1999. Relationships between Leaf Area Index and Landsat TM Spectral Vegetation Indices across Three Temperate Zone Sites. *Remote Sensing of Environment,* Vol. 70, pp. 52–68.
Turner, W., S. Spector, N. Gardiner, M. Fladeland, E. Sterling, and M. Steininger. 2003. Remote Sensing for Biodiversity Science and Conservation. *Trends in Ecology and Evolution.* Vol. 18, pp. 306–314.
Walker, J. J., K. M. de Beurs, and R. H. Wynne. 2012. Evaluation of Landsat and MODIS Data Fusion Products for Analysis of Dryland Forest Phenology. *Remote Sensing of Environment,* Vol. 117, pp. 381–393.
Wessman, C. A., J. D. Aber, and D. L. Peterson. 1989. An Evaluation of Imaging Spectrometry for Estimating Forest Canopy Chemistry. *Remote Sensing of Environment,* Vol. 10, pp. 1293–1316.
Wynne, R. H., T. M. Lillesand, M. K. Clayton, and J. J. Magnuson. 1998. Satellite Monitoring of Lake Ice Breakup on Laurentian Shield (1980–1994). *Photogrammetric Engineering and Remote Sensing,* Vol. 64, pp. 607–617.
Zhang, Y., L. Guanter, J. A. Berry, J. Joiner, C. van der Tol, A. Huete, . . . and P. Köhler. 2014. Estimation of Vegetation Photosynthetic Capacity from Space-Based Measurements of Chlorophyll Fluorescence for Terrestrial Biosphere Models. *Global Change Biology,* Vol. 20, pp. 3727–3742.
Zhang, Y., J. Joiner, S. H. Alemohammad, S. Zhou, and P. Gentine. 2018. A Global Spatially Contiguous Solar-Induced Fluorescence (CSIF) Dataset Using Neural Networks. *Biogeosciences,* Vol. 15, No. 19, pp. 5779–5800.

17 Agricultural Remote Sensing



MAJOR TOPICS TO UNDERSTAND

- Croplands
- Technical Limitations for Agricultural Applications
- Common Agricultural Practices
- Important Crops
- Monitoring the Growth Stage
- Irrigation
- A View of an Irrigated Landscape, Western Kansas
- Crop Calendar
- Crop Damage
- Conservation Tillage
- USDA Cropland Data Layer
- Biophysical Dimensions of Agricultural Remote Sensing
- Precision Agriculture
- Remote Sensing of Plant Pathology

17.1 INTRODUCTION

In the early days of manned flight, aviators recognized the value of the aerial view to observe croplands when they were quick to appreciate the value of aviation's unique overhead perspective. However, until aerial cameras and films were available, they could provide little practical value for the farmer. In the United States, by the 1930s, aerial cameras surveyed croplands to estimate agricultural production and to support soil survey and erosion control. Despite the value of such imagery for agricultural inventory, it was decades more before the technology and analysis progressed to the point where we could record and monitor agricultural systems, assess crop status throughout the growing season, and routinely acquire aerial imagery of the agricultural landscape.

In the 1950s, applications of Robert Colwell's research (introduced in Chapter 1) formed the foundation for practical applications within U.S. agriculture and forestry. The availability of improved aerial cameras, as well as the availability of infrared films, facilitated the effectiveness of Colwell's contributions. By the early 1970s, the U.S. Department of Agriculture (USDA), in collaboration with other agencies, began systematic collection of agricultural data using remotely sensed imagery. Shortly thereafter, the availability of Landsat imagery nurtured interest and vision for systematic use of satellite imagery for agricultural analysis. Later, improved spatial and spectral detail of Enhanced Thematic Mapper (ETM) imagery of Landsats 4, 5, and 7 provided the basis for analytical strategies for conducting agricultural analysis of satellite imagery (Landsat Legacy Project Team, 2017). These efforts defined and validated the effectiveness of many of the analytic strategies that we now routinely use to examine digital satellite data, not only in agriculture, but also in many other applications.

17.2 CROPLANDS

Croplands identify land devoted to *cultivated* row crops, or close-grown crops, perhaps rotated with hay or pasture. The alternative, *noncultivated cropland,* includes land permanently devoted to hay or horticultural cropland (a subcategory of cropland used for growing fruits, nuts, berry, vineyard, and other bush fruit and similar crops). Here we focus on croplands because of their role in the worldwide production of food (Figure 17.1) and the key roles that remote sensing plays in monitoring croplands and forecasting the economic dimensions of food production.

Figure 17.1 represents croplands worldwide (at a nominal 30-m resolution), derived from Landsat imagery for the year 2015. The pattern highlights both the immense land areas devoted to croplands, and simultaneously, the vast land areas that are not now cultivated (due to factors such as climate, terrain, and urbanization). The image highlights the value of broad-scale multispectral remote sensing as a strategy for routine monitoring of the large region globally devoted to agriculture. At this scale, regional, national, and global patterns in agricultural productivity, land use transition, cultivation, and response to major environmental stresses (i.e., fires, droughts, etc.) can be observed.

While broad patterns in agricultural activity are important, it is also valuable to assess agriculture at the scale of the farm, where the food is grown and the land is managed. Figure 17.2 presents a detailed view of croplands in Noxubee County, Mississippi, providing the farmer with an aerial perspective from imagery transmuted in real time. Here, the local region's agricultural profile includes corn, soybeans, grains, cotton,

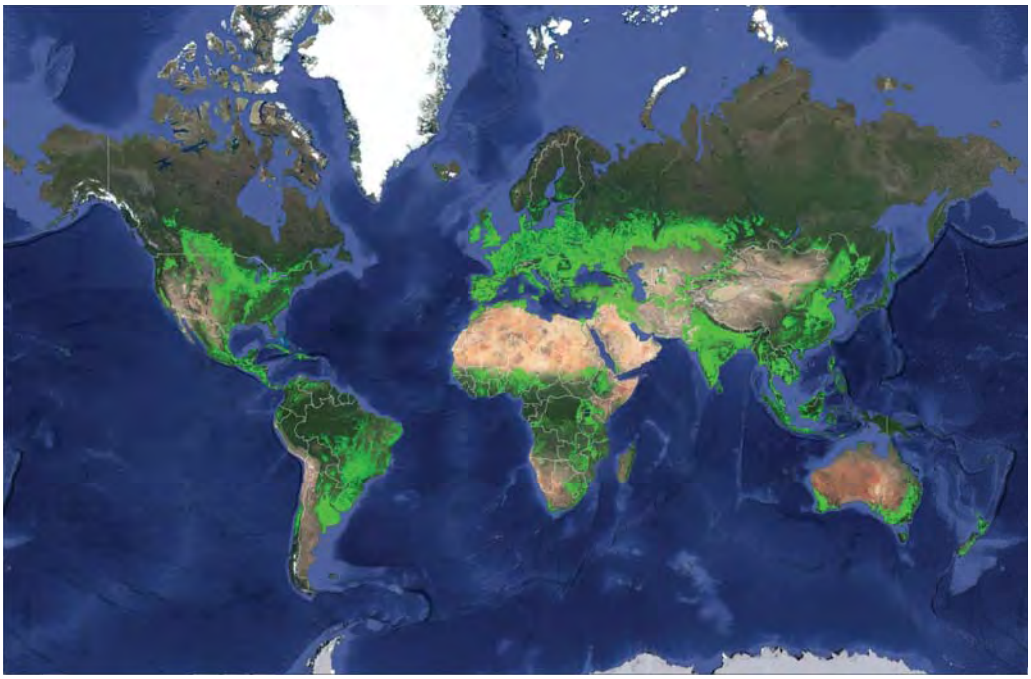


FIGURE 17.1 Map showing croplands (bright green) recorded by Landsat imagery for the year 2015. The croplands depicted here represent different crops that vary with latitude, season, and climate. From U.S. Geological Survey (USGS), GFSAD30 Project.

woodlands, pasture, and aquaculture. It is now possible for each farmer to view their cropland in the context of nearby landscapes.

Thus, the practice of remote sensing forms a vehicle for understanding agricultural concerns through its ability to examine landscapes at varied scales, record temporal changes, and detect spectral anomalies that reveal, for example, incipient signs of drought, diseases, or insect infestations. Agricultural systems are characterized by broad geographic scope, seasonal changes, local variation, and intimate interactions with the

FIGURE 17.2 Aerial image (Noxubee County, Mississippi) showing crops at differing growth stages. This image was transmitted directly from aircraft by a high-speed wireless system that enables farmers to download aerial images directly to their personal computers as they are acquired. Crops here include corn, soybeans, and forage. From the U.S. Department of Agriculture (USDA), Agricultural Research Service (ARS), no. K10384.

landscape, so remote sensing's inherent capabilities match well to observation of agricultural systems.

Remote sensing provides information for significant portions of the world's agricultural production, including for crops that enter international commerce for distribution of food to regions that have experienced crop failures, flooding, conflict, and similar food crises. Remote sensing contributes to the following (see Atzberger, 2013):

1. Biomass and yield estimation, including estimation of crop yields and dates of harvest.
2. Vegetation vigor and stress monitoring, including surveying of crop damage, storm, drought impact, flooding, and frost.
3. Assessment of crop phenological development (Chapter 16).
4. Crop acreage estimation and cropland mapping, including assessment of tillage practices, crop rotation, harvests, and broad-scale monitoring of agricultural lands to estimate extent of agricultural harvest.
5. Mapping of disturbances and land use/land cover changes.

Combining these capabilities of remote sensing over time allows for the long-term monitoring of harvests and the projection of future harvests. This provides the basis for forecasting prices and stabilizing commodity markets, as well as predicting the nature and locations of food shortages.

17.3 TECHNICAL LIMITATIONS FOR AGRICULTURAL APPLICATIONS

Despite the many strengths that remote sensing systems bring to our ability to observe agricultural productivity, important limitations and shortcomings deserve our attention.

Spatial resolution: If we take satellite systems as an example, there are important variations in pixel characteristics within a single image due to differences between nadir and off-nadir perspectives. Pixels acquired by an optical sensor viewing from a nadir perspective will have size and shape as specified by instrument design specifications. Pixels acquired off-nadir will decrease in size as lateral distance from the nadir increases, generating coarser detail and masking details of crops rather than providing an overhead view. Furthermore, a lateral perspective increases the atmospheric path of radiation that reaches the sensor, increasing atmospheric effects, and contaminating spectral signatures relative to a nadir view.

Cross-calibration of varied sensors: As of November 2017, there were 596 active Earth observation satellites, about 327 of which were using optical instruments. Many of these instruments were designed for specific purposes, so they differ in bandwidth, spatial detail, and calibration. For some studies, acquiring imagery to provide the needed temporal resolution may require use of images from several satellites, each with differing optical characteristics (Song, 2004; Liu, Gopal, and Woodcock, 2004). For example, differing viewing geometries, overpass schedules, and calibration gains generate differences in imagery of the same region.

Much of this chapter focuses on applications of remote sensing to broad-scale observation of landscapes, such as the central regions of North America, where fields are large, follow predictable crop cycles, and are dedicated to rather limited selections of crops.

In other regions, fields may be smaller, the crops more varied, and weather and climate characterized by frequent cloud cover. Remote sensing is often less effective in such situations but still provides valuable information through knowledge of the local agricultural scene. The value of remotely sensing imagery in an agricultural context lies not only in its broad scope and ability to collect sequential imagery, but also its ability to provide spatial detail that supports understanding of agricultural management, identification of crops, growth status, diseases, and infestations.

17.4 COMMON AGRICULTURAL PRACTICES

Considered worldwide, significant agricultural crops include wheat, barley, corn (maize), rice, potatoes, sweet potatoes, yams, soybeans, plantains, cassava, and sorghum (milo). All have specific histories and roles in regional agriculture and diets. The significance of such crops can be ranked by acres under cultivation, dietary significance, and economic value, among other considerations.

Except at the very finest levels of detail, remote sensing does not normally permit reliable identification of specific plants or even identification of fields of the same plant. As a result, crops are typically classified as *large-grain* (such as corn and sorghum) or *small-grain* (such as wheat, barley, oats, or rye). From a remote sensing perspective, the distinction is not so much about sizes of the grains themselves, but rather about the structures of mature plants. Large-grain crops are typically planted in rows; plants grow to be large enough and separated by rows wide enough to be recognizable on aerial imagery through their distinctive texture. In contrast, small-grain crops have thinner, less rigid stalks that, once mature, appear on imagery as even, smooth textures that can be reliably separated from large-grain crops. Furthermore, as noted below, analysts familiar with the local scene can understand local crops, together with planting and harvesting schedules.

For large areas in the midlatitudes, crop production is often based on raising a single crop (or, perhaps, two crops in rotation from one year to the next) within the same field, with little year-to-year variation. Under such conditions, identification of crops from aerial imagery may present few challenges, as most analyses can rely on familiar crops, agricultural practices, and established crop calendars (discussed below). In other circumstances, crop identification by aerial observation can be complicated by alternative practices that intermingle varied crops within the same field. For example, *intercropping* identifies practices that mix separate plants within the same field, often used in community gardens and as routine practice in tropical regions. Likewise, aerial observations of croplands can be complicated by practices that produce two or more crops within the same fields.

Multicropping refers to the practice of producing two or more crops within a single growing season. *Double cropping/triple cropping* refers to production of two or three different crops in succession within the same fields in a single growing season. In tropical regions, especially irrigated fields, multiple plantings of the same crop can be grown in the same field. This is common practice for growing rice in tropical regions or cotton in irrigated regions of the southeastern United States. (*Monocropping* refers to production of a single crop year after year on the same land, without rotation, whereas *polycropping* is growing multiple crops in the same fields in order to protect against disease and insects.)

Crop rotation (distinct from double cropping) is the practice of alternating different crops within the same field from year to year, typically following similar crop calendars. This is done to maintain soil structures, restore nutrients, reduce pathogens, and maintain soil health. For example, legumes contain nitrogen-fixing bacteria, and so they are often alternated with cereal crops that require nitrates. In the U.S. Corn Belt, a common crop rotation sequence is to alternate soybeans and corn crops. (Soybeans, as legumes, restore the nitrogen that corn crops take from the soil.)

Dryland farming: In regions of low or highly variable precipitation, farmers must cultivate crops without irrigation, so they rely on a suite of practices that can improve the environment for crops by tilling fields after harvest, suppressing weeds, reducing moisture loss, and retaining tillage debris in the field after harvest (*no-till agriculture*) to reduce erosion, capture snow, and reduce runoff.

Drought-resistant crops, such as sorghum (**Figures 17.3 and 17.4**), are well matched to environments with low or uncertain moisture, through lower transpiration, slower growth, or dormancy during drought conditions, which often can reduce transpiration and may lessen moisture demands during periods of moisture shortage, resuming growth when conditions again become favorable. Drought-evasive crops achieve their main growth during those times of the year when heat and drought conditions are not severe. Crops adapted to dryland farming are usually smaller and quicker to mature than those grown under more humid conditions and are usually allotted more space.

Fallow land: Farmers sometimes specifically leave some of their fields without seeding for the next growing season, either plowed or unplowed. Fallow lands can restore fertility, possibly as part of a crop rotation strategy, accumulate soil moisture, or avoid surplus production. Fallow fields may be treated with pesticides or herbicides to reduce weeds or insects.

Agricultural burning: In some areas, such as the western plains, farmers may periodically burn their fields to remove plants or residue to prepare fields for the next season. Often such burning, known as "prescribed" fire programs, is regulated by local laws that require safety permits, specified practices, and regulated timing. Although practitioners may use unmanned aerial vehicles (UAVs) to monitor prescribed burning, it is not typically the subject of remote sensing analysis. However, large regions of Africa and India,

FIGURE 17.3 Coarse-grain crop: sorghum, approaching harvest. Finney County, Kansas.

FIGURE 17.4 Sorghum crop, approaching harvest. Finney Country, Kansas. From USDA, ARS.

for example, are characterized by extensive agricultural burning, monitored by Moderate Resolution Imaging Spectroradiometer and Visible Infrared Imaging Radiometer Suite, which provide seasonal reviews of aerial imagery (Pereira, 2003).

17.5 IMPORTANT CROPS

In this section, we are not so much concerned with a complete inventory of significant crops, but rather with noting that many of the food crops most easily assessed by remote sensing include many of the most significant crops essential to world food security, including rice, wheat, corn, barley, soybeans, and sorghum. As we review this list, it's important to consider that (1) this list is a concise selection from a broad inventory; (2) worldwide, many other crops are significant locally or regionally and are not easily monitored using approaches discussed in this chapter; (3) other crops not listed above (such as cotton, for example) have important industrial applications; and (4) still others (corn, for example) have both nutritional and industrial significance.

Corn/maize: Corn, a cereal grain native to Mexico, forms one of the world's most widely grown crops (from 58°N latitude in Canada and Russia to 40°S latitude in South America). Corn has many uses, including fodder, human food, industrial material, and biofuels. In the United States, the major uses include livestock feed (30%), ethanol (40%), and export (10–20%, much of which is also used for livestock), with the remainder used mainly for food and beverages. In midlatitudes, corn is planted in spring and harvested in late summer or autumn. Although corn is efficient in its use of soil moisture, it is shallow-rooted, and so it can be sensitive to late-season drought. Maize used for silage is harvested late in the growing season, while the leaves are green; field maize is left in the field until late autumn and sometimes is not harvested until winter.

Sorghum (also known as milo, grain sorghum, and other names) is a tall annual, resembling corn in some respects; it is known for its drought resistance, in part because of its extensive root system (**Figures 17.3 and 17.4**). Relative to other crops, sorghum requires relatively little moisture and so is favored in arid climates. Sorghum for export from the United States includes fodder, fuel (ethanol), human food, flour, syrup, and molasses. In the United States, sorghum is grown rather widely, but it is especially prominent in the western edges of the Great Plains and other regions such as Arizona and California, where rainfall is limited. There, it is used for cover crops, forage, and feed grains. Sorghum is widely used in India, China, and Africa.

Soybeans, native to Asia, were introduced to North America in the late 1700s and are now widely grown in the United States and elsewhere (**Figure 17.5**). They grow as a grass-like plant—a species of legume valuable for its edible bean. As forage for livestock, initially in the southeastern part of the United States and later in the midwestern states, soybeans became a significant U.S. agricultural crop. In 1904, the American agricultural chemist George Washington Carver recognized that soybeans are a valuable source of nutrients that thrive in a variety of soils. Like many legumes, soybeans fix atmospheric nitrogen, which can restore nitrogen to soils. As a result, soybeans are often rotated with a corn crop to retain soil productivity and increase yields of both corn and maize. Today, additional soybean producers include China, Uruguay, Bolivia, Brazil, and Argentina.

Wheat is a grass that is grown globally and is often widely regarded as the world's second-most important cereal crop (after maize and ahead of rice). Wheat is regarded as the leading source of vegetal protein in human food and is raised in many varieties

FIGURE 17.5 Geneticist Thomas Carter working to develop drought-tolerant soybean breeding lines that can grow over an expanded geographic range, Sandhills Research Station, North Carolina. From USDA. Photograph by Thomas Carter, no. d1257-1.

throughout midlatitudes, including Argentina, the United States, Canada, China, Central Asia, and Europe. *Winter wheat* is a strain of wheat planted in autumn to germinate and sprout before winter weather sets in (**Figures 17.6 and 17.7**). During winter, the crop lies dormant and resumes growth in early spring for a summer harvest.

Rice, a cereal grain of Asian origin, is recognized as the most widely consumed staple food for a large proportion of the world's population. In its many varieties, it is raised, sometimes as an annual, in a wide range of locations and landscapes, provided the climate is relatively warm and a reliable access to water resources is available, includ-

FIGURE 17.6 Example of small-grain crop: winter wheat. USDA Extension agent Wayne Cooley, ARS agronomist Randy Anderson, and farmer Gilbert Lindstrom examine a wheat crop as they consider a wheat/corn/fallow rotation within a dryland cropping system. From USDA, ARS.

FIGURE 17.7 Small-grain crop: winter wheat, approaching harvest, Finney County, Kansas.

ing coastal zones, deltas, estuaries, and river floodplains. Upland rice is produced in quite different environments, often rain-fed fields, and is seeded when dry, similar to cultivation of maize or wheat. In steep terrain, construction supports rice cultivation in uneven terrain distant from the coastal regions often associated with rice production. Worldwide, rice crops are susceptible to numerous diseases (including bacteria, fungi, and insects), as well as to extremes of weather and climate and other hazards. Effective management therefore requires both expertise and experience.

In Asia, rice paddy fields are prepared by plowing, fertilization, and bed preparation (smoothing the soil to prepare the seed bed to ensure uniform water depth and accurate control of irrigation). By tradition, it is common practice for rice farmers to plant seedlings in beds outside fields and then, 30 to 50 days later, transplant to fields flooded by rain or irrigation. It is now common in some regions for farmers to plant the crop directly, without transplantation, to well-prepared, irrigated fields as a means of avoiding weather delays in the planting schedule, which can increase the risks of disease and other hazards.

During the growing season, farmers carefully control irrigation manually or by gates in local dikes. Farmers drain fields before harvest (**Figure 17.8**). There is a wide variation in the specific varieties of rice; some agricultural practices are tailored to support rice production in varied locations. Outside Asia, leading exporters, such as the United States, Australia, and Brazil, rely on mechanized production in large inland fields, with reliable access to water. In some remote regions, rice may be cultivated on a subsistence basis, a practice largely outside the scope of remote sensing analysis.

Figure 17.9 illustrates the significance of rice production in the United States, forming the nation's third largest source of cereal production, after corn and wheat. Here we can see the scope of mechanized harvest for Fort Bend County, Texas, with additional production in Arkansas, California, Louisiana, Mississippi, and Missouri. Typically the United States harvests a single rice crop each year from rice fields, although Texas and southwest Louisiana may harvest a second crop from a single planting (known as a "ratoon" crop), due to the effect of a longer growing season.

Remote sensing forms an important asset for monitoring the progress of the international rice crop throughout South Asia, especially Vietnam, the Philippines, India, China, and Korea. Both synthetic aperture radar (SAR) imagery and Landsat/Sentinel optical imagery have proven effective in assessing rice cultivation, especially in examining double-cropping and triple-cropping, which are common practices in some regions (Son et al., 2013; Lam-Dao et al., 2017).

FIGURE 17.8 Rice fields maturing as harvest date approaches, Son La Province, Vietnam. Photograph by Hoa Tran. Used by permission.

FIGURE 17.9 Aerial view of mechanized rice harvest, Fort Bend County, Texas. From USDA. Photograph by David Nance.

17.6 MONITORING THE GROWTH STAGE

The *growth stage* of any crop influences its appearance on aerial imagery. Seldom can most of us recognize crops as individual plants, but rather by the overall appearance of fields during specific seasons, by the growth stages of crops, and by characteristics such as the spatial arrangement of fields and clusters of plants. Specific crops rarely display distinct spectral signatures because all vegetation appears relatively similar in spectral response compared to other features on the landscape. Furthermore, the spectral response differs according to the agricultural calendar and even within the same species. Often, it may not be possible to distinguish between specific crops within each class, so remotely sensed imagery is seldom the principal source of information regarding the specific identity of crops. However, some broad distinctions can be made using aerial imagery, provided image scale is large enough, and plants are approaching maturity.

Agricultural scientists often define many stages for specific crops. These vary from crop to crop, and not all stages can be determined from aerial photography. However, it is often possible to interpret principal growth stages using large-scale aerial photography. Typically, when the crop is planted, the field is mainly bare, tilled soil. Initially, there is little evidence of the presence of plants.

Emergence indicates that seedlings have sprouted, providing visible evidence of the plants. As crops emerge after planting and germination, they present leafy growth above ground. Initially, this growth is usually barely detectable on aerial photography (on color-infrared [CIR] photography, possibly by a light pink tint). At this stage the field, as seen from above, is primarily bare soil with only slight evidence of the emerging crop. Although plants at this stage are usually too small to identify as individual crops, their identities may be indicated by planting pattern, the presence or absence of irrigation, and other observations. Further growth progressively presents larger areas of leaf, shielding the soil from view and indicating the increasing maturity of the crop. As crops mature, leaves begin to shade the soil that was formerly visible between plants, presenting an increasingly complete cover as seen from above. Depending on the specific crop and its method of cultivation, the linear pattern of rows from initial planting may be visible. Assuming that the crop matures without drought or attack by disease or insects, the ground surface will likely be completely covered as the crop matures.

As the crop reaches maturity, many grains will exhibit *senescence* (i.e., aging of the plant, when growth slows or stops, usually indicated by drying of the leaves). As seen on CIR aerial photography, mature crops eventually attain a lighter color and often a grayish or greenish cast as near-infrared (NIR) reflectance decreases. At maturity, coarse grains usually can be recognized by their rougher textures as they appear in the field, due to their larger stalks, coarser leaves, complex structure, and increased micro-shadowing. Small grains, once mature, often have an even, smooth texture that tends to mask rows and irregularities in terrain. Row crops, such as soybeans, are usually recognizable by the latent linear appearance that remains even after the crops are mature enough to mask the rows themselves. *Harvest* occurs, weather permitting, at or near maturity. After harvest, the crop residue often remains on the field until the field is tilled just prior to planting the next year's crop. This gives fields a greenish cast in a CIR image, although the bare soil usually brightens the overall tone of harvested fields.

Although remotely sensed imagery can provide valuable information about the agricultural landscape, it generally is not used as the principal means of identifying specific crops. Thus, aside from the ability of aerial imagery to provide a broad sketch of crops

grown in a region, its value is usually to assess acreages planted to croplands or left as fallow; mark the progress of crops as they mature through the growing season; detect disease and insect infestation; quantify the effects of drought; and obtain other data relating to crop status.

Throughout the growing season, agronomists monitor the progression of the crop regionally through the agricultural calendar to assess disruptions from the usual timetable. Aerial imagery can form a valuable aid in documenting damage for insurance purposes, guiding response to infestations, and assisting commodity markets.

17.7 IRRIGATION

Irrigation is the practice of controlled application of water to agricultural lands to promote the production of crops. This practice dates from antiquity and is evidence of humanity's ingenuity in devising varied technologies for collecting and distributing water for agriculture. Today, wide varieties of irrigation strategies are in use in the United States (Figure 17.10) and throughout the world. They rely on access to a reliable supply of

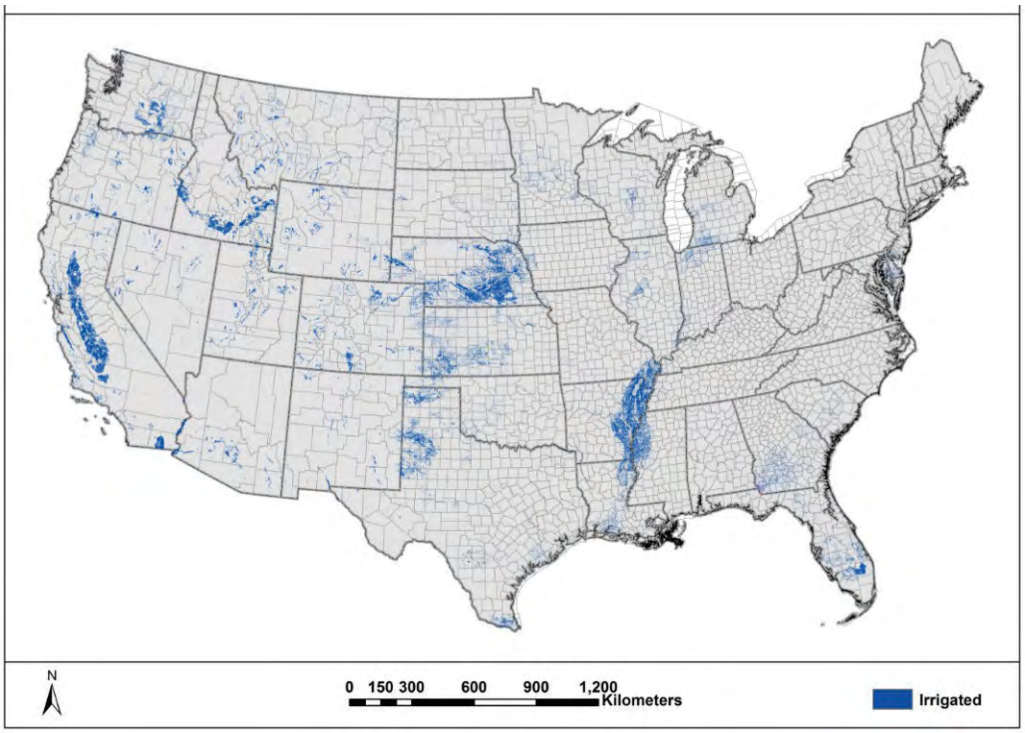


FIGURE 17.10 Irrigated agriculture, United States, presented as a map based on 2012 Moderate Resolution Imaging Spectroradiometer data. Much of this pattern relates to the availability of surface water or to reliable groundwater supplies. Some areas require irrigation because of arid climates, others require irrigation, not because of the prevailing aridity, but because of seasonal needs for water at specific times of the year (e.g., late summer, when evaporation may be high). From USGS.

water, including surface water (pumped from river channels and diverted from rivers and streams), groundwater (pumped by wind or natural gas), dams, and numerous varieties of alternatives.

Ditch/gravity/flood/furrow irrigation is a traditional low-tech irrigation strategy that redirects water from surface streams to fields by a series of channels ("ditches," which are sometimes lined to prevent seepage), controlled by a series of gates (**Figures 17.11 and 17.12**). Ditch irrigation may require expensive land-leveling to remove terrain irregularities and attain the desired grade to permit uniform distribution of water and prevent ponding.

Drip irrigation distributes water through a system of tubes or conduits perforated to allow water to seep into the soil surface. Farmers position such tubes within furrows at the soil surface. Drip irrigation makes economical use of the water supply, although buried conduits are expensive to install.

As mentioned above, the irregular terrain of Asia's upland regions have supported, sometimes for many centuries, *stone terraces* carefully engineered to provide even surfaces for cultivating rice and for controlling distribution of irrigation water. Such terraces are expensive to construct and maintain, and so they represent substantial investments in labor and productivity.

Sprinkler/spray/center-pivot irrigation systems distribute water by pressure through the air using sprinklers, tubes, and spray guns. Although they are still favored in many areas, sprinkler systems can be inefficient because of water losses through wind and evaporation, especially in arid regions where wind and dry air can deplete water before it reaches plants. Also included in the sprinkler system category, as a form of spray irrigation, is *center-pivot irrigation* (**Figure 17.13**), which uses a line of A-shaped metal frames mounted on rolling wheels that direct pressurized water through sprinkler heads to dis-

FIGURE 17.11 Aqueduct channel, with drop structure and sluice gates feeding irrigation water to lateral channels and croplands, Phoenix, Arizona.

FIGURE 17.12 Furrow irrigation, Phoenix, Arizona. Cement-lined channel in foreground carries irrigation water to furrows within fields. Curved siphon tubes transport water over the edge of the channel to the furrows. Photograph by Baojuan Zheng. Used by permission.

tribute to the fields as the line rotates slowly around a central point at the center of the field. Electric motors slowly power the frames to rotate in a large circle, spraying water as they pivot. Center-pivot systems are widely used throughout the world, usually relying on groundwater supplies. This approach is efficient in the sense that it does not require land-leveling and has permitted agriculture to extend into lands that otherwise would not be cultivated because of lack of access to water or because of uneven terrain. This type of irrigation is obvious from above, owing to a distinct circular shape.

FIGURE 17.13 Left: Center-pivot sprinkler irrigation in use, Nebraska. From USGS. Right: View of terminal truss at the end of a different but similar system.

Irrigation infrastructure is expensive both to install and to maintain, so it represents a sizable and continuing investment. Irrigation is, of course, significant in arid regions where water is absent, insufficient, or unreliable. It is also important in humid climates where annual precipitation totals are adequate but poorly distributed throughout the growing season. So, for example, irrigation may protect a corn crop as it approaches maturity late in the growing season (August, for example) when evapotranspiration may be high and rainfall low or unreliable.

From a broader remote sensing perspective, irrigation can complicate the timing of remote sensing analysis of agricultural systems. Because irrigation permits farmers to control the growing cycle, the observed seasonal patterns may not match expected crop calendars, which may create challenges for temporal assessments.

17.8 A VIEW OF AN IRRIGATED LANDSCAPE, WESTERN KANSAS

The scene represented in Figure 17.14 provides a retrospective glimpse of irrigation practices in the western edge of the Great Plains. Here we see a portion of the High Plains landscape west of Garden City, Kansas, a region noted for its varied crops, some of which are dependent on irrigation, while others are suited for dryland farming of this semiarid region. This area is characterized by its fertile soils, even topography, access to rail communications, and proximity to the Arkansas River.

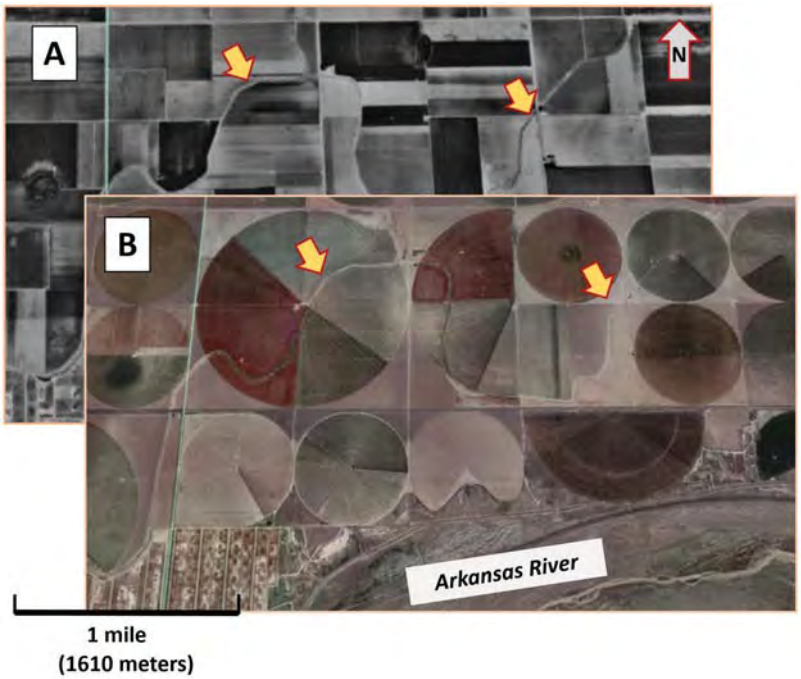


FIGURE 17.14 Irrigation systems near Holcomb, Finney County, Kansas, from Google Earth Pro. (A) September 1991; (B) June 2014.

Here we also see snapshots of a current center-pivot irrigation system (B), superimposed over the footprint of the previous flood irrigation system. Recall that the center-pivot system relies on groundwater, whereas the earlier flood irrigation system used surface water redirected from the Arkansas River, channeled along the canal (marked by the yellow arrows) originating from the Arkansas River (just off the lower left-hand corner of the maps).

Beginning in the 1880s, a growing population of pioneering farmers sought to stabilize the agricultural scene by constructing irrigation systems redirecting water from the nearby Arkansas River to adjacent lands (Hanks, 2014). Farmers upstream (to the west, in Colorado) also used the Arkansas River for irrigation water, so downstream farms in western Kansas soon experienced reduced irrigation capabilities. To expand local irrigation efforts, entrepreneurs in the downstream sections, near Garden City, began to use pumps and windmills to provide local sources of groundwater to support irrigation agriculture, and constructed ambitious irrigations systems.

The two snapshots in Figure 17.14 show the same landscape at different dates. The 1991 image (A) provides a black-and-white image of the agricultural landscape west of Holcomb, Kansas, and just north of the Arkansas River (flowing west to east). Here, we see the gridded field boundaries, and the irrigation canal that carries water channeled from the Arkansas River (just upstream, off the lower left edge of the image). The canal, as mentioned above, originates from the days of early local irrigation projects, following local terrain to minimize costs. (The red/yellow arrows mark the path of the channel as it follows local terrain.) Its function is to distribute irrigation water, using siphons to distribute water to field furrows (Figure 17.12). These fields are also irrigated by pumps (not visible on our imagery) powered by natural gas. The lower right of Figure 17.14 shows the same area, but for June 2014, now with center-pivot irrigation system. Center-pivot irrigation was introduced to the broader region as early as the 1970s and was used at this specific site beginning in about 2002.

17.9 CROP CALENDAR

A *crop calendar* describes the cycle of crops grown during the agricultural year, in harmony with regional climate, local practices and landscapes, and economic incentives. Each farmer adopts a specific sequence of preparing the field, planting, and harvesting to minimize labor and risk for the crop, and to maximize agricultural production. Although each farmer decides the specifics for a given farm, over broad regions, farmers as a group tend to follow a common sequence of agricultural activities for their region. As an example, the crop calendar for principal crops of western Nebraska can be broadly sketched as is depicted in Figure 17.15.

Analysts who examine images of agricultural landscapes apply knowledge of a region's agricultural calendar to understand the meaning of the patterns they see on imagery. Over decades, crop calendars may change as new crops and crop varieties are introduced, new agricultural technologies are installed, and weather and climate vary. Thus, at any given time, the landscape shows a variety of crops, each at different growth stages, due to varied planting dates as determined by the local agricultural calendar.

Sequential coverage provided by Earth observation satellites permits monitoring of agricultural systems over time, allowing analysts to apply knowledge of local crop calendars to identify cropping patterns. For example, the winter wheat crop may be character-

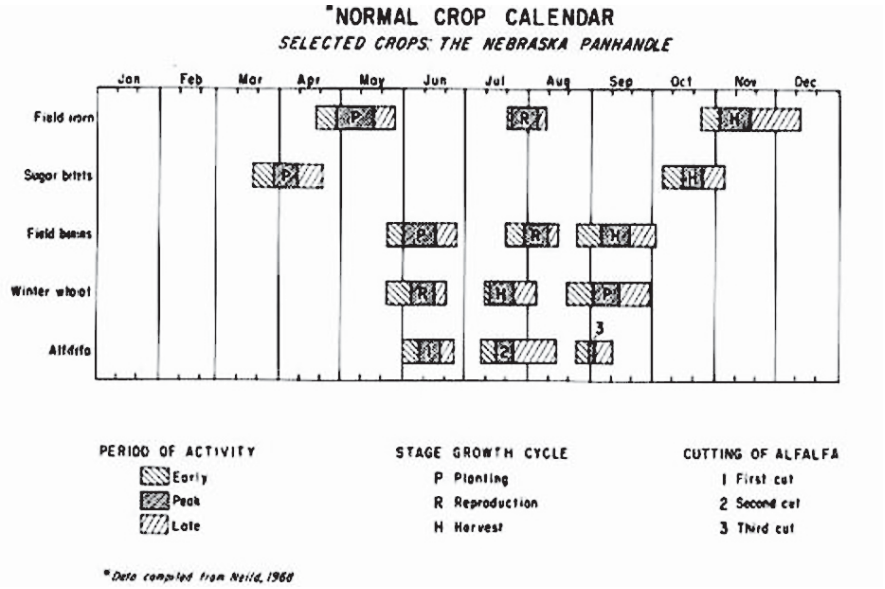


FIGURE 17.15 Local agricultural calendar for western Nebraska, 1968. From the Center for Advanced Land Management Information Technologies. Used by permission.

ized by open fields in September and October, and by a mixture of soil and newly emerged vegetation in mid-to-late autumn. The crop is dormant during the winter months; then, during spring and early summer, crop cover increases, completely masking the soil by late May or early June. As the crop matures, spectral evidence of senescence signals the approach of harvest in late June or early July.

In contrast, the corn crop (**Figure 17.15**) is typically not planted until late spring, emerging in early summer, attaining complete coverage until June, and maturing in August. Thus, each crop displays a characteristic behavior over time, permitting tracking progress of crop growth and (in many cases) recognition of its identity.

Note that **Figure 17.15** presents the crop calendar for a midlatitude region where there is a small range of crops, with planting and harvest cycles that follow a rather simple structure. In other regions of the world, climate, soils, and terrain dictate forms of agriculture based on a wider variety of crops, intercropping, and varied planting and harvesting cycles. For comparison, **Figure 17.16** presents a crop calendar for the Malaprabha Basin in southern India, which presents a much more intricate sequence of crops. The difference illustrates the challenge of relying on broad-scale remote sensing to assess such complex environments. Todoroff and Kemp (2016) specifically examined the effectiveness of remote sensing in a tropical context (mainly agriculture), noting the difficulties presented by the prevalence of small holdings.

Knowledge of the local crop calendar provides an analyst with the ability to examine the patchwork of local crops to identify differing growth stages and permits selection of a single date that will provide maximum contrast between two crops. In the instance of discrimination between corn and wheat, for example, selection of a date in late spring should show complete vegetative cover in wheat fields, whereas cornfields show only newly emerged plants against a background of bare soil.

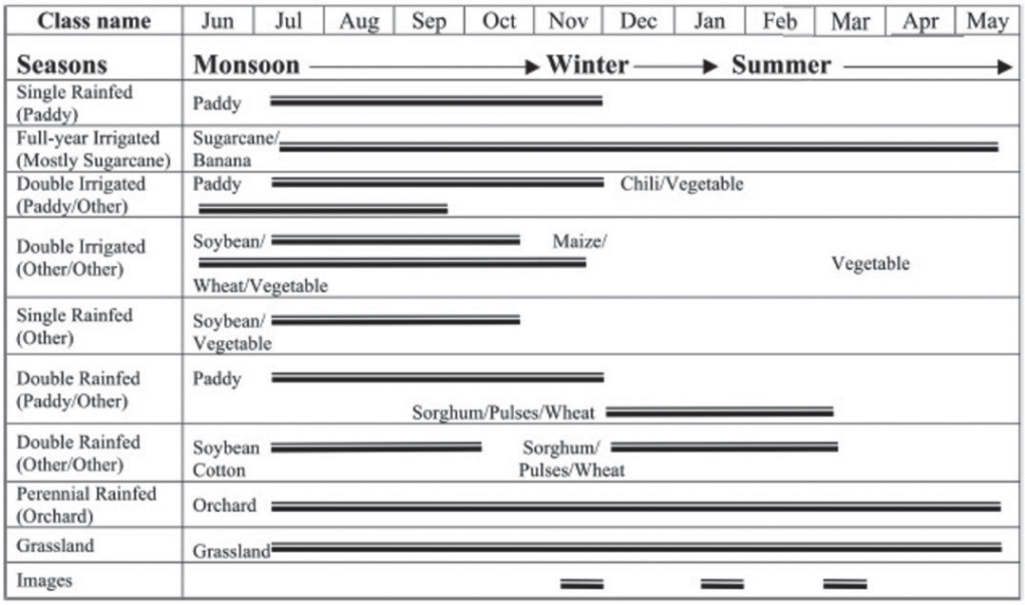


FIGURE 17.16 Crop calendar for cropping systems in the Malaprabha Basin of southern India, illustrating the complexity of crops and their temporal patterns relative to typical midlatitude agricultural systems. From Heller et al. (2012). Copyright © 2012 American Society for Photogrammetry and Remote Sensing. Used by permission.

Harvard Marsh, Nebraska

Figure 17.17 offers a practical example of how the local crop calendar appears on the landscape. Here we see a CIR aerial photograph of an agricultural landscape, Harvard Marsh, in south central Nebraska, as observed in mid-August. In this image, red is indicative of vegetation (mainly corn), due to relatively high reflectance in the NIR. The CIR film eliminates shorter wavelengths from the image, highlighting living vegetation and creating dark, sharply defined, shadows. Clear, open, water is imaged as dark black, with sharp, well-defined edges. Winter wheat fields, planted in the autumn of the previous year, have been harvested early in summer (perhaps in June or July) and so are now clearly distinctive as harvested in midsummer, with open, bare, fields and some crop residue left from the harvest earlier in the summer.

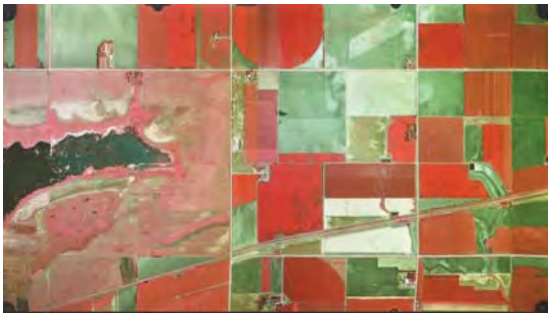


FIGURE 17.17 Harvard Marsh, Nebraska. Near-infrared imagery, August 1983. From U.S. Fish and Wildlife Service.

Much of the remainder of the image shows the corn crop, which requires ample moisture, supplied here by irrigation, needed in this region because of long periods of daylight, high transpiration, and dry weather as summer progresses. The corn crop is visible on this image as bright red fields, planted in spring, now approaching harvest time. Fields irrigated by center-pivot irrigation are visible (top center) as bright red, with generally even textures on the aerial photograph. Elsewhere, the corn crop has been irrigated by ditch irrigation, channeling the water flow in open channels, confined by field boundaries.

Figure 17.18 depicts the same scene with annotations to indicate notable features of this landscape scale:

1. Farmstead (likely well established over many years), with equipment sheds, and forest cover serving as windbreak.
2. Center-pivot irrigation. Maize (with variations in crop maturity) approaching harvest. Darker reds here and elsewhere suggest maize crop approaching harvest time.
3. Maize crop, ditch irrigation.
4. Unplanned release of irrigation water.
5. Winter wheat, after harvest earlier in the summer, to be planted for the next year, likely later in August or in September.
6. Grazing and water for cattle.
7. Paved highway; the culvert channels irrigation water below highway.
8. Wetlands. This section of the image is a protected wildlife refuge, which expands in spring with snowmelt and drainage. At this time in August, much of the wetland region has retreated to its smaller dimensions.
9. Natural channel.

17.10 CROP DAMAGE

Aerial imagery can be a significant asset for observing the effects of disease, weather damage, and insect infestation. *Lodging* is a common agricultural feature—a severe form of storm damage that occurs when strong winds and hail damage mature crops (see Figure 17.19, right). Lodging damages the stems of mature wheat, for example, as they are beaten down by heavy rain, hail, or winds. Mature crops are especially vulnerable

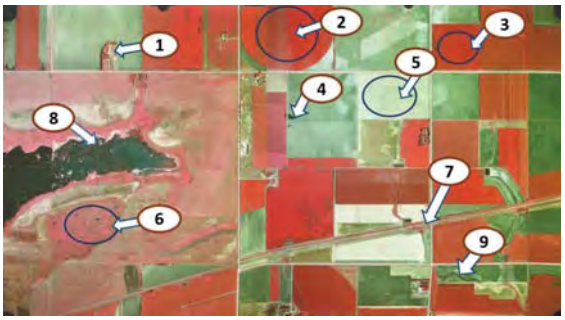


FIGURE 17.18 Harvard Marsh, Nebraska. Near-infrared radiation photography, August 1983. From U.S. Fish and Wildlife Service. Annotated to illustrate topics discussed in text. See the text for labels and key.

FIGURE 17.19 Storm-damaged crops. Left: Hail-damaged corn crop, Kansas. Right: Wind-damaged (lodged) wheat, Virginia.

to lodging; as crops approach maturity, the plants are larger and the stems weaker. Even if the lodged plants survive to continue growth, the broken stems prevent harvesting. *Hail damage* to some crops completely shreds the leaves, physically destroying the plant beyond hope of recovery (Figure 17.19, left). Yet some plants can survive hail damage, especially if they are young enough, because smaller, more compact leaves and stems are strong in comparison with those of the mature plant.

Aerial photography and photointerpretation are especially important tools for assessing crop status. Although farmers and agricultural analysts are familiar with crops planted in their own regions, it is difficult to assess the nature and status of crops over much larger regions at a given time because of the regional differences in crops, agricultural practices, climate, and terrain. The spatial perspective and the use of the nonvisible spectrum offered by aerial photography provide the agricultural community with a valuable means of understanding the crop status.

17.11 CONSERVATION TILLAGE

Tillage is the agricultural practice, dating from antiquity, of disturbing the soil surface at the time of planting in order to prepare the seedbed, disturb weeds, aerate the soil, accelerate the drying of wet soils, and, at high latitudes, promote springtime warming of soil. Tillage was practiced, initially with hand tools, later with wooden plows and draft animals, as early as 100 BCE in China, and 643 CE in Europe. Later, by 1836, in the United States, development of the John Deere plow provided farmers the ability to precisely adjust depth, width, and incline for plowing (Schmitz and Moss, 2015).

By the early 1900s, industrialization introduced steam-powered and then gasoline-powered tractors. Mechanization led to larger farm sizes and increased agricultural output. As mechanization increased, the nature and scope of tillage practices changed as it became possible to till larger areas, steeper slopes, and sod too tough to plow prior to mechanization (Figure 17.20). By the 1940s, increased awareness of the detrimental aspects of tillage (Faulkner, 1943), combined with the availability of herbicides, led to alternative practices to minimize adverse aspects of tillage, including increased erosion and loss of nutrients.

FIGURE 17.20 Example of *intensive tillage,* less than 30% residue cover remaining in the field after planting. Iowa, May 2017. Photograph by Michael Graham. Used by permission.

In this context, some specialists recommended *conservation tillage* practices to minimize cultivation and retain crop debris from the previous season's crop, reduce the use of herbicides, retain soil moisture, protect against erosion, and retain organic matter within the soil (Figure 17.21). Although conservation tillage can present disadvantages, its use has been encouraged to improve sustainability and soil management relative to more conventional practices.

Separation of *no-till* fields from *conventional tillage* practices (turning the soil surface to bury organic debris from the previous year's crop) provides the ability to assess broad-scale erosion risk and threats to water quality by runoff of fertilizers, pesticides, and sediment transport. Field sampling procedures and roadside surveys provide only rough estimates. In this context, Landsat satellite imagery provides an effective strategy for tillage assessment.

Tillage intensity can be characterized by the fraction of soil surface covered by crop residue. The Conservation Technology Information Center (CTIC) defined the following categories of tillage, based on the crop residue cover on the soil surface shortly after planting: *conventional tillage* has <15% residue cover; *reduced tillage* has 15–30% resi-

FIGURE 17.21 Example of *conservation tillage;* more than 30% residue cover remains after planting (in this example, strip-tillage). Iowa, May 2017. Photograph by Michael Graham. Used by permission.

due cover; *conservation tillage* has >30% residue cover (Claassen et al., 2018). Assessment of tillage intensity includes visual assessment, field measurements, agricultural census, and remote sensing.

The standard technique to characterize tillage intensity used by the USDA's National Resources Conservation Service is the *line-point transect method*. For a specific field, a 15- to 30-m line with 100 evenly spaced markers is positioned diagonally across crop rows. Field crews count points where the line intersects crop residue. The effectiveness of the line-point transect method depends on the length of the line, the number of points per line, and the skill of the observer. At least 500 points must be observed to determine corn residue cover to within 15% of the mean (Laflen et al., 1981). Significant modifications to the line-point transect method include use of measuring tapes, meter sticks, and wheels with pointers (Morrison et al., 1993).

Although aerial imagery, properly timed and collected at suitable resolutions, offers the capability to assess soil tillage status, the broad-scale surveys require areal coverage, revisit capabilities, and spectral channels that are, as a practical matter, available only through satellite observation.

By the 1980s, agronomists recognized the potential of satellite imagery for broad-scale survey of tillage practices, but it required several decades to define an effective strategy (outlined below) for use of satellite imagery. Here we outline the two main classes of satellite systems offering potential for routine broad-scale tillage assessment:

1. Optical remote sensing (visible, near-infrared, and mid-infrared imaging sensors)
2. Microwave remote sensing (SAR, Chapter 8)

Practitioners of remote sensing have proposed a variety of multispectral indices tailored to identify the presence of crop debris at the soil surface. As an example, the normalized difference tillage index (NDTI) (van Deventer et al., 1997) is defined as:

$$\mathrm{NDTI} = \frac{\mathrm{TM5} - \mathrm{TM7}}{\mathrm{TM5} + \mathrm{TM7}} \qquad \text{(EQ. 17.1)}$$

Thematic Mapper spectral bands are not optimized to separate crop residues from soil background, so many studies have been done to evaluate the effectiveness of such indices. Whereas the normalized difference vegetation index (NDVI) and related vegetation indices can exploit the sharp contrast in brightness of vegetation in the red and near-infrared regions, there is no such dramatic contrast between crop debris and the soil background. Moreover, the contrast that does exist decreases as crop debris weathers and decomposes, so the effectiveness of tillage indices varies regionally in part because of variations in the color and brightness of the soil background. A variety of other measures have been proposed, some of which use other sensors; Serbin et al. (2009), Sullivan et al. (2008a), Arsenault and Bonn (2005), and Daughtry et al. (2005) review many of these indices.

The effectiveness of tillage assessment depends not only on the tillage index, but also on the timing of the acquisition of the imagery (Zheng et al., 2012) Accurate assessment of tillage status requires a sequence of images acquired late in the spring, when crop residue is still present, but before the new crop has emerged to mask residue cover (Figure 17.22). Zheng et al. (2012) demonstrated that a single "snapshot" Landsat image is inad-

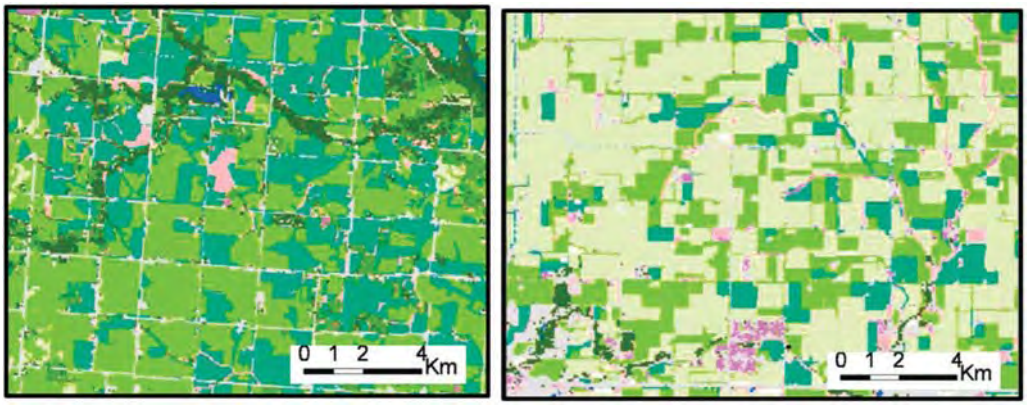


FIGURE 17.22 Remote sensing of tillage status derived from Landsat 5 and Landsat 7. Left: Marshall County, Iowa. Right: Champaign County, Illinois. Darker shades of green symbolize fields characterized by conservation tillage. Lighter shades of green symbolize fields characterized by conventional tillage. Tillage status based on methods and validation as reported by Zheng et al. (2013a). From Baojuan Zheng. Used by permission.

equate to record tillage status because of the wide variety of planting dates encountered within even a small area. A single image will record only a fraction of the fields needed to assess tillage status for a region. Thus, assessment of tillage status requires a sequence of images acquired at the beginning of the growing season. This task can be challenging because of the difficulty of acquiring several cloud-free images within the rather narrow temporal window available to assess tillage status.

At the time of this writing, the current availability of Landsat 8, Sentinel-2a, and Sentinel-2b imagery (Chapter 7) should, with their compatible orbits, similar spectral channels, and coordinated revisit intervals, improve practical application of the NDTI tillage assessment strategy, as opportunities for acquiring cloud-free sequences will increase. Such developments will greatly improve our ability to assess tillage in regions where agriculture is characterized by large fields, monoculture (single crops within fields), and favorable cloud cover (such as North America, Central Asia, China, Argentina, South Africa, and Australia). In other regions, especially in tropical zones, cloud cover presents a serious obstacle to acquisition of the sequential imagery mentioned above; field sizes are often smaller or irregularly shaped, and crops can be planted at intervals throughout the growing season and intermixed within fields.

17.12 USDA CROPLAND DATA LAYER

In 1997, the USDA began systematic preparation of the *Cropland Data Layer (CDL)*, which includes site-specific inventories of U.S. croplands for the 48 contiguous states (Johnson and Mueller, 2010). The USDA's National Agricultural Statistics Service (NASS) has responsibility for preparing the survey each year, with each year's results released in January of the following year. USDA's analysts generate the CDL through analysis of moderate-resolution satellite imagery, supported by field validation and organized as a

raster at 30-m detail (note that earlier versions were at 56-m resolution), using the Albers Equal-Area Conic projection (Figure 17.23).

Analysts identify croplands on a field-by-field basis, based on local agricultural calendars, satellite imagery, and field verification, supplemented by comparable imagery from similar systems when cloud cover masks the use of primary imagery. The CDL program began in 2006 and initially focused on five states in the central United States (Iowa, Arkansas, Illinois, Indiana, and North Dakota). Then, as analytical techniques were refined and validated, its scope was expanded. By about 2010, all 48 contiguous U.S. states were included.

The CDL's ability to provide national, regional, and state-level maps reveals the overall pattern of agricultural activities, providing valuable insight into agricultural systems. At finer scales, such as the county-level detail of the central plains, field sizes are large enough that cropping patterns and, over time, changes in cropping patterns, are distinct. The basic CDL product is displayed online as the *NASS CropScape* tool *(https://nassgeodata.gmu.edu/CropScape),* which provides users with the ability to roam, magnify, and select data for a specific calendar year.

17.13 BIOPHYSICAL DIMENSIONS OF AGRICULTURAL REMOTE SENSING

The discussion on the biophysical measure of vegetation cover in Chapter 16 is obviously relevant for agriculture. The three common biophysical measures discussed previously (i.e., vegetation fraction, leaf area index, and net primary production) are widely used to monitor crop cover, condition, and productivity. The vegetation indices introduced in Chapter 16 (simple ratio, NDVI, moisture stress index, normalized difference moisture index), among others, are widely used for agricultural monitoring. Numerous other indices are useful for agriculture but were not discussed in Chapter 16. We discuss two additional indices below which incorporate information about the soil.

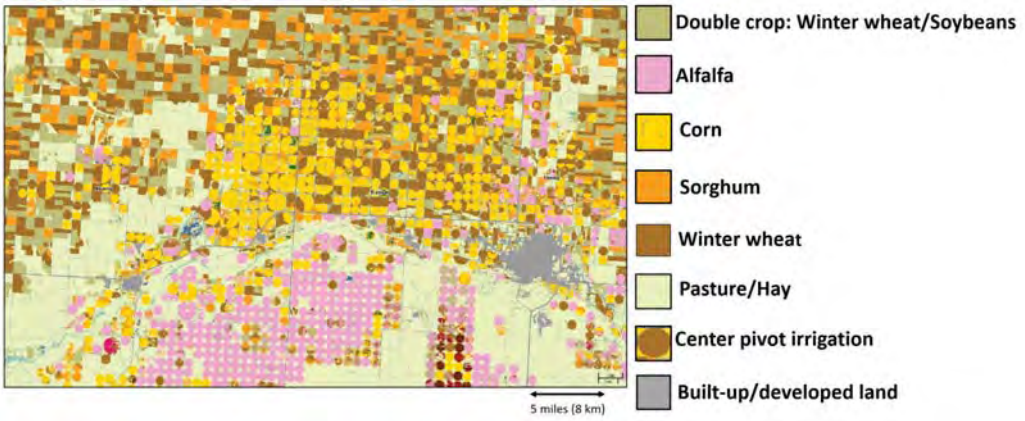


FIGURE 17.23 Segment of USDA Cropland Data Layer, Garden City, Finney County, Kansas, 2014. This region is characterized by a variety of crops and agricultural practices, including center-pivot irrigation, ditch irrigation, and dryland farming. From USDA.

When assessing large areas, the pixel resolution of satellites often results in a mix of vegetation reflectance with that of the soil surface. Furthermore, in agricultural scenes, reflection from individual plants or rows of plants is closely intermingled with bare soil between plants and between rows of plants, such that reflectance is mixed, even at the finest resolutions. Such mixing is especially important after plants have emerged because large proportions of soil are exposed to the sensor.

We can examine the relationship between vegetation and soil, based on their reflectance in red and NIR wavelengths (Figure 17.24). Dry soils tend to be bright in both spectral regions and appear at the high end of the line (C); wet soils tend to have dark colors and are positioned at the low end (B). Richardson and Wiegand (1977) defined this relationship, known as the *soil brightness line,* and recognized that the spectral response of living vegetation will always have a consistent relationship to the line. Soils typically have high or modest responses in the red and infrared regions, whereas living vegetation must display low values in the red (due to the absorption spectra of chlorophyll) and high values in the near infrared (due to infrared [IR] brightness of mesophyll tissue).

Points representing a "pure" vegetation response will be positioned in the upper left of Figure 17.24, where values on the red axis are low and those on the IR axis are high. Furthermore, Richardson and Wiegand (1977) defined an index to portray the relative magnitudes of soil background and vegetative cover to a given spectral response. Point *X* typifies a "pure" vegetation pixel, with a spectral response determined by vegetation alone and with no spectral contribution from the soil surface. In contrast, point *Y* typifies a response from a partially vegetated pixel; it is brighter in the red and darker in the near infrared than is *X*. Richardson and Wiegand quantified this difference by defining the *perpendicular vegetation index* (PVI) as a measure of the distance of a pixel (in spectral data space) from the soil brightness line. The PVI is simply a Euclidean distance measure similar to those discussed in earlier chapters:

$$\mathrm{PVI} = \sqrt{\left(S_R - V_R\right)_{IR}^2 + \left(S_{IR} - V_{IR}\right)^2} \qquad \text{(EQ. 17.2)}$$

where *S* is the soil reflectance, *V* is the vegetation reflectance, subscript *R* represents red radiation, and subscript *IR* represents near-infrared radiation.

In practice, the analyst must identify pixels known to be composed of bare soil to identify the local soil brightness line and pixels known to be fully covered by vegetation to identify the local value for full vegetative cover (point *X* in Figure 17.24). Then, inter-

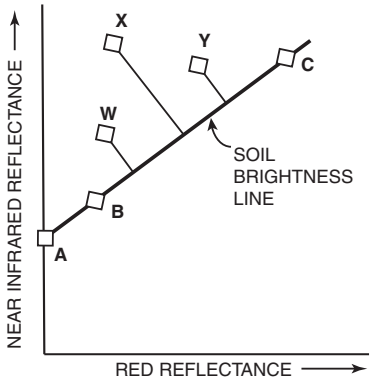


FIGURE 17.24 Perpendicular vegetation index. Based on Richardson and Weigand (1977, p. 1547). Copyright © 1977 the American Society for Photogrammetry and Remote Sensing. Used by permission.

mediate values of PVI indicate the contributions of soil and vegetation to the spectral response. The soil line has been used to indicate soils that are free of vegetation at the time that the image was acquired and to separate partially vegetated pixels from those that have complete vegetative cover. The position of any specific pixel on the soil line is likely determined by the nature of the soil surface, to include soil properties (physical and chemical), roughness, and moisture status (Baret, Jacquemond, and Hancoq, 1993). Because such properties can be transient, it is not likely that a single soil line can precisely represent all soils.

Baret et al. (1993) investigated the soil brightness line from both experimental and theoretical perspectives. They concluded that the theoretical basis for the soil brightness line is sound, although the effects of variations in soil moisture and surface roughness are still not well understood. Baret et al.'s experimental data revealed that it is not feasible to define a single, universally applicable soil brightness line, as local variations in soil types lead to spectral variations. However, these variations were found to be minor in the red and infrared regions. They concluded that the use of a single soil brightness line is a reasonable approximation, especially in the context of analysis of coarse-resolution satellite data. Nanni and Dematte (2006) and Ladoni et al. (2009) provide reviews of research on these topics.

Soil-Adjusted Vegetation Index (SAVI). In regions where vegetative cover is sparse enough to expose soil surfaces (e.g., <40%), reflectance of red and near-infrared light can affect NDVI values. This effect can be problematic when the goal is to compare differing surfaces, because altered brightness can distort NDVI values. Thus, SAVI was specifically developed to adjust for the effects of soil brightness when vegetative cover is sparse (Huete, 1988):

$$\text{SAVI} = \frac{(\text{NIR} - \text{RED})}{(\text{NIR} + \text{RED} + L)} \times (1 + L) \qquad \text{(EQ. 17.3)}$$

where NIR is the reflectance value of the near-infrared band, RED is reflectance of the red band, and L is the soil brightness correction factor. The value of L varies by the amount or cover of green vegetation: in highly vegetated situations, L = 0, and SAVI = NDVI. In areas without green vegetation, L = 1. Generally, an L = 0.5 works well in most situations and is the default value.

17.14 PRECISION AGRICULTURE

Previous sections have introduced aerial observations of terrain and agricultural landscapes, offering opportunities to examine important regions at varied scales and perspectives. Here, we introduce unmanned aerial systems (UAS), often known as drones—basically, aircraft operated without pilots or passengers, guided by experienced controllers at ground level, within range of the instrument. In the United States, the Federal Aviation Administration (FAA) publishes requirements for drone operations, requiring that pilots pass an examination, register aircraft, and carefully follow restrictions. Examples of applications include aerial photography, agriculture, forestry, construction, monitoring traffic, and forest fires. For the most part, we should think of drone imagery as characterized by rather small or modest features.

Drones are widely used in the United States, Canada, and most other countries to varying degrees, most specifically for optimizing agricultural operations, improving crop production, and monitoring crop status. Drones can therefore empower farmers to examine fields without direct on-site observation, enhancing crop yields and efficiency. Furthermore, aerial views acquired by drones can monitor agricultural production, such as irrigation, diseases, soil variation, pests, fungal infestations, and related characteristics.

In this context, we should think of UAS imagery as tailored to observe small or modest-sized features. So, whereas aerial and satellite systems may observe very broad regions, perhaps many miles in size, drone operations focus on observing small fields. Such fields enable farmers to observe fine detail, and to frequently monitor crops. Drones recording detail (sometimes known as "agricultural drones") enable farmers to examine locations without on-site investigation, for example, storm damage, local flooding, and infestation. Because healthy plants often reflect light differently than do less vibrant plants, some forms of light might enable farmers to recognize nutrient deficiencies and perhaps to detect diseases.

Furthermore, we note that farmers can tailor selected crops to isolate specific rows or regions to meet local requirements. As a practical matter, UASs (drones) can apply low-altitude aerial imagery to monitor the progress of small crops such as corn, soybeans, and varied forms of small grains. Drones can often be effective in observing crop development, especially when crops such as soybeans, corn, or small-grain crops are considered. In contrast, other crops, especially tobacco, can be difficult to monitor because of the exposed soil present between crop rows. In this context, we note the effectiveness of Austin Hayes's drone research, which focuses on the effects of tobacco canopies on masking vegetation canopies (**Figure 17.25**). Tobacco crops often present exposed soil surfaces such that generally low-flight altitudes of drones can expose soils dominating aerial views.

Figure 17.25 presents aerial views of crops, specifically the right-hand view of graphics in **Figure 17.25**. Here, the image in the lower right corner presents a detailed

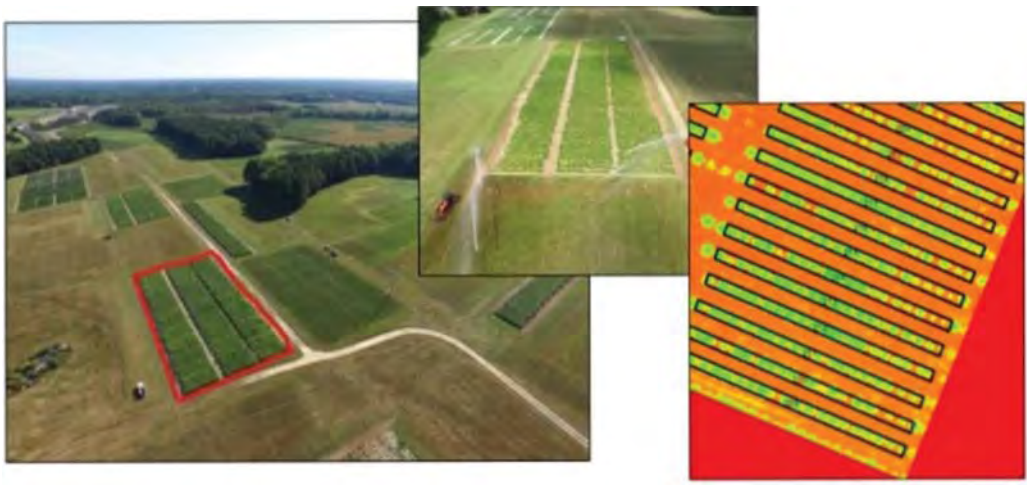


FIGURE 17.25 Unmanned aerial vehicle systems in support of agricultural management. Note the within-row detail and variation made possible with the high spatial resolution of this system. Photographs by Austin Hayes. Used by permission.

illustration of the agricultural landscape, outlined to highlight the area marked to identify precise areas to be analyzed, separating key targets (inside the black stripes) from the broader regions of the range background of the larger region.

17.15 REMOTE SENSING OF PLANT PATHOLOGY

The Disease Triangle

Plant pathology is the study of plant diseases and their occurrence. *Plant pathologists* study the biochemistry and biology of pathogens and the plants they infect to understand the life cycle of the pathogens and the means by which they infect hosts. Plant pathologists counter the effects of pathogens by proposing management techniques, designing pesticides, and recommending countermeasures to minimize the effects of the disease and to prevent or slow its spread to other regions.

A few well-known examples can remind us of the significance of plant diseases. During 1846–1850, the Irish potato famine killed at least one million Irish citizens, caused immense social disruption, and created emigration that itself influenced social systems in the United States and elsewhere. In the United States, Dutch elm disease and the chestnut blight have destroyed trees that once constituted North America's most distinctive and beautiful trees. Specifically, the decline of the American chestnut tree changed the ecology of large regions of Appalachian forests. The tree once constituted perhaps one-fourth of the trees of the Appalachian forests and provided food for both humans and wildlife. Both tree species exist now only as isolated remnants. Often such catastrophes are closely connected to human behavior, especially through the introduction of plants and diseases from other regions, the tendency to depend on single crops, usually densely planted, and the uses of genetically uniform hybrids that may increase susceptibility to attacks by pathogens.

Infection depends on the co-occurrence of three interrelated factors (known as the *disease triangle*):

1. A pathogen (bacteria, viruses, fungi, mycoplasmas, nematodes)
2. A susceptible host plant (a plant's resistance to disease)
3. Favorable environmental conditions (including temperature, wind, solar illumination, and soil)

Pathogens spread through the environment by wind, water, and soil and by biologic transmission, including through insects, animals, and humans.

Remote sensing offers many ways to supplement and reinforce the efforts of the detailed, ground-based expertise of the plant pathologist. Applications of remote sensing employ the plant pathologist's understanding of the dynamics of an infestation in its spatial context at the landscape level and the use of ground-based surveys to understand the behavior of the pathogen at the level of the plant and the cell.

Remote sensing offers the ability to conduct systematic surveys to detect mortality associated with diseases or declines in vegetation health and to document spread from one region to another and the effectiveness of countermeasures. Specifically, remote sensing has been useful in the following:

- Detection of infestation; symptoms from aerial imagery
- Detection of host plants, identification of crop residue and volunteer growth that can harbor pathogens
- Phenology: understanding the onset of seasonal growth in hosts or vectors
- Assessment of environmental conditions favorable for the pathogen
- Spatial perspective on the occurrence and spread of the disease and monitoring of its spread

A few case studies can highlight some of the issues pertaining to the role of remote sensing in assessing and monitoring the occurrence of plant diseases.

The Corn Blight Watch Experiment, 1971

One of the earliest applications of remote sensing to a broad-scale infestation occurred during the 1971 Corn Blight Watch Experiment (see **Figures 17.26 and 17.27**). In previous years, corn breeders had introduced varieties of maize that had been optimized for hybridization but that were vulnerable to specific varieties of the fungus (*Bipolaris maydis*) that causes southern corn leaf blight (SCLB).

SCLB exists in nature, mainly as a minor disease of corn. SCLB produces elongated brown spots on leaves and stalks during the interval from mid-June to mid-October. If the disease attacks before the ears have formed, crop losses can be severe. The fungus resides in soil and in crop debris, remains after harvest, and survives the winter season to infest the crops of the succeeding year. SCLB thrives in warm, moist weather, so it spreads easily in the summer weather of central North America. Its occurrence is very difficult to control.

FIGURE 17.26 Corn Blight Watch Experiment flight lines. Initiated in April 1971, this experiment evaluated the effectiveness of remote sensing and focused on the following: evaluating the use of advanced remote sensing techniques and concepts; detecting development and spread of corn blight during the growing season in the Corn Belt region; amplifying information acquired by ground visits to better assess the current blight status and probable impacts on crop production; estimating the applicability of these techniques to similar future situations. Orange stripes signify zones for photographic coverage flown at 60,000 ft, green stripes include a multispectral scanner at 5,000 ft. From NASA (1974).

Once the vulnerable hybrid became widely used, SCLB rose from a minor nuisance to a major threat to U.S. agriculture. The 1970 corn crop had declined by 15% of the previous crop, creating havoc in commodities markets and creating political pressures for agricultural officials. By the early months of 1971, seeds carrying the genetic weakness had already been widely distributed to farmers throughout the continent. As the 1971 growing season began, agricultural scientists became increasingly concerned about the vulnerability of the North American corn crop. Because of the widespread use of the vulnerable seed and the dense occurrence of maize cultivation throughout central North America, the right weather conditions could lead to rapid spread of the blight and destruction of the nation's corn production, with ominous implications for the agricultural economy and a major portion of the nation's food production.

The Corn Blight Watch Experiment was based on the use of remotely sensed data to monitor the occurrence and spread of SCLB during the 1971 growing season. A series of flight lines was designed to assess the spread of the infection throughout the growing season. Biweekly flights collected CIR photography, as well as multispectral scanner data, for the principal corn-growing regions in the United States and Canada. Ground data were collected in coordination with the collection of aerial imagery. The sampling design allowed ground data to be extended to broad-scale estimates for the principal corn-producing regions of the United States and Canada.

Photointerpretation estimates varied in accuracy when compared against the data collected at ground level. Photointerpretation was effective in distinguishing between healthy or slight levels of infection for more severely infected fields, but was not effective in assessing slight or mild levels of infection. Early in the growing season, there was a large discrepancy between estimates of crop damage from ground-based observations and those derived from photointerpretation (Figure 17.27). Some difficulties were experienced in assessing the effects of blight when other causes influenced the health of the corn crop.

Significantly, the Corn Blight Watch Experiment is a clear example of a broad-scale attempt to use remote sensing as a tool to contribute to understanding the spatial behavior of a plant pathogen. As globalization and concerns about bioterrorism raise awareness about the possible impacts of introduced pathogens, this experiment from several decades past may merit examination.

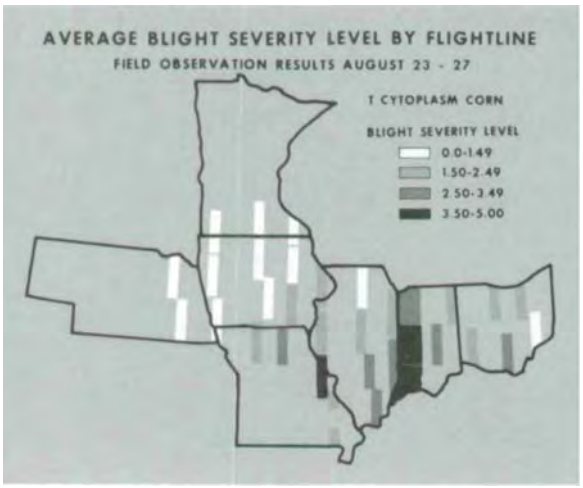


FIGURE 17.27 Results of the 1971 Corn Blight Watch Experiment at the conclusion of its seasonal program, illustrating the effectiveness of remote sensing's capabilities. Dark panels in this image signify the presence of high blight severity. The Corn Blight Watch Experiment was the first to apply a sound statistical design for broad-scale remote sensing. Here, the accuracy of corn identification by remote sensing exceeded 90% throughout the experiment. From NASA (1974).

17.16 SUMMARY

Given the pressures on global food security, with increasing population and changing environmental stressors, the techniques outlined in this chapter (as well as Chapter 16) are an integral component of decision making for the well-being of humankind and the sustainability of global ecosystems. The ability to examine agriculture patterns using vegetation indices, combined with the synoptic view and repetitive coverage of satellite sensors, provides an opportunity to survey agricultural patterns and inform decision making at multiple scales, from the individual farm to national and global policy structures. If the information gathered by remote sensing systems can be better integrated into decision making and the transportation and distribution systems that are so important for agricultural production, there is a prospect for improvements in the effectiveness of food production.

SOME TEACHING AND LEARNING RESOURCES

- Additional teaching and learning resources on this topic can be viewed in the video series "Science for a Hungry World":
  www.nasa.gov/feature/goddard/science-for-a-hungry-world

REVIEW QUESTIONS

1. Remote sensing to monitor crop development is much more difficult than it might appear initially. List some of the practical problems you might encounter as you plan an experiment to use satellite data to study the development of the winter wheat crop in western Kansas (or another crop region specified by your instructor).
2. Connecting the concepts of this chapter to those presented in Chapter 16, list some of the ways in which monthly phenologic information might be useful for the study of cropped agricultural land.
3. Why are multispectral satellites particularly well suited for studying agriculture patterns?
4. What are some ways that remote sensing could be used to study the impacts of drought on agriculture?

REFERENCES

Arsenault, E., and F. Bonn. 2005. Evaluation of Soil Erosion Protective Cover by Crop Residues Using Vegetation Indices and Spectral Mixture Analysis of Multispectral and Hyperspectral Data. *Catena,* Vol. 62, pp. 157–172.

Atzberger, C. 2013. Advances in Remote Sensing of Agriculture: Context Description, Existing Operational Monitoring Systems and Major Informational Needs. *Remote Sensing,* Vol. 5, pp. 949–981.

Baret, F., S. Jacquemond, and J. F. Hancoq. 1993. The Soil Line Concept in Remote Sensing. *Remote Sensing Reviews,* Vol. 7, pp. 65–82.

Bauer, M. 1975. The Role of Remote Sensing in Determining the Distribution and Yield of Crops. *Advances in Agronomy,* Vol. 27, pp. 271–304.

Boryan, C., Z. Yang, R. Mueller, and M. Craig. 2011. Monitoring US Agriculture: The US Department of Agriculture, National Agricultural Statistics Service, Cropland Data Layer Program, *Geocarto International,* Vol. 26, No. 5, pp. 341–358.

Brandt, C. S. 1966. Agricultural Burning. *Journal of Air Pollution Control Association,* Vol. 16, No. 2, pp. 85–86.

Brown, M. E. 2015. Satellite Remote Sensing in Agriculture and Food Security Assessment. *Procedia Environmental Sciences,* Vol. 29, p. 307.

Brown, M. E., K. Grace, G. Shively, K. B. Johnson, and M. Carroll. 2014. Using Satellite Remote Sensing and Household Survey Data to Assess Human Health and Nutrition Response to Environmental Change. *Population and Environment,* Vol. 36, No. 2, pp. 48–72.

Claassen, R., M. Bowman, J. McFadden, D. Smith, and S. Wallander. 2018, September. *Tillage Intensity and Conservation Cropping in the United States* (Economic Information Bulletin, No. 197). Washington, DC: U.S. Department of Agriculture, Economic Research Service.

Cohen, W. B. 1991. Response of Vegetation Indices to Changes in Three Measures of Leaf Water Stress. *Photogrammetric Engineering and Remote Sensing,* Vol. 57, pp. 195–202.

Collins, W. 1978. Remote Sensing of Crop Type and Maturity. *Photogrammetric Engineering and Remote Sensing,* Vol. 44, pp. 43–55.

Curran, P. 1980. Multispectral Remote Sensing of Vegetation Amount. *Progress in Physical Geography,* Vol. 4, pp. 315–341.

Daughtry, C. S. T., E. R. Hunt, P. C. Doraiswamy, and J. E. McMurtrey. 2005. Remote Sensing the Spatial Distribution of Crop Residues. *Agronomy Journal,* Vol. 97, pp. 864–871.

Faulkner, E. H. 1943. *Plowman's Folly.* Norman, OK: University of Oklahoma Press.

Gibson, G. R., J. B. Campbell, and R. H. Wynne. 2012. Three Decades of War and Food Insecurity in Iraq. *Photogrammetric Engineering and Remote Sensing,* Vol. 78, pp. 885–895.

Haboudane, D., J. R. Miller, N. Tremblay, P. J. Zarco-Tejada, and L. Dextraze. 2002. Integrated Narrow-Band Vegetation Indices for Prediction of Crop Chlorophyll Content for Application to Precision Agriculture. *Remote Sensing of Environment,* Vol. 81, pp. 416–426.

Han, W., Z. L. Yang, L. Di, and R. Mueller, R. 2012. CropScape: A Web Service Based Application for Exploring and Disseminating US Conterminous Geospatial Cropland Data Products for Decision Support. *Computers and Electronics in Agriculture,* Vol. 84, pp. 111–123.

Han, W., Z. L. Yang, A. Di, S. Yagci, and S. Han. 2014. Making Cropland Data Layer Data Accessible and Actionable in GIS Education. *Journal of Geography,* Vol. 113, No. 3, pp. 129–238.

Hanks, K. 2014, July 19. Relentless quest for a precious resource. *The Hutchinson News.* Available at *https://hutchnews.com/story/news/local/kansas-agland/2014/07/19/relentless-quest-for-precious-resource/20711030007.*

Hatfield, J. L., and J. H. Prueger. 2010 Value of Using Different Vegetative Indices to Quantify Agricultural Crop Characteristics at Different Growth Stages under Varying Management Practices. *Remote Sensing,* Vol. 2, pp. 562–578.

Hayes, A. C. 2019. *Evaluating the Potential of Aerial Remote Sensing in Flue-Cured Tobacco.* Unpublished masters thesis, Department of Crop and Soil Environmental Science, Virginia Polytechnic Institute and State University.

Heller, E., J. M. Sharachchnadra Lele, M. Kalacska, S. Badiger, R. Sengupta, and N. Ramakutty. 2012. Mapping Crop Types, Irrigated Areas, and Cropping Intensities in Heterogeneous Landscapes of Southern India Using Multi-Temporal Medium-Resolution Imagery: Implications for Assessing Water Use in Agriculture. *Photogrammetric Engineering and Remote Sensing,* Vol. 78, pp. 815–827.

Huete, A. R. 1988. A Soil-Adjusted Vegetation Index (SAVI). *Remote Sensing of Environment,* Vol. 25, pp. 295–309.

Huete, A. R., H. Q. Liu, L. Batchily, and W. van Leewen. 1997. A Comparison of Vegetation Indices over A Global Set of TM Images for EOS-MODIS. *Remote Sensing of Environment,* Vol. 59, pp. 440–451.

Johnson, D. 2014. An Assessment of Pre- and Within-Season Remotely Sensed Variables for Forecasting Corn and Soybean Yields in the United States. *Remote Sensing of Environment,* Vol. 141, pp. 116–128.

Johnson, D., and R. Mueller. 2010. The 2009 Cropland Data Layer. *Photogrammetric Engineering and Remote Sensing,* Vol. 76, pp. 1201–1205.

Johnson, D. M. 2013. A 2010 Map Estimate of Annually Tilled Cropland within the Conterminous U.S.A. *Agricultural Systems,* Vol. 114, pp. 95–105.

Ladoni, M., H. A. Bahrami, S. K. Alvaipanah, and A. A. Norouzi. 2009. Estimating Soil Organic Carbon from Soil Reflectance: A Review. *Precision Agriculture,* Vol. 11, pp. 82–99.

Laflen, J. M., M. Amemiya, and E. A. Hintz. 1981. Measuring Crop Residue Cover. *Journal of Soil and Water Conservation*, Vol. 36, No. 6, pp. 341–343.

Lam-Dao, N., P. Hoang-Phi, J. Huth, and P. Cao-Va. 2017. Estimation of the Rice Yield in the Mekong Delta Using SAR Dual Polarisation Data. 2017. *International Journal of Applied Earth Observation and Geoinformation,* Vol. 57, pp. 75–85.

Landsat Legacy Project Team. 2017. *Landsat's Enduring Legacy: Pioneering Global Land Observations from Space*. Bethesda, MD: American Society for Photogrammetry and Remote Sensing.

Liu, W., S. Gopal, and C. E. Woodcock. 2004. Uncertainty and Confidence in Land Cover Classification Using a Hybrid Classifier Approach. *Photogrammetric Engineering and Remote Sensing,* Vol. 70, pp. 963–971.

Lobell, D. B. 2013. The Use of Satellite Data for Crop Yield Gap Analysis. *Field Crops Research,* Vol. 143, pp. 56–64.

Mazoyer, M., and L. Roudart. 2006. *A History of World Agriculture: From the Neolithic Age to the Current Crisis*. New York: New York University Press, 496 pp.

Milas, A. S., M. Romanko, P. Reil, T. Abeysinghe, and A. Marambe. 2018. The Importance of Leaf Area Index in Mapping Chlorophyll Content of Corn under Different Agricultural Treatments using UAV Imagery. *International Journal of Remote Sensing,* Vol. 39, pp. 5415–5431.

Morrison, J. E., C.-H. Huang, D. T. Lightle, and C. S. T. Daughtry. 1993. Residue Measurements Techniques. *Journal of Soil and Water Conservation*, Vol. 48, No. 6, pp. 478–483.

Moulin, S., A. Bondeau, and R. Delecolle. 1998. Combining Agricultural Crop Models and Satellite Observations: From Field to Regional Scales. *International Journal of Remote Sensing,* Vol. 19, pp. 1021–1036.

Myneni, R. B., C. D. Keeling, C. J. Tucker, G. Asrar, and R. R. Nemani. 1997. Increased Plant Growth in the Northern Latitudes from 1981–1991. *Nature,* Vol. 386, pp. 698–702.

Nanni, M. R., and J. A. M. Dematte. 2006. Spectral Reflectance Methodology in Comparison to Traditional Soil Analysis. *Soil Science Society of America Journal,* Vol. 67, pp. 258–267.

National Aeronautics and Space Administration (NASA). 1974. *Corn Blight Watch Experiment: Summary Report*. Washington, DC: Scientific and Technical Information Office, National Aeronautics and Space Administration, 15 pp.

Nellis, M. D., K. P. Price, and D. Rundquist. 2009. Remote Sensing of Cropland Agriculture. Chapter 26 in the *SAGE Handbook of Remote Sensing*. London: SAGE, pp. 368–380.

Ozdogan M., Y. Yang, G. Allez, and C. Cervantes. 2010. Remote Sensing of Irrigated Agriculture: Opportunities and Challenges. *Remote Sensing,* Vol. 2, pp. 2274–2304.

Pereira, J. M. C. 2003. Remote Sensing of Burned Areas in Tropical Savannas. *International Journal of Wildland Fire,* Vol. 23, pp. 259–270.

Peters, A. J., S. C. Griffin, A. Vina, and L. Ji. 2000. Use of Remotely Sensed Data for Assessing Crop Hail Damage. *Photogrammetric Engineering and Remote Sensing,* Vol. 66, pp. 1349–1355.

Pinter, P., J. Hatfield, J. Schepers, E. Barnes, S. Moran, C. Daughtry, and D. Upchurch. 2003. Remote Sensing for Crop Management. *Photogrammetric Engineering and Remote Sensing,* Vol. 69, pp. 647–664.

Prasad, A., L. Chai, R. Singh, and M. Kafatos. 2006. Crop Yield Estimation Model for Iowa Using Remote Sensing and Surface Parameters. *International Journal of Applied Earth Observation and Geoinformation,* Vol. 8, No. 1, pp. 26–33.

Reitsma, K. D., D. E. Clay, S. A. Clay, B. H. Dunn, and C. Reese. 2016. Does the U.S. Cropland Data Layer Provide an Accurate Benchmark for Land-Use Change Estimates? *Agronomy Journal,* Vol. 108, pp. 266–272.

Richardson, A. J., and C. L. Wiegand. 1977. Distinguishing Vegetation from Soil Background. *Photogrammetric Engineering and Remote Sensing,* Vol. 43, pp. 1541–1552.

Saho, Y. 2016. Characterizing Major Agricultural Land Changes in the Western Corn Belt. *ISPRS Journal of Photogrammetry and Remote Sensing,* Vol. 127, pp. 116–125.

Schmitz, A., and C. B. Moss. 2015. Mechanized Agriculture: Machine Adoption, Farm Size, and Labor Displacement. *AgBioForum,* Vol. 18, No. 3, pp. 278–296.

Serbin, G., C. S. T. Daughtry, E. R. Hunt, D. J. Brown, and G. W. McCarty. 2009. Effect of Soil Spectral Properties on Remote Sensing of Crop Residue Cover. *Soil Science Society of America Journal,* Vol. 73, pp. 1545–1558.

Shao, Y. 2016. Characterizing Major Agricultural Land Changes in the Western Corn Belt. *ISPRS Journal of Photogrammetry and Remote Sensing,* Vol. 127, pp. 116–125.

Son, N. T., C. F. Chen, C. R. Chen, L. Y. Chang, H. N. Duc, and L. D. Nguyen. 2013. Prediction of rice crop yield using MODIS EVI–LAI data in the Mekong Delta, Vietnam. *International Journal of Remote Sensing*, Vol. 34, No. 20, pp. 7275–7292.

Song, C. 2004. Cross-Sensor Calibration between IKONOS and Landsat TM+. IEEE *Geoscience and Remote Sensing Letters,* Vol. 1, No. 4, pp. 272–276.

Sorensen, C. 1968. *A History of Irrigation in the Arkansas River Valley in Western Kansas. 1880–1910*. Unpublished masters thesis, University of Kansas, 186 pp. Available at *https://kuscholarworks.ku.edu/handle/1808/11088*.

Stern, A. J., P. C. Doraiswamy, and E. R. Hunt. 2012. Changes of Crop Rotation in Iowa Determined from the United States Department of Agriculture, National Agricultural Statistics Service Cropland Data Layer Product. *Journal of Applied Remote Sensing,* Vol. 6, No. 1, 063590–063590.

Sullivan, D. G., D. Lee, J. Beasley, S. Brown, and E. J. Williams. 2008a. Evaluating a Crop Residue Cover Index for Determining Tillage Regime in a Cotton–Corn–Peanut Rotation. *Journal of Soil and Water Conservation,* Vol. 63, pp. 28–36.

Sullivan, D. G., T. C. Strickland, and M. H. Masters. 2008b. Satellite Mapping of Conservation Tillage Adoption in the Little River Experimental Watershed, Georgia. *Journal of Soil and Water Conservation,* Vol. 63, pp. 112–119.

Thenkabail, P. S. 2010. Global Croplands and Their Importance for Water and Food Security in the Twenty-First Century: Towards an Evergreen Revolution That Combines a Second Green Revolution with a Blue Revolution. *Remote Sensing,* Vol. 2, pp. 2305–2312.

Thenkabail, P. S., C. M. Biradar, P. Noojipady, V. Dheeravath, Y. Li, M. Velpuri, . . . R. Dutta. 2009. Global Irrigated Area Map (GIAM), Derived from Remote Sensing, for the End of the Last Millennium. *International Journal of Remote Sensing,* Vol. 30, pp. 3679–3733.

Todoroff, P., and J. Kemp. 2016. Contribution of Remote Sensing to Crop Monitoring in Tropical Zones. Chapter 5 in *Land Surface Remote Sensing in Agriculture and Forest* (N. Baghdadi and M. Zribi, eds.). London: ISTE Press, pp. 179–220.

Tucker, C. J. 1979. Red and Photographic Infrared Linear Combinations for Monitoring Vegetation. *Remote Sensing of Environment,* Vol. 8, pp. 127–150.

van Deventer, A. P., A. D. Ward, P. H. Gowda, and J. G. Lyon. 1997. Using Thematic Mapper Data to Identify Contrasting Soil Plains and Tillage Practices. *Photogrammetric Engineering and Remote Sensing,* Vol. 63, pp. 87–93.

Whitcraft, A. K., I. Becker-Reshef, and C. O. Justice. 2015. A Framework for Defining Spatially Explicit Earth Observation Requirements for a Global Agricultural Monitoring Initiative (GEOGLAM), *Remote Sensing,* Vol. 7, No. 2, pp. 1461–1481.

Wright, C. K., and M. C. Wimberley. 2012. Recent Land Use Change in the Western Corn Belt Threatens Grasslands and Wetlands. *Proceedings of the National Academy of Sciences,* Vol. 110, No. 10, pp. 4134–4139.

Yeom, J., J. Jung, A. Chang, A. AshaPure, M. Murilo, A. Maeda, and J. Landivar. 2019. Comparison of Vegetation Indices Derived from UAV Data for Tillage Treatment in Agriculture. *Remote Sensing,* Vol. 11, No. 13, p. 1548.

Zheng, B., J. B. Campbell, and K. M. de Beurs. 2012. Remote Sensing of Crop Residue Cover Using Multi-Temporal Landsat Imagery. *Remote Sensing of Environment,* Vol. 117, pp. 177–183.

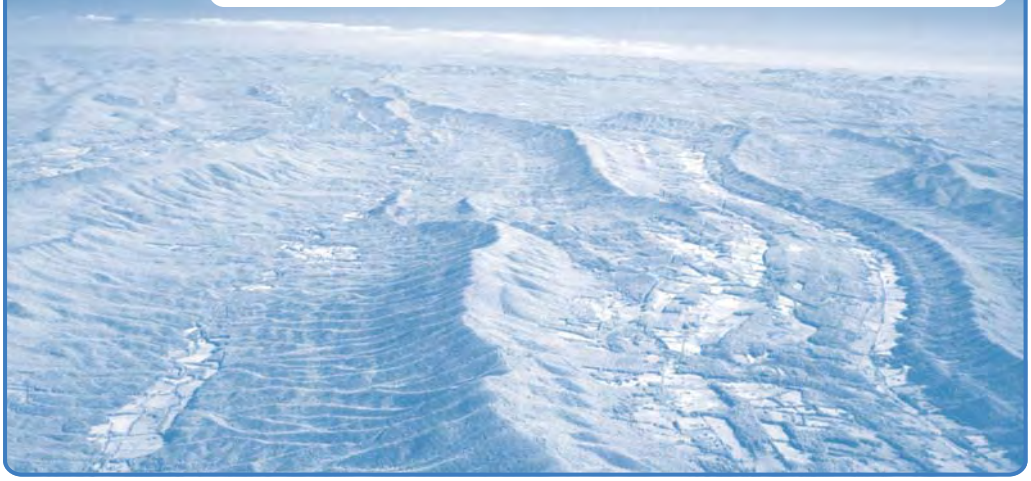
Zheng, B., J. B. Campbell, G. Serbin, and J. M. Galbraith. 2014. Remote Sensing of Crop Residue and Tillage Practices: Present Capabilities and Future Prospects. *Soil and Tillage Research,* Vol. 138, pp. 26–34.

Zheng, B., J. B. Campbell, Y. Shao, and R. H. Wynne. 2013a. Broad-Scale Monitoring of Tillage Practices Using Sequential Landsat Imagery. *Soil Science Society of America Journal,* Vol. 77, pp. 1755–1764.

Zheng, B., J. B. Campbell, G. Serbin, and C. Daughtry. 2013b. A Validation of the minNDTI Strategy in the United States. *Journal of Soil and Water Conservation,* Vol. 68, pp. 120–131.

Zheng, B., J. B. Campbell, G. Serbin, C. S. T. Daughtry, H. McNairn, and A. Pacheco. 2015. Remote Sensing of Tillage Status. Chapter 7 in *Land Resources Monitoring, Modeling, and Mapping with Remote Sensing: Remote Sensing Handbook*, Vol. 2. (P. S. Thenkabail, ed.). Boca Raton, FL: CRC Press, 849 pp.

18 Forestry



MAJOR TOPICS TO UNDERSTAND

- Silviculture
- Fire
- Inventory

18.1 INTRODUCTION

Forestry is "the profession embracing the science, art, and practice of creating, managing, using and conserving forests and associated resources for human benefit and in a sustainable manner" (Helms, 1998, p. 72). Human benefit encompasses a wide range of management objectives, including timber, nontimber forest products, water quality and yield, carbon sequestration, and biodiversity and wildlife conservation. The last objective in this list was the first historically, as the Middle English origin of the term *forest* denoted a wooded area kept for hunting (Fernow, 1907; Stevenson and Lindberg, 2017). This management objective remains vital today, with, as just one example, an active market in hunt leases.

Silviculture is the "art and science of controlling the establishment, growth, composition, health, and quality of forests and woodlands to meet the diverse needs and values

of landowners and society on a sustainable basis" (Helms, 1998, p. 167). More simply, and akin to the agriculture analog discussed in the preceding chapter, silviculture is the growing and cultivation of trees (Steven and Lindberg, 2017). Etymologically, silviculture is "wood cultivation" and agriculture is "field cultivation."[1]

Forest measurements "supply the numerical data needed to make prudent management decisions" (Burkhart, Avery, and Bullock, 2019). The fundamental unit of forest management is the forest *stand* (also known as a *compartment*), a contiguous group of trees with similar composition and growth characteristics. Forest stand mapping originated in medieval Germany (Fernow, 1907). The first stand map in the United States was made by Gifford Pinchot for the Biltmore Estate (Figure 18.1). A subset of a modern stand map for a U.S. national forest is shown as Figure 18.2. Use of remote sensing and related spatial data processing is a vital, near-daily, part of professional forestry practice. Satellite images and digital aerial photographs (from either piloted or remotely piloted aircraft) are used most frequently, but use of lidar data and lidar-derived products is becoming increasingly common.

18.2 SILVICULTURE

The range of a tree species is largely climatically driven. Recent work elucidating the specific climatic drivers of range for temperate trees (Körner et al., 2016) has revealed a "triangular interaction of inherent freezing tolerance of foliage in spring, that selects for

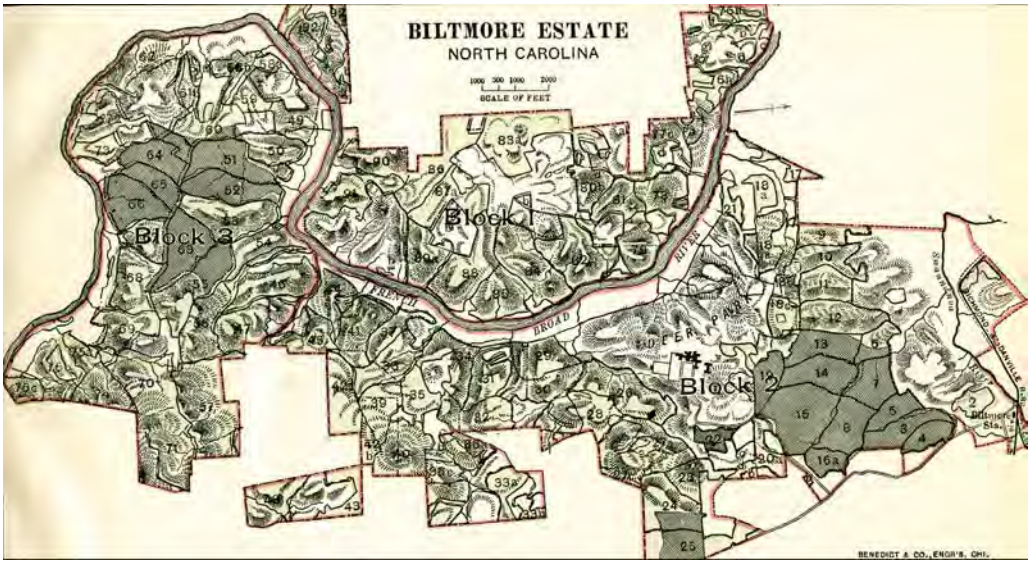


FIGURE 18.1 First stand map in the United States. From Pinchot (1893). Used by permission of the Forest History Society, Durham, North Carolina.

[1]The reader with a strong interest in forestry is encouraged to peruse modern silviculture texts such as Ashton and Kelty (2018) and Nyland et al. (2016) for more information than can be presented here.

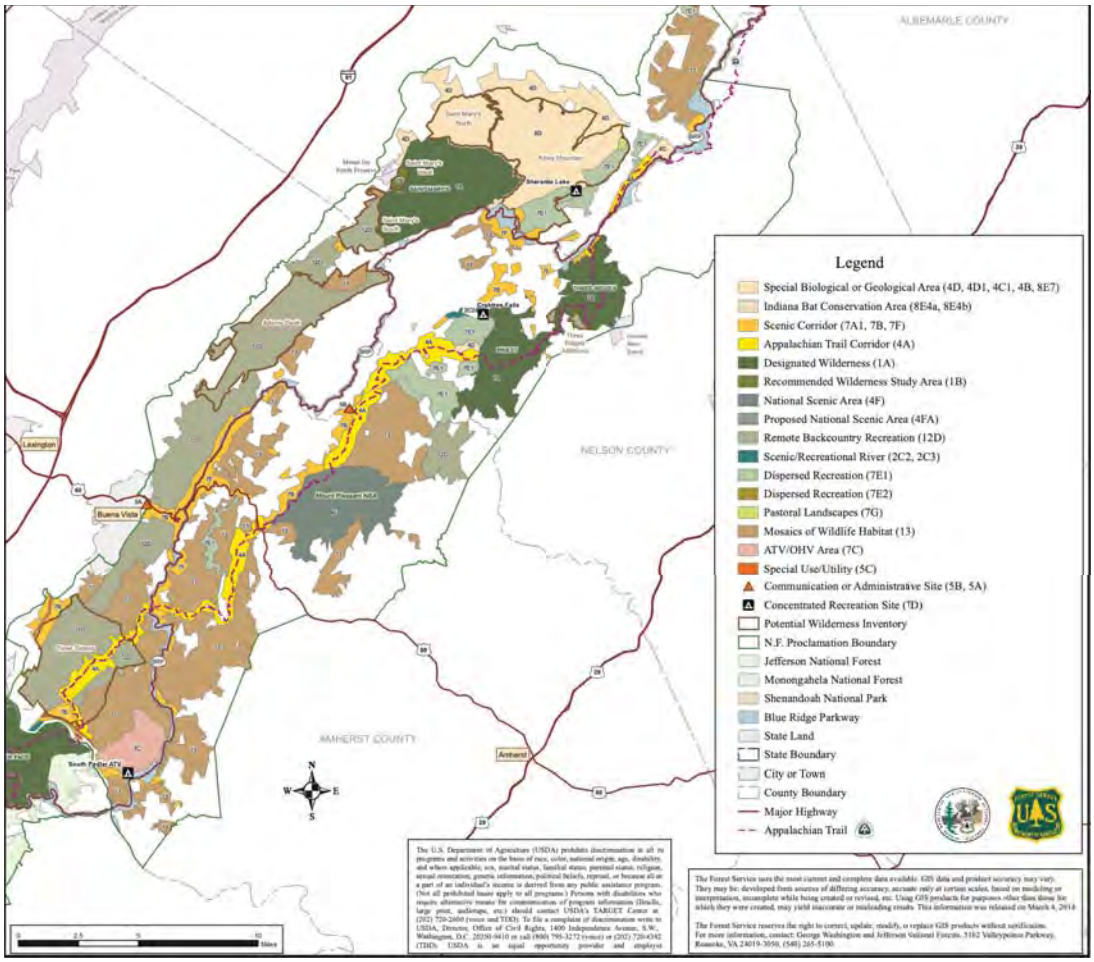


FIGURE 18.2 Subset of map from George Washington National Forest's 2014 Land and Resource Management Plan, south half, alternative I. From United States Forest Service (2014).

a certain phenological control of spring flushing, which in turn truncates the length of the growing season, and thus, the time available for tissue maturation" (p. 1085). To simplify: the tree species that *could* grow in a particular area is climatically driven. However, the actual distribution (presence at a given site) of a given temperate tree species is largely edaphically controlled (Walthert and Meier, 2017).

To better understand the hierarchy of silvicultural decision making, for simplicity, let's start with a backyard gardening analogy. You are going to plant tomatoes but have to decide which varieties to plant. That decision is based in part on the soils you either have or (for smaller areas) choose to create combined with the prevailing climate. These together comprise the edaphic ("of, produced by, or influenced by the soil" per Stevenson and Lindberg, 2017) and climatic drivers, the abiotic environment. Once you have found the suitable tomato varieties for both your personal tastes and environmental constraints, you will usually then want to grow as many as you can—the focus of the next section.

Intensive silviculture requires an understanding of *genetic* × *environment* × *silviculture* (G × E × S) interactions (Rubilar et al., 2018) to improve stand productivity. Recall from Chapter 16 that net (primary) productivity is the net production of biomass/unit area/unit time (also *sensu* Powers, 1999). Since leaves grow on trees, stand productivity is directly related to the available leaf area. Powers (1999, p. 273) introduced the pivotal concept of *potential productivity,* which is the site's potential for biomass production when "at full carrying capacity for leaf area" (Figure 18.3). The constraints on potential productivity are climate, soils, and genetics. Powers (1999) further, and importantly, notes that each of these three factors is also mutable (through, for example, irrigation, fertilization, and improved growing stock; Figure 18.3), so potential productivity is not fixed for a given site/species. Actual current productivity, in contrast, is also directly related to the actual, rather than potential, leaf area for a given stand/age.

Reaching, and increasing, potential productivity is the goal of silvicultural treatments. Per Rubilar et al. (2018), there are four broad categories of treatments, as follows: (1) managing carrying capacity (initial planting density, thinning), (2) managing radiation interception (pruning, thinning), (3) reducing site limitations (bedding, release from competing vegetation), and (4) adding limiting resources (irrigation, fertilization).

Given that productivity, whether it be actual or potential, is directly related to leaf area, remote sensing applications for silvicultural decision support focus on precise estimation of leaf area (usually leaf area index [LAI], Chapter 16). Potential productivity is established through either process-based modeling or field trials (such as twin plots) where it can be directly measured and/or increased (Leite da Silva et al., 2016). However, as noted by Reich (2012), once leaves (estimated by remotely sensed LAI) harvest light, they must biochemically fix carbon using the photosynthetic chemistry of the foliage (estimated by percent of nitrogen). If we also add the length of the growing season (estimated by land surface phenology metrics, Chapter 16), almost 90% of the variance in productivity is explained.

In this section, we focus on estimation of LAI, not nitrogen. The reasons for this are fourfold. First, foliar nitrogen estimation using remote sensing requires imaging spectros-

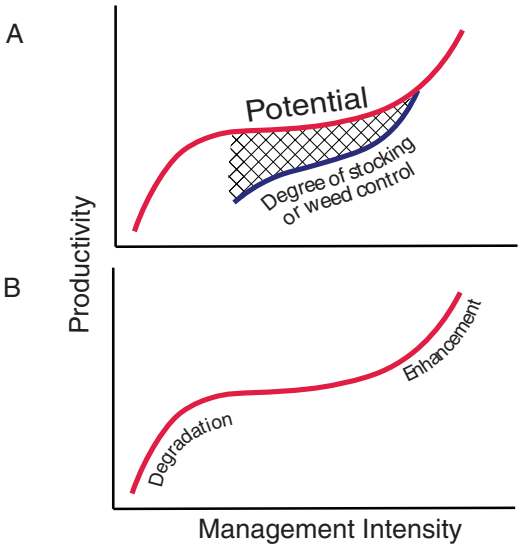


FIGURE 18.3 Relationship of management intensity to site productivity: (A) managing to potential productivity; (B) improving potential productivity. Adapted from Powers (1999). Image by Susmita Sen.

copy (Martin et al., 2008), and the requisite operational satellite sensors with the appropriate signal-to-noise ratio and spatial resolution will likely not be available until the launch of the German Environmental Mapping and Analysis Program (EnMAP) hyperspectral satellite mission (Guanter et al., 2015), followed half a decade later by the NASA Surface Biology and Geology mission (National Academies of Sciences, Engineering, and Medicine, 2018). Second, high interband correlations in hyperspectral data require tailored analytical approaches, such as partial least squares regression, which are beyond the scope of this book. Third, the methods by which foliar nitrogen can be estimated using remote sensing have focused on interspecific models; monospecific estimation accuracy using remote sensing may not yet be sufficient for tailored silvicultural prescriptions. Fourth, and most important, a stand's "response" to increased nitrogen availability is often to increase total LAI (Albaugh et al., 1998) but not change foliar nitrogen percentages (Mayor, Wright, and Turner, 2014).

LAI

Leaves grow trees, and LAI is how foresters quantify leaf area. Forest productivity and a wide range of productivity-related ecosystem goods and services are thus associated with remote estimation of LAI.

Passive

LAI can be estimated passively, using passive optical sensors, or actively, using either synthetic aperture radar (less common) or lidar (more common). Given their widespread availability and use, we start our discussion with optical estimation.

In Chapter 16 we introduced some important vegetation indices, including the simple ratio (SR), normalized difference moisture index (NDMI), and normalized difference vegetation index (NDVI). All three of these indices are linearly related to LAI before saturation occurs, which varies by both index and forest composition. The reduced simple ratio (RSR) is also widely utilized in empirical models of LAI and has been shown to outperform SR in some instances, particularly in boreal forests. RSR is defined as follows:

$$\mathrm{RSR} = \mathrm{SR}\left(1 - \frac{\mathrm{SWIR1} - \mathrm{SWIR1}_{\mathrm{min}}}{\mathrm{SWIR1}_{\mathrm{max}} - \mathrm{SWIR1}_{\mathrm{min}}}\right) \qquad \textbf{(EQ. 18.1)}$$

where *SWIR1* (shortwave infrared 1) is the reflectance from the wavelength band centered around 1.6 µm (band 6 on Landsat 8 Operational Land Imager [OLI] and band 11 on Sentinel-2 MultiSpectral Instrument [MSI]). The minimum and maximum are defined as the SWIR1 values for a completely closed and completely open canopy, respectively. In practice, though, these are often defined as the maximum and minimum for the forested area under consideration. The analyst must be careful to exclude water pixels in this calculation. The scene and forest specificity of the minimum and maximum SWIR1 reflectances in RSR are both its strength and its weakness. The specificity strengthens local relationships with biophysical variables like LAI. However, since the RSR varies with local conditions (because the range in SWIR1 does), scientists and managers have difficulty in comparing RSR levels across space and time.

The scale of management requires remotely sensed data at a comparable scale. This, at a minimum, means use of the well-calibrated and radiometrically consistent moder-

ate resolution Earth resource satellites such as Sentinel-2 MSI, Landsat 8 OLI, and the *Satellite Pour l'Observation de la Terre (SPOT)* series. High spatial resolution commercial satellite data are also increasingly utilized by forest scientists and managers, albeit with foreknowledge of potentially reduced radiometric consistency due to different look angles, inherent sensor characteristics, and the like.

Ganguly et al. (2012) proposed a physically based algorithm for LAI retrieval using Landsat based on canopy spectral invariants theory (Huang et al., 2007) to parameterize the bidirectional reflectance factor. In addition, of course, to high-quality sensor-observed surface reflectances (now routinely available, as noted in Chapter 11, for both Sentinel-2 MSI and Landsat), global application of this or similar algorithms also requires accurate pixel-specific land cover to be available at least annually. While the details of either radiative transfer or spectral invariants theory is beyond the scope of this book, it is likely that this or another physically based LAI retrieval algorithm will eventually be the basis of a higher-order LAI product, but at the time of this writing was not yet available.

Given the lack of availability of an accurate, global, physically based LAI product, empirical modeling of LAI has dominated research and applications for decades. Most such efforts relate one or more vegetation indices to field-measured LAI to develop regionally applicable (and usually species-specific) models. The most common indices used are SR, NDVI, and RSR. Two examples are shown below. The first is for Norway spruce (*Picea abies* (L.) Karst.) stands in Finland (Stenberg et al., 2004):

$$\mathrm{LAI} = 0.18 * \mathrm{RSR} + 1.04 \qquad \textbf{(EQ. 18.2)}$$

The second is for loblolly pine (*Pinus taeda* L.) stands in the southern United States (Blinn et al., 2019):

$$\mathrm{LAI} = 0.33 * \mathrm{SR} - 0.0021 \qquad \textbf{(EQ. 18.3)}$$

Many other examples abound and are in local to regional use by practitioners.

Active

Both lidar and synthetic aperture radar data are utilized to estimate LAI. Lefsky et al. (1999) found that LAI could be estimated using metrics (filled canopy volume, closed gap volume, and number of canopy classes per unit height) derived from waveform lidar data. They note that after crown closure, LAI increases come from distributing leaves over a larger volume. The potential volume available increases with canopy height. As such, early studies using the more commonly available small-footprint airborne laser scanner data could relate height metrics directly to LAI for specific study sites. One example is the study by Lim et al. (2003) in which log-transformed LAI was shown to be well related to the log-transformed mean laser height derived from all returns.

Discrete return airbone laser scanner data enable the calculation of leaf penetration metrics that are biophysically related to leaf area (rather than using height as a predictor; Zhao and Popescu, 2009). While these are many and varied, two that stand out are the laser penetration index (LPI; Barilotti et al., 2005) and the above and below ratio index (ABRI; Sumnall et al., 2016b). There are numerous formulations of both indices, but we present only the original concepts here. The LPI is the ratio of the number of returns that hit the ground divided by the total number of returns (i.e., both those that hit vegetation and those that hit the ground). ABRI is defined as the ratio of the number of vegetation

returns divided by the number of ground returns. Importantly, if pulse energy and/or density are sufficient, one can partition the LAI vertically, enabling, for example, detection of competing understory vegetation (Sumnall et al., 2016a).

Synthetic aperture radar can also be used to estimate LAI. Frequencies with wavelengths roughly corresponding to leaf size (i.e., the X and C bands) are commonly used. Peduzzi et al. (2012) found that variables derived from both X-band interferometric height and backscatter were important for loblolly pine LAI estimation. Stankevich et al. (2017) were able to estimate LAI accurately and parsimoniously using Sentinel-1 (C-band) data through use of the relative difference polarization index (RDPI), defined as follows:

$$\mathrm{RDPI} = \frac{\sigma^0_{VH}}{\sigma^0_{VH} + \tau \cdot \sigma^0_{VV}} \quad \text{(EQ. 18.4)}$$

where σ^0_{VH} and σ^0_{VV} are backscatter coefficients (normalized radar cross sections) from different polarization bands and τ is the degree of polarization. Unfortunately, however, the RDPI must be further modified using orientation information computed from the radar heading angle, radar incidence angle, terrain slope, and terrain aspect.

Chen et al. (2017) compared both optical and synthetic aperture radar–derived (SAR-derived) indices for their utility in estimating LAI in a mixed forest in Inner Mongolia. They found that the optically derived SR (the so-called ratio vegetation index, or RVI) produced the best model out of 20 compared sensors/indices. However, two SAR-derived indices (Figure 18.4), the radar RVI (RRVI) and the radar NDVI (RNDVI), computed using L-band data from ALOS PALSAR, outperformed all other indices but one, the atmospherically corrected vegetation index (ARVI; Kaufman and Tanré, 1992). These two radar vegetation indices are defined as follows:

$$\mathrm{RRVI} = \frac{\sigma^0_{HH}}{\sigma^0_{HV}} \quad \text{(EQ. 18.5)}$$

$$\mathrm{RNDVI} = \frac{\sigma^0_{HV} - \sigma^0_{HH}}{\sigma^0_{HV} + \sigma^0_{HH}} \quad \text{(EQ. 18.6)}$$

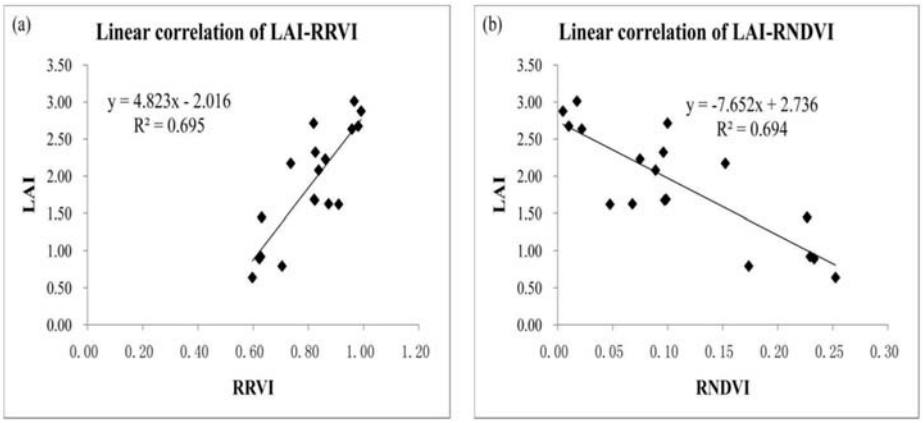


FIGURE 18.4 Relationship between LAI and (L-band) SAR-derived vegetation indices: (a) radar ratio vegetation index, (b) radar normalized difference vegetation index. From Chen et al. (2017).

where, as above, σ^0_{HV} and σ^0_{HH} are backscatter coefficients from different polarization bands.

Polarimetric SAR-based estimation of forest LAI thus shows great promise and is effective independent of cloud occlusion. While X, C, and L bands have all been used to estimate LAI successfully, as noted above, seminal early work (Paloscia, 1998) highlighted the L band as being preferred. Thus, the NASA-ISRO SAR (NISAR), slated at the time of this writing for launch in 2023, will have strong potential for LAI estimation. Technical challenges still remain, however, and lidar or optical measurements are thus more routinely used by practitioners for LAI estimation.

Competing Vegetation

Control of competing vegetation is important to maximize survival and growth immediately postestablishment. This has been definitively shown for many target species, including birch (Hytönen and Jylhä, 2005), poplar (Böhlenius and Övergaard, 2015), oak (Jensen and Löf, 2017), and Douglas fir (Roberts, Harrington, and Terry, 2005). Control of competing vegetation is also important for growth after the establishment phase. This, too, has been well demonstrated for a wide variety of target species, including loblolly pine (Lauer and Glover, 1999; Yin and Sedjo, 2001; Carter and Foster, 2006), Douglas fir, Eucalptus spp., and radiata pine (Wagner et al., 2006), and white fir and sugar pine (Zhang et al., 2017).

Competing vegetation is aptly named, since whatever site resources go to the nontarget species are not going to the target species. Specific mechanisms have been demonstrated ecophysiologically for numerous site resources. As such, control of competing vegetation can lead to increased nitrogen availability (Slesak, Harrington, and Schoenholtz, 2010), increased water availability (Vargas et al., 2018), and increased light availability (Hytönen and Jylhä, 2005).

Controlling competing vegetation is thus quite important, and remote sensing is increasingly being used to determine whether control is needed at a given stand. Efficacious methods vary depending on stand composition and vertical structure. In cases where deciduous understory species are competing with overstory evergreens, for example, vegetation indices can be compared across seasons to assess the amount of leaf area in the understory. Spring is a particularly important time (Knight et al., 2004; Blinn et al., 2012) since leaves start appearing on the deciduous trees while overstory evergreens are still at or near their leaf area minimum. In instances when the understory is also evergreen (such as ericaceous understory in the flatwoods of the U.S. South), winter is usually the best season. Peduzzi, Allen, and Wynne (2010), for example, showed a strong correlation between winter SR and ericaceous understory LAI in slash pine plantations.

In some instances, the competing vegetation is at or near the same vertical position in the canopy. In these instances (which include stand establishment and immediately thereafter), classification using visible, near-infrared, and shortwave infrared spectral bands (including vegetation indices derived therefrom) has been effective in separating target species from competing species (Prasad and Brooks, 1989; Knight et al., 2004) as long as the spatial resolution is sufficiently fine. Tree species can be separated using point cloud data alone (e.g., for lidar, Brandtberg, 2007) or, with better results, in combination with optical data (Ke, Quackenbush, and Im, 2010). Given that species are often easily separated using hyperspectral data (van Aardt and Wynne, 2007), the combination of canopy height and hyperspectral data is particularly powerful in identifying the presence and amount of competing vegetation (Tuominen et al., 2018).

Keep in mind, however, that the silvicultural decision is often whether to release the stand (mechanically or chemically) from competing vegetation. As such, just knowing the relative magnitude of the competition with respect to leaf area is often sufficient. Approaches include seasonal vegetation index analysis (as noted above) and, increasingly, vertical partitioning of leaf area using lidar data (Sumnall et al., 2016a). The latter is particularly efficacious, though it has been shown to be less accurate if the difference between the height of the target and competing vegetation is low (Korpela et al., 2008), including at establishment, or if the distance between the ground and vegetation height is low (because of discrete return lidar dead zones; Nayegandhi and Brock, 2008).

18.3 FIRE

There have been forest fires as long as there have been forests (Bowman et al., 2009), and humans have long influenced forest fire patterns in time and space. As of this writing, Australia just recorded its worst fire season (Tarabay, 2020), the rate of rainforest loss to fire in the Brazilian Amazon was higher than that in any year in the preceding decade (Kraus et al., 2019), and a reforesting Europe was primed to ignite (Sengupta, 2020). A warming world is lengthening the fire seasons in temperate and boreal forests (Flannigan et al., 2009), and the wildland-urban interface continues to expand (Intini et al., 2020). In the United States, the federal costs for fighting fires exceed 2 billion dollars annually and are by far the largest part of the U.S. Forest Service's budget (United States Forest Service, 2017).

Remote sensing is vital to every aspect of wildland fire monitoring and management. Remote sensing is used to (1) map fuel types (groups of characteristics for both live and dead biomass; Arroyo, Pascual, and Manzanera, 2008), (2) conduct active fire monitoring, (3) map burned areas, and (4) characterize postfire ecological effects (Lentile et al., 2006). We focus on the first two of these in this chapter.

Fuel-Type Mapping

A *fuelbed* is a relatively homogeneous landscape unit with fuel characteristics that represent a distinct combustion environment (Riccardi et al., 2007; United States Forest Service, 2020b). Excerpts from the characteristics are shown in Table 18.1 for two different species and in Table 18.2 for the same species with different conditions. Remote sensing can be used (almost always in conjunction with field and other ancillary data) in three principal ways with respect to fuelbeds, as follows: (1) selection of the appropriate fuelbed for a given pixel or object, (2) partial characterization of a newly identified fuelbed, and (3) development of new fuelbeds (e.g., Pettinari and Chuvieco, 2016).

Characterization of fuels thus requires information on structure (both horizontal and vertical) as well as composition. Stereo aerial photography, since it provides both structure and spectral information, has been long used (in concert with field data) for fuel characterization, acquired using both manned (Oswald et al., 1999) and unmanned (Shin et al., 2018; Figure 18.5) aircraft. Three-dimensional structure can also be obtained using airborne (Mutlu et al., 2008; Erdody and Moskal, 2010) and terrestrial (Rowell et al., 2020) laser scanning. In addition to aerial photography, multispectral (Mutlu et al., 2008) and hyperspectral (Lasaponara, Lanorte, and Pignatti, 2006; Varga and Asner, 2008) data (from air- or spaceborne sensors) can be used to determine composition.

TABLE 18.1 Select Fuelbed Characteristics for Two Different Species

	Overstory cover (%)	DBH (inches)	Overstory height (feet)	Primary herb. cover (%)	Herb. cover primary species
Douglas-fir forest	55	12	70	30	*Polystichum munitum*
Loblolly-slash pine plantation	90	7	70	10	*Toxicodendron radicans*

Note: DBH = diameter at breast height; herb. = herbaceous. Shown are Fuel Characteristics Classification System (FCCS) fuelbed 3, Douglas-fir forest, and fuelbed 161, Loblolly-slash pine plantation. FCCS version 3.0 as accessed from Fire and Fuel Tools version 2.0.

TABLE 18.2 Select Fuelbed Characteristics for the Same Species, *pinus contorta* (Lodgepole Pine)

	Overstory cover (%)	Overstory height (feet)	Density (trees/acres)	DBH (inches)
21	50	5	5400	0.5
22	70	50	402	4.6
23	50	63.2	218	8.2

Note: DBH = diameter at breast height. Shown are Fuel Characteristics Classification System (FCCS) fuels 21, young lodgepole pine forest; 22, mature lodgepole pine forest; and 23, mature lodgepole pine forest with bark beetle damage. FCCS version 3.0 as accessed from Fire and Fuel Tools version 2.0.

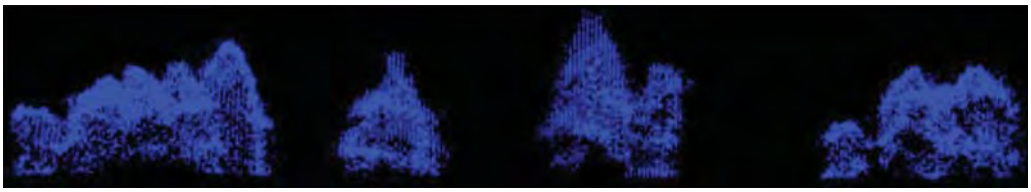


FIGURE 18.5 Structure from motion point cloud excerpted from the acquisition used to estimate forest canopy fuels (canopy cover and tree density) in a *Pinus ponderosa* (ponderosa pine) stand. The images from which the point cloud was derived were acquired using a multispectral camera on the SenseFly eBee fixed-wing unmanned aerial vehicle (minimum endlap and sidelap 85%, maximum flight altitude 120 m). From Shin et al. (2018).

Active Fire Monitoring

The core principle behind active fire monitoring is that as fires get hotter, their emitted energy increases and the peak of their emitted energy decreases in wavelength. This is shown conceptually in Figure 18.6, which shows the Planck curves for a 780 K fire and a normal surface (280 K) not undergoing combustion. The bandwidth for the I4 band from the Visible Infrared Imaging Radiometer Suite (VIIRS) is shown in green. At night, a simple brightness temperature threshold is often used. For example, the VIIRS 375-m active fire detection data product (Schroeder et al., 2014) uses 320 K as the nighttime brightness temperature threshold (meaning any pixel hotter than that is flagged as burn-

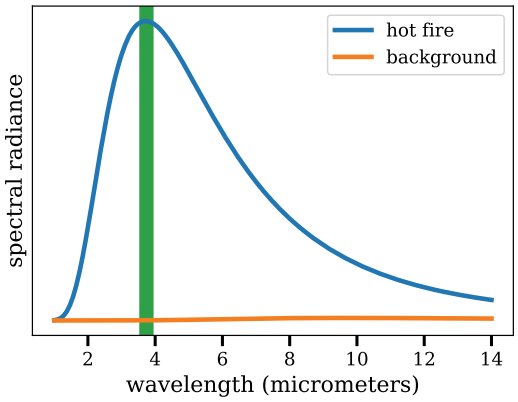


FIGURE 18.6 Core principle behind active fire monitoring illustrated using Planck curves for an active fire (at 780 K) and typical surface temperature (at 280 K). Units for the *y* axis are $W\,m^{-2}\,sr^{-1}\,\mu m^{-1}$. Bandwidth of VIIRS I4 band shown in green.

ing if the observation is cloud-free and of sufficient quality). The precursor Moderate Resolution Imaging Spectroradiometer (MODIS) product used 330 K as the threshold for a comparable portion of the electromagnetic spectrum, and the Sentinel-3 Sea and Land Surface Temperature Radiometer uses 326 K (with a large dynamic range specifically designed for active fire monitoring; Xu et al., 2020). False alarms can occur if the same threshold is used in the daytime (since the ambient surface temperature is hotter), so the absolute threshold is increased and an alternative flagging technique, the difference between the brightness temperatures at 4 μm and 11 μm, is also used. We note that, while conceptually simple, the implemented operational algorithms are complex, able to handle a variety of fire types, background land covers, and background temperatures. An example of the VIIRS 375-m active fire detection product from February 2, 2021, is shown as Figure 18.7.

FIGURE 18.7 Example of VIIRS active fires detected over North America on February 2, 2021.

Even 375-m resolution is insufficient for a tactical response, however, so aircraft-mounted thermal sensors have long been used (since the 1960s in the United States) to aid firefighting. Sensor types include (1) gimbal-mounted camera ball systems (combining optical wavelengths via a complementary metal oxide semiconductor and the medium-wave infrared with a microbolometer) and (2) scanners (United States Forest Service, 2020c). The use case is slightly different for each. Data from camera ball systems can be acquired using both manned and unmanned (e.g., Christensen, 2015) aircraft and have higher spatial resolution but typically less coverage. As such, they are particularly good for detecting humans and other mammals (especially for cooled sensors), spotting fires, and providing support for ground operations. The thermal scanners, in contrast, are typically flown at least 10,000 feet (3,048 m) above ground, have a high scan volume, provide much greater areal coverage, and are best for mapping large fires (**Figure 18.8**) or multiple fires spanning a large geographic area.

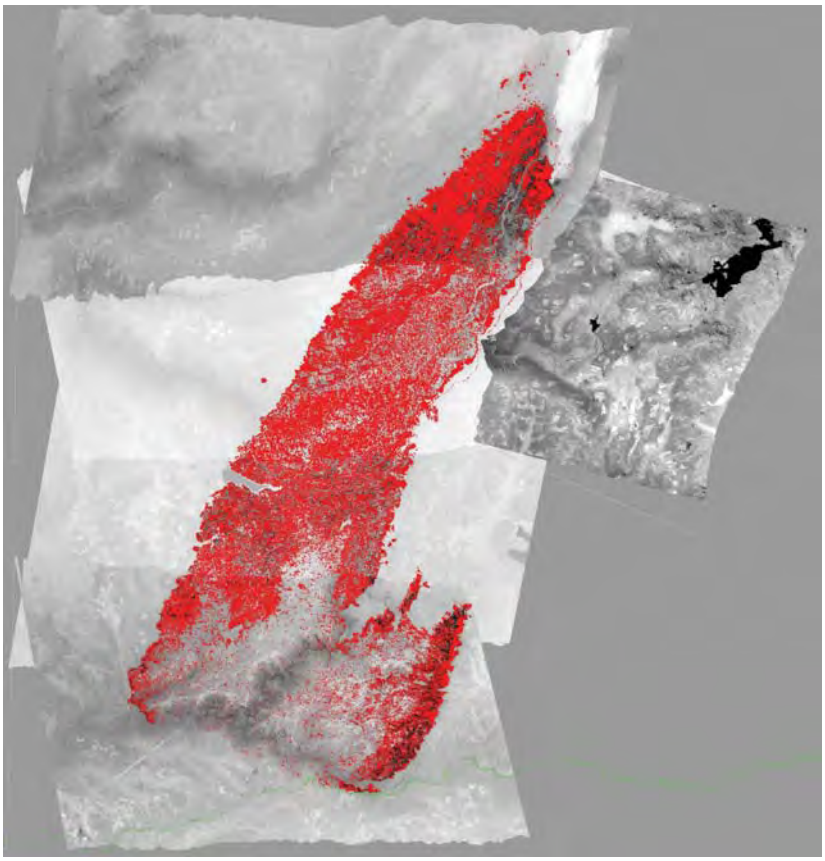


FIGURE 18.8 Phoenix image mosaic of the King Fire (El Dorado County, California) acquired September 18, 2014, at or near 1:18 A.M. local time (07:18 UTC), 5 days into the 26-day fire that consumed 97,717 acres (39,545 ha). Both red and black pixels indicate active fire. The Phoenix line scanner has an 8-bit quantization level, two wavelength bands (one spanning 3–5 μm for fire detection and the other spanning 8–14 μm for background terrain temperature), and a spatial resolution of approximately 3.5 m at nadir at typical flying heights (10,000 ft above ground). Image from United States Forest Service (2020a); fire statistics from California Department of Forestry and Fire Protection (2020).

18.4 INVENTORY

Species Identification

Identifying trees, whether on the ground or using remotely sensed data, has always been critical to the practice of forestry. As Gifford Pinchot noted in 1914 (p. 18), "The trained Forester must know the forest as a doctor knows the human machine. First of all, he must be able to distinguish the different trees of which the forest is composed, for that is like learning to read."

Interpretation of Species

Identification of individual plants can be accomplished by close examination of crown size and shape (Figure 18.9). At the edges of forested areas, the shadows of trees can form clues to their identification. At smaller scales, individual plants are not recognizable, and the interpreter or machine vision algorithm must examine patterns formed by the aggregate appearance of individual stands, in which individual crowns form a distinctive tone and texture. Identification may be relatively straightforward if stands are pure or are composed of only a few species that occur in consistent proportions. If many species are present and their proportions vary, then specific identification may not be possible, and the use of such broad designations as *mixed deciduous forest* may be necessary (as supported by extant classification schemes; analysts are cautioned to work carefully within the context of an appropriate scheme). The level of categorical specificity that can be achieved in a given classification scheme is scale dependent, as shown in Table 18.3.

At these smaller scales, cover types are distinguished using differences in image tone, image texture, site, and shadow rather than spectral differences alone. Cover classes can be considered as rather broad vegetation classes, based perhaps on the predominant species present, their age, and their degree of crown closure. Thus, forest-cover-type classes might include *aspen/mixed conifer* and *Douglas fir,* indicating the predominant species without precluding the presence of others. Subclasses or secondary descriptors could indicate sizes of the trees, crown closure, and presence of undergrowth.

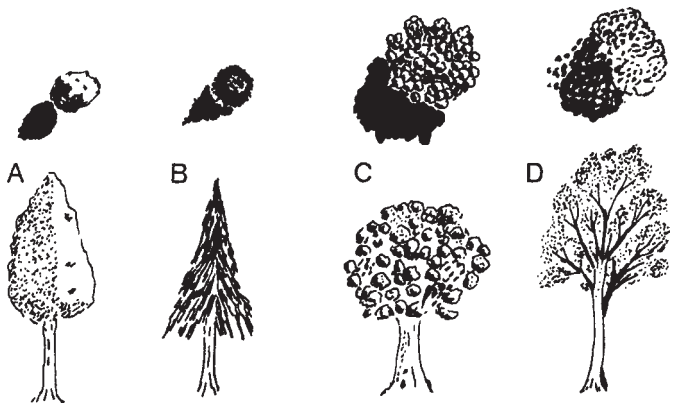


FIGURE 18.9 Identification of individual plants by crown size and shape. Sizes and shapes of individual tree canopies form distinctive shadows, as do the structures and densities of the individual branches.

TABLE 18.3 Examples of Cover Types

a. At small scale (1:1,000,000 to 1:50,000), cover types must be defined at coarse levels of detail because it will not be possible to distinguish or to delineate classes that are uniform at the species level:
   "evergreen forest"
   "deciduous forest"
   "chaparral"
   "mixed broadleaf forest"
   "mangrove forest"

b. At large scale (larger than about 1:20,000), cover types may be defined at fine levels of detail. Sometimes, when stands are very uniform, cover types can be mapped at the detail of individual species:
   "balsam fir"
   "shortleaf pine"
   "aspen and white birch"
   "oak–hickory forest"

Forest Photogrammetry

Foresters desire to identify specific *stands* of timber, areas of forest with uniform species composition, age, and density that can be treated as homogeneous units. Stands are the basic unit of forest management, so forest managers wish to monitor their growth over time to detect the effects of disease, insect damage, and fire or drought. Even when stands have been planted from seedlings by commercial foresters, aerial photography, by virtue of its map-like perspective and its wealth of environmental information, provides accurate and economical information concerning the condition of the stand at a specific time.

If large-scale, high-quality images are available (preferably in stereo), it is possible to apply the principles of photogrammetry to measurement of factors of significance in forestry. In most instances, these assessments are made using algorithmic approaches beyond the scope of this book. We note, however, that drone-acquired structure from motion photogrammetry is leading to a resurgence in photogrammetrically derived forest measurements (e.g., Figure 18.5).

Crown Closure

One of the most important variables that contributes to estimates of stand density is assessment of *crown closure,* the proportion of the area of a stand that that is covered by crowns of trees. Crown closure measures the density of trees in a stand and indicates the degree of competition between trees as a stand matures. Because crown closure is also related to stand volume, it assists predictions of economic dimensions of a stand at harvest. Because the size of the crown of a tree is closely related to many of its physiological characteristics (such as its ability to conduct photosynthesis), crown closure, monitored over time, permits foresters to assess the growth of the forest.

Although crown closure can be measured at ground level using any of several procedures, aerial photography forms a valuable tool for assessing crown closure. It assesses the percentage of a stand that is covered by the crowns of the dominant and co-dominant species. (Crowns of understory trees are not easily resolved with photogrammetry, though understory trees and shrubs are often detectable using airborne laser scanning, *sensu* Gopalakrishnan et al., 2018.) Crown closure is reported as a percentage (i.e., 60%), or as a decimal in the form 0.6, or 0.7.

With experience, crown closure can be accurately estimated by eye, though analog interpretation is rare in current workflows. The use of stereo is especially effective when understory crowns are visible. The quality of the photography can be important, including scale, shadows, and haze. So also can variables related to the stand, including the presence of irregular topography, the nature of spacing of trees, and the backgrounds of other vegetation.

Timber Volume Estimation

Foresters are routinely interested in estimating timber volume for a specific stand as a means of monitoring its growth over time, assessing management practices, and determining the amount of timber to be obtained at harvest. Volume measurement consists of estimates of board-foot or cubic-foot volumes to be obtained from a specific tree or, more often, from a specific stand. In the field, the forester measures the diameter of the tree at breast height (DBH) and the height of the straight section of the trunk (bole) as the two basic components to volume estimation.

There are many approaches to estimation of volume from measurements derived from aerial photographs; there can be no universally applicable relationship between photo measurements and timber volume, as species composition, size, age, soil, and climate vary so greatly from place to place.

Lidar Data

Chapter 9 introduced the application of laser altimetry to acquisition of data characterizing vegetation structure, and we noted its clear utility to LAI estimation earlier in this chapter. Because lidar data directly characterize the physical structure of a forest (heights, crown closure, crown size, etc.), they provide an opportunity to directly assess the three-dimensional structure of vegetation formation in ways that are not possible with other sensors. As such, when available, they have for decades been "the gold standard" of remotely sensed data used for forest assessment and inventory. They form the remote sensing "backbone" of many corporate and national forest inventory systems (always in concert with high-quality, well-geolocated field measurements). Furthermore, as was noted earlier, lidar pulses penetrate to lower canopy layers, which is not possible using photogrammetric approaches (classical or structure from motion; **Figure 18.10**). Discrete-return airborne laser scanning is by far the workhorse lidar data type for forestry applications, but sampling approaches using terrestrial laser scanning, drone-acquired lidar, and spaceborne lidars (ICESat-2, GEDI) are employed, depending on the forest science or management objective.

18.5 SUMMARY

Ensuring healthy forests is vital for humans and the earth system. Forests are managed for various uses, including commodity wood and fiber, water quality, and carbon sequestration. Forests remain under threat globally from conversion to agriculture and a changing climate. Remotely sensed data from both active and passive sensors provide the data needed for monitoring, modeling, and management. Information derived from advances in forest remote sensing improves decisions at the management and policy scales.

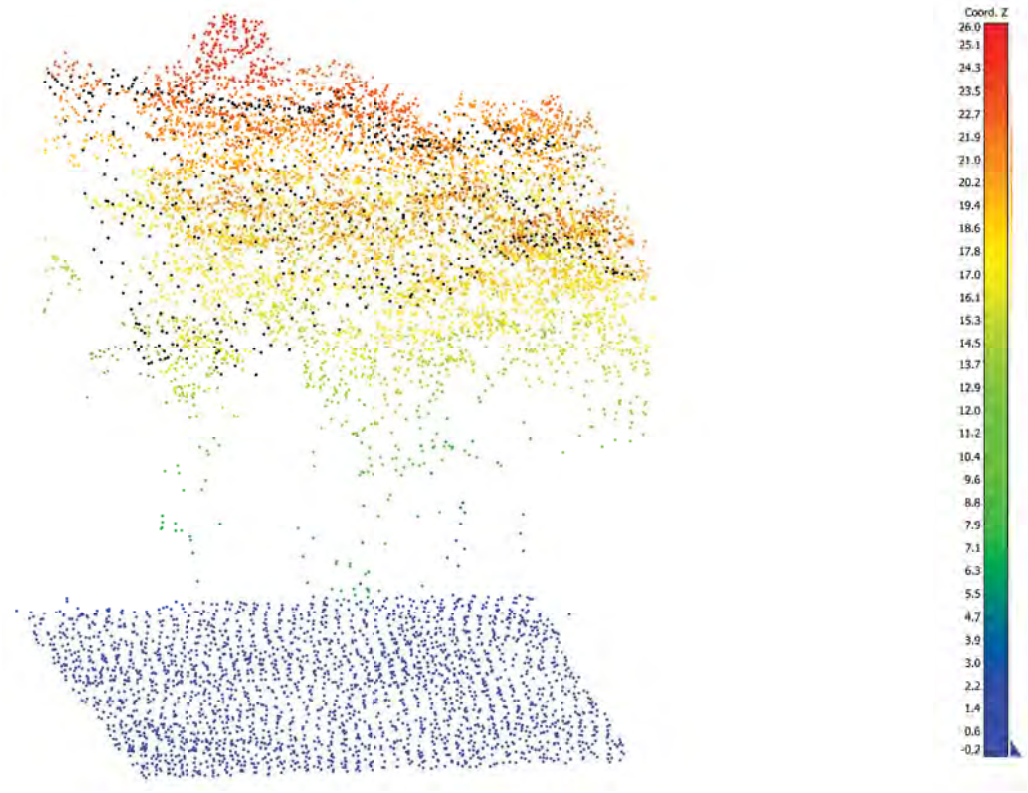


FIGURE 18.10 Normalized lidar point cloud (also shown as **Figure 9.9**) captures the whole forest structure, from ground level (dark blue) to the very top of the canopy (red). The photogrammetric point cloud (black; new to this figure) depicts the upper to midstory canopy, but it is unable to detect the ground. The lidar is sourced from the National Ecological Observatory Network Mountain Lake Biological Station. The photogrammetric point cloud is sourced from the National Agriculture Imagery Program. Both data sets were acquired in 2018. Area is approximately 1410 sq m. From Elizabeth M. Prior. Used by permission.

REVIEW QUESTIONS

1. Summarize differences between classification of forests from ground observations versus classification from aerial images. Consider such factors as (a) the basis of the classification and (b) the units classified. Identify distinctions for which remote sensing is especially well suited and those for which it is not likely to be useful.
2. Why is LAI important for forestry?
3. Why might a forester choose to use drone-based structure from motion photogrammetry rather than airborne laser scanning from a manned aircraft?
4. What effect will shadows have on the optical assessments of LAI?

5. Why is detection of forest fires more challenging in daytime than at night? If you have to implement a daytime acquisition, what thermal bands will be needed to improve detection?
6. Why has lidar data become so important for forestry applications of remote sensing?

REFERENCES

Albaugh, T. J., H. L. Allen, P. M. Dougherty, L. W. Kress, and J. S. King. 1998. Leaf Area and Above- and Belowground Growth Responses of Loblolly Pine to Nutrient and Water Additions. *Forest Science,* Vol. 44, pp. 317–328.

Arroyo, L. A., C. Pascual, and J. A. Manzanera. 2008. Fire Models and Methods to Map Fuel Types: The Role of Remote Sensing. *Forest Ecology and Management,* Vol. 256, pp. 1239–1252.

Ashton, M. S., and M. J. Kelty. 2018. *The Practice of Silviculture: Applied Forest Ecology* (10th edition). New York: John Wiley, 776 pp.

Barilotti, A., S. Turco, R. Napolitano, and E. Bressan. 2005. La Tecnologia LiDAR per lo Studio Della Biomassa Negli Ecosistemi Forestali. Proceedings of the 15th meeting of the Italian Society of Ecology, Torino, Italy, September 12–14.

Blinn, C. E., T. J. Albaugh, T. R. Fox, R. H. Wynne, J. L. Stape, R. A. Rubilar, and H. L. Allen. 2012. A Method for Estimating Deciduous Competition in Pine Stands Using Landsat. *Southern Journal of Applied Forestry,* Vol. 36, pp. 71–78.

Blinn, C. E., M. N. House, R. H. Wynne, V. A. Thomas, T. R. Fox, and M. Sumnall. 2019. Landsat 8 Based Leaf Area Index Estimation in Loblolly Pine Plantations. *Forests,* Vol. 10, Article No. 222, 18 pp.

Böhlenius, H., and R. Övergaard. 2015. Growth Response of Hybrid Poplars to Different Types and Levels of Vegetation Control. *Scandinavian Journal of Forest Research,* Vol. 30, pp. 516–525.

Bowman, D. M. J. S., J. K. Balch, P. Artaxo, W. J. Bond, J. M. Carlson, M. A. Cochrane, . . . and S. J. Pyne. 2009. Fire in the Earth System. *Science,* Vol. 324, pp. 481–484.

Brandtberg, T. 2007. Classifying Individual Tree Species under Leaf-Off and Leaf-On Conditions Using Airborne Lidar. *ISPRS Journal of Photogrammetry and Remote Sensing,* Vol. 61, pp. 325–340.

Burkhart, H. E., T. E. Avery, and B. P. Bullock. 2019. *Forest Measurements* (6th edition). Long Grove, IL: Waveland Press, 434 pp.

California Department of Forestry and Fire Protection. 2020. King Fire. *www.fire.ca.gov/incidents/2014/9/13/king-fire.*

Carter, M. C., and C. D. Foster. 2006. Milestones and Millstones: A Retrospective on 50 Years of Research to Improve Productivity in Loblolly Pine Plantations. *Forest Ecology and Management,* Vol. 227, pp. 137–144.

Chen, W., H. Yin, K. Moriya, T. Sakai, and C. Cao. 2017. Retrieval and Comparison of Forest Leaf Area Index Based on Remote Sensing Data from AVNIR-2, Landsat-5 TM, MODIS, and PALSAR Sensors. *ISPRS International Journal of Geo-Information,* Vol. 6, Article No. 179, 16 pp.

Christensen, B. R. 2015. Use of UAV or Remotely Piloted Aircraft and Forward-Looking Infrared in Forest, Rural and Wildland Fire Management: Evaluation Using Simple Economic Analysis. *New Zealand Journal of Forestry Science,* Vol. 45, No. 1, Article No. 16, 9 pp.

Collins, W. 1978. Remote Sensing of Crop Type and Maturity. *Photogrammetric Engineering and Remote Sensing,* Vol. 44, pp. 43–55.

Condit, H. R. 1970. The Spectral Reflectance of American Soils. *Photogrammetric Engineering,* Vol. 36, pp. 955–966.

Eidenshink, J. C. 1992. The 1990 Conterminous U.S. AVHRR Data Set. *Photogrammetric Engineering and Remote Sensing,* Vol. 58, pp. 809–813.

Erdody, T. L., and L. M. Moskal. 2010. Fusion of LiDAR and Imagery for Estimating Forest Canopy Fuels. *Remote Sensing of Environment,* Vol. 114, pp. 725–737.

Fernow, B. E. 1907. *A Brief History of Forestry in Europe, the United States and Other Countries*. Toronto: University of Toronto Press, 438 pp.

Flannigan, M. D., Krawchuk, M. A., de Groot, W. J., Wotton, B. M., and Gowman, L. M. 2009. Implications of Changing Climate for Global Wildland Fire. *International Journal of Wildland Fire,* Vol. 18, No. 5, p. 483.

Ganguly, S., R. R. Nemani, G. Zhang, H. Hashimoto, C. Milesi, A. Michaelis, . . . R. B. Myneni. 2012. Generating Global Leaf Area Index from Landsat: Algorithm Formulation and Demonstration. *Remote Sensing of Environment,* Vol. 122, pp. 185–202.

Gopalakrishnan, R., V. A. Thomas, R. H. Wynne, J. W. Coulston, and T. R. Fox. 2018. Shrub Detection Using Disparate Airborne Laser Scanning Acquisitions over Varied Forest Cover Types. *International Journal of Remote Sensing,* Vol. 39, pp. 1220–1242.

Guanter, L., H. Kaufmann, K. Segl, S. Foerster, C. Rogass, S. Chabrillat, . . . and B. Sang. 2015. The enMAP Spaceborne Imaging Spectroscopy Mission for Earth Observation. *Remote Sensing,* Vol. 7, pp. 8830–8857.

Helms, J. A. 1998. *The Dictionary of Forestry.* Bethesda, MD: Society of American Foresters, 210 pp.

Huang, D., Y. Knyazikhin, R. E. Dickinson, M. Rautiainen, P. Stenberg, M. Disney, . . . and R. B. Myneni. 2007. Canopy Spectral Invariants for Remote Sensing and Model Applications. *Remote Sensing of Environment,* Vol. 106, pp. 106–122.

Hytönen, J., and P. Jylhä. 2005. Effects of Competing Vegetation and Post-Planting Weed Control on the Mortality, Growth and Vole Damages to *Betula pendula* Planted on Former Agricultural Land. *Silva Fennica,* Vol. 39, No. 3, 365–380.

Intini, P., E. Ronchi, S. Gwynne, and N. Bénichou. 2020. Guidance on Design and Construction of the Built Environment Against Wildland Urban Interface Fire Hazard: A Review. *Fire Technology,* Vol. 56, 1853–1883.

Jensen, A. M., and M. Löf. 2017. Effects of Interspecific Competition from Surrounding Vegetation on Mortality, Growth and Stem Development in Young Oaks (*Quercus robur*). *Forest Ecology and Management,* Vol. 392, pp. 176–183.

Kaufman, Y. J., and D. Tanre. 1992. Atmospherically Resistant Vegetation Index (ARVI) for EOS-MODIS. *IEEE Transactions on Geoscience and Remote Sensing,* Vol. 30, pp. 261–270.

Ke, Y., L. J. Quackenbush, and J. Im. 2010. Synergistic Use of QuickBird Multispectral Imagery and Lidar Data for Object-Based Forest Species Classification. *Remote Sensing of Environment,* Vol. 114, No. 6, pp. 1141–1154.

Knight, T. C., A. W. Ezell, D. R. Shaw, J. D. Byrd, and D. L. Evans. 2004. Identifying Loblolly Pine and Four Common Competing Hardwood Species Using Multispectral Reflectance Analysis (General Technical Report No. SRS-71). Asheville, NC: USDA Forest Service.

Körner, C., D. Basler, G. Hoch, C. Kollas, A. Lenz, C. F. Randin, . . . and N. E. Zimmermann. 2016. Where, Why and How? Explaining the Low-Temperature Range Limits of Temperate Tree Species. *Journal of Ecology,* Vol. 104, pp. 1076–1088.

Korpela, I., T. Tuomola, T. Tokola, and B. Dahlin. 2008. Appraisal of Seedling Stand Vegetation with Airborne Imagery and Discrete-Return LiDAR—An Exploratory Analysis. *Silva Fennica,* Vol. 42, pp. 753–772.

Kraus, C., D. Yaffe-Bellany, and M. Simões. 2019, October 10. Why Amazon Fires Keep Raging 10 Years after a Deal to End Them. *The New York Times,* p. 8.

Lasaponara, R., A. Lanorte, and S. Pignatti. 2006. Characterization and Mapping of Fuel Types

for the Mediterranean Ecosystems of Pollino National Park in Southern Italy by Using Hyperspectral MIVIS Data. *Earth Interactions,* Vol. 10, Article No. 13, 11 pp.
Lauer, D. K., and G. R. Glover, 1999. Stand Level Pine Response to Occupancy of Woody Shrub and Herbaceous Vegetation. *Canadian Journal of Forest Research,* Vol. 29, pp. 979–984.
Leite da Silva, R. M., R. E. Hakamada, J. H. Bazani, M. S. Gentil Otto, and J. L. Stape. 2016. Fertilization Response, Light Use, and Growth Efficiency in Eucalyptus Plantations across Soil and Climate Gradients in Brazil. *Forests,* Vol. 7, Article No. 117.
Lefsky, M. A., W. B. Cohen, S. A. Acker, G. G. Parker, T. A. Spies, and D. Harding. 1999. Lidar Remote Sensing of the Canopy Structure and Biophysical Properties of Douglas-Fir Western Hemlock Forests. *Remote Sensing of Environment,* Vol. 70, pp. 339–361.
Lentile, L. B., Z. A. Holden, A. M. S. Smith, M. J. Falkowski, A. T. Hudak, P. Morgan, . . . and N. C. Benson. 2006. Remote Sensing Techniques to Assess Active Fire Characteristics and Post-Fire Effects. *International Journal of Wildland Fire,* Vol. 15, pp. 319–345.
Lim, K., P. Treitz, K. Baldwin, I. Morrison, and J. Green. 2003. Lidar Remote Sensing of Biophysical Properties of Tolerant Northern Hardwood Forests. *Canadian Journal of Remote Sensing,* Vol. 29, pp. 658–678.
Martin, M. E., L. C. Plourde, S. V. Ollinger, M. L. Smith, and B. E. McNeil. 2008. A Generalizable Method for Remote Sensing of Canopy Nitrogen across a Wide Range of Forest Ecosystems. *Remote Sensing of Environment,* Vol. 112, pp. 3511–3519.
Mayor, J. R., S. J., Wright, and B. L. Turner. 2014. Species-Specific Responses of Foliar Nutrients to Long-Term Nitrogen and Phosphorus Additions in a Lowland Tropical Forest. *Journal of Ecology,* Vol. 102, pp. 36–44.
Mutlu, M., S. Popescu, C. Stripling, and T. Spencer. 2008. Mapping Surface Fuel Models Using Lidar and Multispectral Data Fusion for Fire Behavior. *Remote Sensing of Environment,* Vol. 112, pp. 274–285.
National Academies of Sciences, Engineering, and Medicine. 2018. *Thriving on Our Changing Planet: A Decadal Strategy for Earth Observation from Space*. Washington, DC: National Academies Press, 694 pp.
Nayegandhi, A., and J. C. Brock. 2008. Assessment of Coastal Vegetation Habitats Using Airborne Laser Remote Sensing. In X. Yang (ed.), *Lecture Notes in Geoinformation and Cartography—Remote Sensing and Geospatial Technologies for Coastal Ecosystem Assessment and Management*. New York: Springer, pp. 365–389.
Nyland, R. D., L. S. Kenefic, K. K. Bohn, and S. L. Stout. 2016. *Silviculture: Concepts and Applications* (3rd ed.). Long Grove, IL: Waveland Press, 680 pp.
Oswald, B. P., J. T. Fancher, D. L. Kulhavy, and H. C. Reeves. 1999. Classifying Fuels with Aerial Photography in East Texas. *International Journal of Wildland Fire,* Vol. 9, pp. 109–113.
Paloscia, S. 1998. An Empirical Approach to Estimating Leaf Area Index from Multifrequency SAR Data. *International Journal of Remote Sensing,* Vol. 19, No. 2, pp. 359–364.
Peduzzi, A., H. L. Allen, and R. H. Wynne. 2010. Leaf Area of Overstory and Understory in Pine Plantations in the Flatwoods. *Southern Journal of Applied Forestry,* Vol. 34, pp. 154–160.
Peduzzi, A., R. H. Wynne, V. A. Thomas, R. F. Nelson, J. J. Reis, and M. Sanford. 2012. Combined Use of Airborne Lidar and DBInSAR Data to Estimate LAI in Temperate Mixed Forests. *Remote Sensing,* Vol. 4, pp. 1758–1780.
Pettinari, M. L., and E. Chuvieco. 2016. Generation of a global fuel data set using the Fuel Characteristic Classification System. *Biogeosciences,* Vol. 13, No. 7, pp. 2061–2076.
Pinchot, G. 1893. *Biltmore Forest, the Property of Mr. George W. Vanderbilt: An Account of Its Treatment, and the Results of the First Year's Work*. Chicago: Lakeside Press, 60 pp.
Pinchot, G. 1914. *The Training of a Forester* (3rd ed., rev.). Philadelphia: J. B. Lippincott, 62 pp.
Powers, R. F. 1999. On the Sustainable Productivity of Planted Forests. *New Forests,* Vol. 17, pp. 263–306.
Prasad, S., and C. C. Brooks. 1989. Estimating Phytomass Productivity of Weed Species Associ-

ated with Young Pine Plantations Utilizing Remotely Sensed Reflectance Data. In *12th Canadian Symposium on Remote Sensing Geoscience and Remote Sensing Symposium*. Vancouver, Canada, pp. 1347–1348.

Reich, P. B. 2012. Key Canopy Traits Drive Forest Productivity. *Proceedings of the Royal Society B,* Vol. 279, pp. 2128–2134.

Riccardi, C. L., R. D. Ottmar, D. V. Sandberg, A. Andreu, E. Elman, K. Kopper, and J. Long. 2007. The Fuelbed: A Key Element of the Fuel Characteristic Classification System. *Canadian Journal of Forest Research,* Vol. 37, pp. 2394–2412.

Roberts, S. D., C. A. Harrington, and T. A. Terry. 2005. Harvest Residue and Competing Vegetation Affect Soil Moisture, Soil Temperature, N Availability, and Douglas-Fir Seedling Growth. *Forest Ecology and Management,* Vol. 205, pp. 333–350.

Rowell, E., E. L. Loudermilk, C. Hawley, S. Pokswinski, C. Seielstad, L. Queen, . . . and J. K. Hiers. 2020. Coupling Terrestrial Laser Scanning with 3D Fuel Biomass Sampling for Advancing Wildland Fuels Characterization. *Forest Ecology and Management,* Vol. 462, Article No. 117945.

Rubilar, R. A., H. L. Allen, T. R. Fox, R. L. Cook, T. J. Albaugh, and O. C. Campoe. 2018. Advances in Silviculture of Intensively Managed Plantations. *Current Forestry Reports,* Vol. 4, pp. 23–34.

Schroeder, W., P. Oliva, L. Giglio, and I. A. Csiszar. 2014. The New VIIRS 375m Active Fire Detection Data Product: Algorithm Description and Initial Assessment. *Remote Sensing of Environment,* Vol. 143, pp. 85–96.

Sengupta, S. 2020, February 5. How Europe Turned into a Perfect Landscape for Wildfires. *The New York Times,* p. A6.

Shin, P., T. Sankey, M. Moore, and A. Thode. 2018. Evaluating Unmanned Aerial Vehicle Images for Estimating Forest Canopy Fuels in a Ponderosa Pine Stand. *Remote Sensing,* Vol. 10, Article No. 1266.

Slesak, R. A., T. B. Harrington, and S. H. Schoenholtz. 2010. Soil and Douglas-Fir (*Pseudotsuga menziesii*) Foliar Nitrogen Responses to Variable Logging-Debris Retention and Competing Vegetation Control in the Pacific Northwest. *Canadian Journal of Forest Research,* Vol. 40, pp. 254–264.

Stankevich, S. A., A. A. Kozlova, I. O. Piestova, and M. S. Lubskyi. 2017, August 29–31. Leaf Area Index Estimation of Forest using Sentinel-1 C-band SAR data. IEEE Microwaves, Radar, and Remote Sensing Symposium, Kyiv, Ukraine.

Stenberg, P., M. Rautiainen, T. Manninen, P. Voipio, and H. Smolander. 2004. Reduced Simple Ratio Better Than NDVI for Estimating LAI in Finnish Pine and Spruce Stands. *Silva Fennica,* Vol. 38, pp. 3–14.

Stevenson, A., and C. A. Lindberg (eds.). 2010. *New Oxford American Dictionary* (3rd ed.). New York: Oxford University Press, 2096 pp.

Stevenson, A., and C. A. Lindberg (eds.). 2017. *New Oxford American Dictionary* (4th ed.). New York: Oxford University Press.

Sumnall, M. J., T. R. Fox, R. H. Wynne, C. E. Blinn, and V. A. Thomas. 2016a. Estimating Leaf Area Index at Multiple Heights within the Understorey Component of Loblolly Pine Forests from Airborne Discrete-Return Lidar. *International Journal of Remote Sensing,* Vol. 37, No. 1, pp. 78–99.

Sumnall, M., A. Peduzzi, T. R. Fox, R. H. Wynne, V. A. Thomas, and B. Cook. 2016b. Assessing the Transferability of Statistical Predictive Models for Leaf Area Index Between Two Airborne Discrete Return Lidar Sensor Designs within Multiple Intensely Managed Loblolly Pine Forest Locations in the Southeastern USA. *Remote Sensing of Environment,* Vol. 176, pp. 308–319.

Tarabay, J. 2020, January 21. Why These Australia Fires Are Like Nothing We've Seen Before. *The New York Times,* p. A8.

Tuominen, S., R. Näsi, E. Honkavaara, A. Balazs, T. Hakala, N. Viljanen, . . . and H. Ojanen. 2018. Assessment of Classifiers and Remote Sensing Features of Hyperspectral Imagery and Stereo-Photogrammetric Point Clouds for Recognition of Tree Species in a Forest Area of High Species Diversity. *Remote Sensing*, Vol. 10, No. 5, Article No. 714.

United States Forest Service. 2014. *Revised Land and Resource Management Plan, George Washington National Forest*. R8-MB 143 A.

United States Forest Service. 2017. *United States Forest Service Fiscal Year 2018 Budget Overview*. United States Department of Agriculture.

United States Forest Service. 2020a. *Fire Imaging Technologies for Wildland Fire Operations*.

United States Forest Service. 2020b. *Fuel Characteristics Classification System Tutorial*.

United States Forest Service, 2020c. *National Infrared Operations*. USDA Forest Service. *https://fsapps.nwcg.gov/nirops/pages/imagery*.

van Aardt, J. A. N., and R. H. Wynne. 2007. Examining Pine Spectral Separability Using Hyperspectral Data From an Airborne Sensor: An Extension of Field-Based Results. *International Journal of Remote Sensing,* Vol. 28, pp. 431–436.

Vanden Heuvel, R. M. 1996. The Promise of Precision Agriculture. *Journal of Soil and Water Conservation,* Vol. 51, pp. 38–40.

Varga, T. A., and G. P. Asner. 2008. Hyperspectral and LiDAR Remote Sensing of Fire Fuels in Hawaii Volcanoes National Park. *Ecological Applications,* Vol. 18, pp. 613–623.

Vargas, F., C. Gonzalez-Benecke, R. Rubilar, and M. Sanchez-Olate. 2018. Modelling the Effect of Weed Competition on Long-Term Volume Yield of *Eucalyptus globulus* Labill. Plantations across an Environmental Gradient. *Forests,* Vol. 9, Article No. 480.

Voipio, P., H. Smolander, P. Stenberg, T. Manninen, and M. Rautiainen. 2014. Reduced Simple Ratio Better than NDVI for Estimating LAI in Finnish Pine and Spruce Stands. *Silva Fennica,* Vol. 38, pp. 3–14.

Wagner, R. G., K. M. Little, B. Richardson, and K. McNabb. 2006. The Role of Vegetation Management for Enhancing Productivity of the World's Forests. *Forestry: An International Journal of Forest Research,* Vol. 79, pp. 57–79.

Walthert, L., and E. S. Meier. 2017. Tree Species Distribution in Temperate Forests Is More Influenced by Soil than by Climate. *Ecology and Evolution,* Vol. 7, pp. 9473–9484.

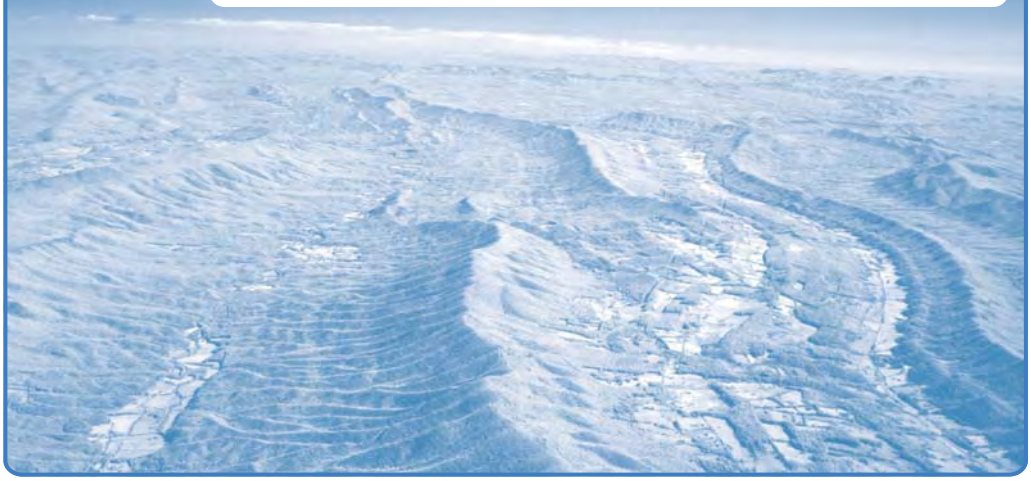
Xu, W., M. J. Wooster, J. He, and T. Zhang. 2020. First Study of Sentinel-3 SLSTR Active Fire Detection and FRP Retrieval: Night-Time Algorithm Enhancements and Global Intercomparison to MODIS and VIIRS AF products. *Remote Sensing of Environment,* Vol. 248, Article No. 111947.

Yin, R., and R. Sedjo. 2001. Is This the Age of Intensive Management? A Study of Loblolly Pine on Georgia's Piedmont. *Journal of Forestry,* Vol. 99, pp. 10–17.

Zhao, K., and S. Popescu. 2009. Lidar-Based Mapping of Leaf Area Index and Its Use for Validating GLOBCARBON Satellite LAI Product in a Temperate Forest of the Southern USA. *Remote Sensing of Environment,* Vol. 113, pp. 1628–1645.

Zhang, J., M. D. Busse, D. H. Young, G. O. Fiddler, J. W. Sherlock, and J. D. TenPas. 2017. Aboveground Biomass Responses to Organic Matter Removal, Soil Compaction, and Competing Vegetation Control on 20-Year Mixed Conifer Plantations in California. *Forest Ecology and Management,* Vol. 401, pp. 341–353.

19 Earth Sciences



MAJOR TOPICS TO UNDERSTAND

- Photogeology
- Galisteo Creek, New Mexico
- Drainage Patterns
- Lineaments
- Lidar's Contributions and Geoscience Information
- Mass Wasting and Debris Flows
- Stream Diversion
- Geobotany
- Direct Multispectral Observation of Rocks and Minerals
- Photoclinometry
- Band Ratios
- Soil and Landscape Mapping
- Integrated Terrain Units
- Wetlands Inventory
- Radar Imagery for Exploration

19.1 INTRODUCTION

This chapter addresses applications of remote sensing in the earth sciences, broadly defined to include geology, geomorphology, soil science, and related topics. Despite their many differences, these disciplines share a common focus on the Earth's shape and structure and the nature of the soils and sediments at its surface. Applications of remote sensing in the earth sciences can be challenging because the subjects of investigation are geologic structures, soil horizons, and other features entirely or partially hidden beneath the Earth's surface. Remote sensing cannot provide direct firsthand observation of these features, but rather gives an overhead perspective that provides the context for firsthand observations and allows for inference of certain characteristics. Furthermore, interpretations of geoscience information are frequently based on subtle distinctions in tone, texture, and spectral response. Even direct examination of many geologic and pedologic materials is subject to error and controversy, so it should be no surprise that applications of remote sensing in these fields can be equally difficult.

Remote sensing offers the ability to observe reflectance and emittance over a range of wavelengths, opening opportunities to study subjects that would not otherwise be possible. The synoptic view of satellite images provides broad-scale perspectives of patterns that are not normally discernible by ground observation. Furthermore, the ability of sequential imagery to record seasonal changes or the impacts of events (such as landslides, fires, and floods) also allows geoscientists to record and analyze them before their character is altered by subsequent events.

Finally, although remotely sensed images cannot replace field and laboratory studies, they can form valuable supplements to more traditional methods and sometimes provide information and perspectives that are not otherwise available.

Historical Context for Aerial Imagery on the Earth Sciences

Aerial imagery assumed an early role in the development and adoption of aerial observation for understanding terrain, physiography, and geomorphic systems. As mentioned in previous chapters, the airplane and camera initially existed as separate, largely unrelated, technologies that were integrated only with considerable effort. World War I formed the context for effective aerial observation, mainly through rapid experimentation and innovation to improve its effectiveness for understanding the battlefield. The firsthand experience of pilots and observers drove the rapid innovation in aerial observation, aerial photography, and accurate interpretation of aerial imagery.

As World War I concluded, aerial photography began its transformation from status as a military implement to a practical scientific tool for examining soils, geological features, coastlines, and river systems. In the interwar years, publications regarding aerial photography focused mainly on the practicalities of aviation equipment. Lee's (1922) outline of aerial photography's potential for civil applications forms a notable landmark in the documentation of aerial photography's role in acquiring data to investigate both social and physical systems.

By the late 1930s, use of aerial photography for geological inquiry was visible through development of aerial cameras, aircraft, and supporting services. Geologists recognized the value of aerial observation for teaching earth sciences, developing instruction in the fundamentals of earth science (including streams, wind, waves, and glaciers).

As World War II began, the belligerents developed training materials that may be considered the precursors of instructional materials, handbooks, and reference works for instruction in quantitative measurements and recognition of enemy vehicles, weapons, and infrastructure. In the United States especially, experienced pilots and photoanalysts were in a position to apply their experience in civil society. Authors such as Ray (1960) and Simpson (1966) prepared texts and reference materials that introduced civil society to the practical value of aerial imagery. Bauer (2004) considered geomorphology as based on landform description and classification, supported by the availability of remote sensing imagery to assess the following:

1. Placement of landforms and surface features
2. Land surface elevation
3. Land surface composition
4. Subsurface characterization

Millington and Townshend (1987) focused on the role of satellite remote sensing as a source of topographic data, and the value of advances in geomorphological inquiry. Smith and Pain (2009) amended Millington and Townshend's schematic diagram to provide a more current graphic (**Figure 19.1**). They noted that advances in geomorphological inquiry had progressed because of the availability of new technologies, specifically (1) the increasing availability of digital elevation models and (2) the introduction of hyperspectral imaging, radiometrics, and electromagnetics. Remote sensing also began to offer improved capacities in terms of close-range (<200 m) techniques for very high-resolution imagery.

19.2 PHOTOGEOLOGY

Geologists study many aspects of the Earth's surface in an effort to understand its structure, to guide the search for minerals and fuels, and to assess geologic hazards. Remote sensing contributes to several dimensions of the geological sciences by providing information concerning lithology, structure, and vegetation patterns. *Lithology* refers to the fundamental physical and chemical properties of rocks, including, for example, the gross distinctions between sedimentary, igneous, and metamorphic rocks. *Structure* defines the kinds of deformation experienced by rocks, including folding, fracturing, and faulting. *Geobotanical studies* focus on relationships between plant cover at the Earth's surface and the lithology of underlying rocks.

Photogeology is the derivation of geological information from interpretation of aerial photography, originating early in the development of aerial photo interpretation. Many of its basic techniques were developed in the 1920s, and then refined and applied into the 1950s and 1960s, when they approached the limits of their capabilities, to be assimilated into newly developing research in geological remote sensing (Ray, 1960). Today, photogeology is routinely applied to good effect. However, because aerial photographs now form only one of many forms of aerial imagery routinely available to geologists, most research and innovation is now likely to occur in the broader context of geological remote sensing.

The practice of photogeology is based on direct application of principles of image interpretation (Chapter 6) to geological problems. Image texture, size, shape, tone, and shadow all continue to have special significance for the geologist's view of terrain. Like-

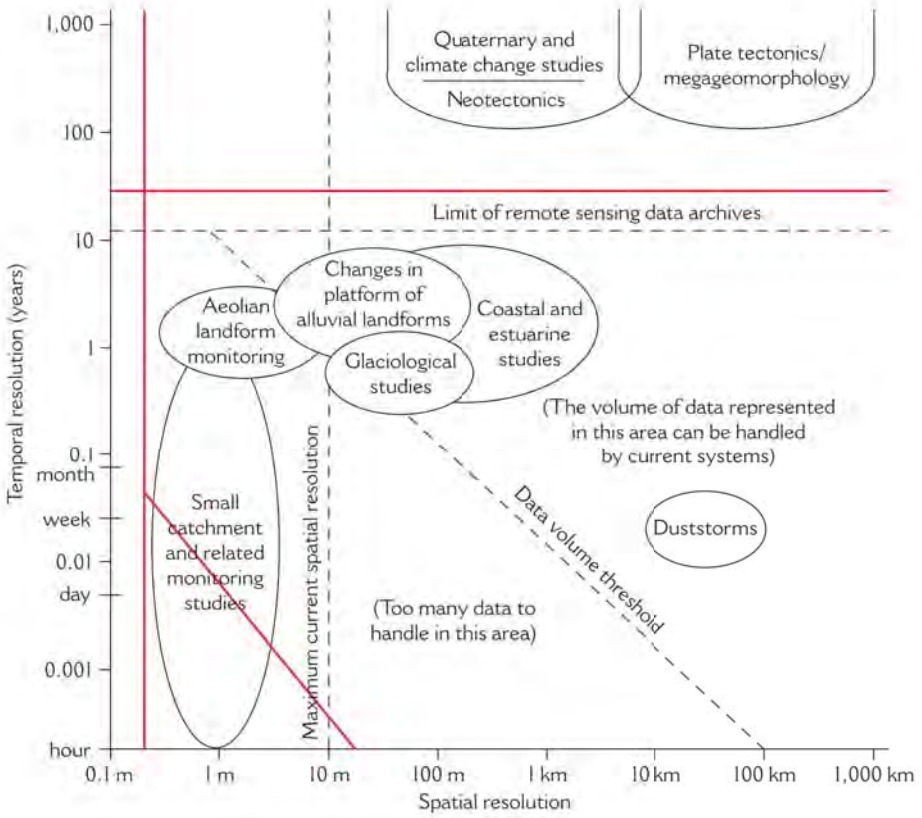


FIGURE 19.1 Changing relationships between spatial and temporal scales as they apply to applications of remotely sensed data for geomorphic analysis. Dashed lines represent thresholds they reported in 1987. Red lines signify current thresholds depicting the expansion of data archives in recent years and improved resolution and data-handling capabilities. From M. J. Smith and C. F. Pain. Used by permission of SAGE.

wise, the principles of photogrammetry (Chapter 4) have special applications in the context of photogeology for calculating the thicknesses of beds and determining strike and dip from aerial photography. Such measurements provide the capability to derive structural information from aerial images.

19.3 GALISTEO CREEK, NEW MEXICO

Remotely sensed images can provide valuable information about structure and terrain, especially if stereoscopic views are available. As an illustration, Figure 19.2 presents a stereo image of an arid landscape in New Mexico (near Galisteo Creek), characterized by a history of tectonic activity and related volcanic events. This landscape, positioned east of the Rio Grande between Santa Fe and Albuquerque, experienced extensive volcanism during the Quaternary and late Tertiary.

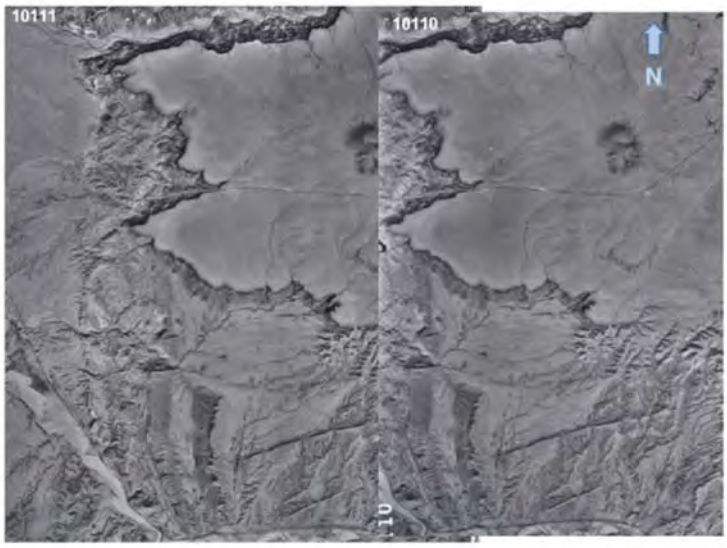


FIGURE 19.2 Stereo aerial photography of the terrain near Galisteo Creek region, north central New Mexico (35° 29.71212′ N, 105° 9.228′ W). From U.S. Geological Survey (USGS) and Army Map Service.

Our image was acquired in 1954. (Recent photographs of this region are available using imagery displayed by Google Earth [35° 29.71212′ N, 105° 9.228′ W]). The following paragraphs highlight some of the distinctive landscape features of this region, illustrating the value of the aerial perspective in understanding the geospatial landscape. Figure 19.3 provides an annotated version of the image, marked to label distinctive features discussed below.

A. **Hogbacks:** These features, known as *hogbacks* (i.e., resistant strata upended to form steeped-sided ridges with jagged crests) are characterized by resistant sandstones forming the jagged ridges, with less resistant shales eroded to form intervening valleys. Figure 19.4 presents a ground view of the hogbacks, taken from a position near the center lower edge of Figure 19.3 (south of the site of the yellow "A").
B. **Volcanic dikes:** These photographs also show evidence of the volcanism that has shaped this landscape. *Volcanic dikes* cut across established topography as narrow, pronounced, steep-sided ridges. Notice that, in Figure 19.3, these dikes cut across north–south-oriented ridges, confirming the recent age of the volcanism relative to the other structures.
C. **Braided streams:** *Braided streams,* a common feature of arid regions, are characterized by varied flow, alternating between a mainly dry riverbed, with minimal flow, and in contrast, high, rapidly flowing currents that can transport large amounts of sediment, coarser stones, and gravels.
D. **Bajada:** A *bajada* is a broad slope of alluvial material at the foot of an escarp-

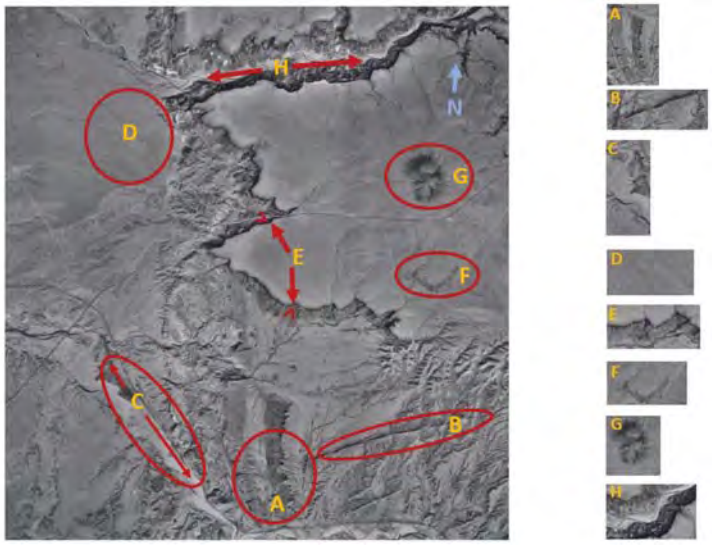


FIGURE 19.3 Annotated aerial photography of the terrain near Galisteo Creek, New Mexico, 1954, labeled to identify some of the distinctive landforms of this region. Explanations of the labels are found in the text. From USGS and Army Map.

FIGURE 19.4 Ground-level photograph of terrain near the Galisteo Creek region, looking northward across the creek toward the hogback visible in the lower center of **Figures 19.2 and 19.3.** Galisteo Creek (a braided stream) is visible in the foreground, flowing from right to left (at low flow stage). The cinder cone marked as "G" in **Figure 19.3**. is recognizable as the tiny feature visible at the right-hand horizon. From Photographic Library, USGS. Photograph by W. T. Lee.

ment, formed in this example by transport of alluvial sediment weathered from the mesa, and eroded sediment transported from the canyon, as marked as "H."

E. **Mesa cliffs:** A *mesa* is an elevated, flat-topped landform, with steep, cliff-like sides, and with an upper, highly resistant, flat surface. Viewed in stereo, the resistant character of the basaltic surface is evident from its sharp relief and the precipitous cliffs at its margins. The cliffs are distinctive for their abrupt, sharp, edges. As viewed in stereo, the thick, rigid, character of the basaltic surface is evident. Below, it is clear that lower strata have been subject to weathering and fracturing, undermining the stronger basalt, leading to collapse of large basaltic blocks (barely) visible at the base of the cliff. (See the markers near the tips of arrows marked at "E" in Figure 19.3.)

F. **Mesa surface:** The *basaltic mesa* surface, as viewed on the aerial photograph, has a distinctive appearance, even in tone, slightly rough in texture, and with a puffy appearance when viewed in stereo. "F" marks one of several stabilized lava flows (created where the movement of molten cooled to freeze in place). Note that the mesa surface is characterized by several such features, not all of which are immediately recognizable. Using Google Earth, we can view where the flow of the mesa's sediment has covered some of the north–south-oriented sediments, indicating the recent origins of the mesa relative to underlying sediments.

G. **Cinder cone:** More or less at the center right of the mesa (marked as "G" in Figure 19.3) is a *cinder cone,* a remnant from events that released the flow of lava, creating the cinder cone and the discharge of lava over the mesa surface.

H. **Canyon:** Deep channels such as this can provide protection from intense heat and have a reliable supply of water. Downstream, the water is used for crops.

At the base of the cliff, the aerial photograph shows contrasting textures of a thick sheet of stone, sand, and gravel, shaped by fluvial processes that transport geologic debris across the slope at the base of the cliff toward the stream flowing right to left across the photograph. This stream has maintained its course across the hogbacks as they were formed, carving a distinct V-shaped notch through the right-hand ridge. As the stream passes through the notch, it enters a terrain formed by erosion of shales exposed at the surface. Here, surface erosion is much more effective than on the surfaces of the basaltic flows, for example, and the topography is much more disorganized and chaotic relative to the surface visible on the left-hand side of the image.

Although discussion of this example is necessarily quite brief, it illustrates the power of aerial imagery to provide information and insight to the Earth scientist, both to derive new information and to organize and interpret information already at hand. Subsequent sections explore some other dimensions of photogeology and the value of remotely sensed imagery for the earth sciences.

19.4 DRAINAGE PATTERNS

Drainage patterns are often the most clearly visible features on remotely sensed imagery, and they are also among the most informative indicators of surface materials and processes. The character of drainage patterns permits geologists to infer valuable information about the surface materials and the geologic structure that prevail for a specific

landscape (Figures 19.5 and 19.6). Drainage patterns as observed in nature may display characteristics of two or more of these patterns, so they may not match precisely to illustrations used to describe them.

A common drainage pattern is characterized by even branching of tributaries, similar to patterns of veins on an oak or maple leaf—a *dendritic* drainage pattern. Dendritic patterns suggest uniformly resistant surface materials, gentle regional slopes, and the absence of major faults or structural systems. *Parallel* drainage is also found on uniform materials, but typically on landscapes that have pronounced regional slopes. Parallel drainage resembles the dendritic pattern but is recognizable by its elongated form derived from the increased typographic slope. In fact, many transitional forms fall between these two forms.

If a landscape has been uplifted by tectonic forces or if base level has been lowered (e.g., by a decline in sea level), then streams may become *entrenched,* or *incised,* into the landscape such that their channels are cut well below the surface of the surrounding landscape. The nature of the incision can reveal the character of the bedrock—sharp, deep, well-defined edges suggest the presence of strong, cohesive surface materials. More gently sloping terrain near the channel suggests the presence of less cohesive, weaker strata.

If the landscape is characterized by linear structural or lithologic features, drainage often develops to form a *trellis* pattern. Tributaries often follow the strike of the structure, whereas the main streams cut across the principal structures. Therefore, tributaries join main streams at right angles to form a system in which the main streams and the tributaries are oriented perpendicular to each other. When a landscape is dominated by a large central peak, the drainage system is organized to drain water away from the central upland, forming a *radial* pattern.

When landscapes are disturbed by broad-scale processes such as *glaciation, faulting, surface mining,* or *volcanic deposition,* often buried or disturbed drainage systems are established. Such landscapes form *incoherent drainage* systems—lakes and marshes connected by a confused, chaotic system of streams without the organized stream patterns that we expect to encounter in most regions. As time passes, erosion and deposition will reestablish the familiar stream networks, but in the meantime, the drainage system will be chaotic and disorganized.

Finally, *braided streams* have several channels that divide and rejoin, separated by ephemeral bars and islands. Braided streams are often found in locales that have variable stream discharge, such as those of arid regions, or are supplied by glacial meltwater (which may have uneven discharge) and high sediment loads. Their appearance is usually quite distinctive—a wide, sparsely vegetated strip of open sand or gravel, with a network of ephemeral, anastomosing channels and elongated bars and islands.

Closer inspection of details of the drainage system can reveal specifics of the textures of the surface materials. Steep slopes are characteristic of coarse, loose, surface materials. Noticeably flat, rounded slopes may indicate cohesive, plastic surface materials. The absence of surface drainage or a very simple drainage pattern often indicates well-drained pervious soil. Highly integrated drainage patterns often indicate impervious, plastic sediments that lose their strength when wet.

The shape of the cross section of a drainage way is controlled largely by the cohesiveness of surface materials. Abrupt changes in stream grade, channel direction, or its cross section indicate changes in the underlying bedrock or the surface materials. Generally, short, V-shaped gullies with steep gradients are typical of well-drained, loosely structured sediments such as coarse sands. U-shaped stream profiles indicate the presence

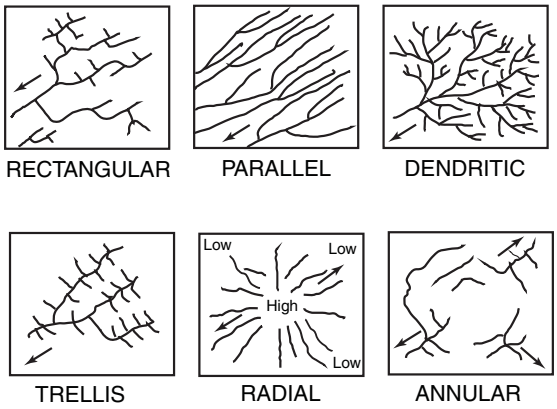


FIGURE 19.5 Sketches of varied drainage patterns. Note that each pattern reveals, without the underlying topography, relief information. Image by Susmita Sen.

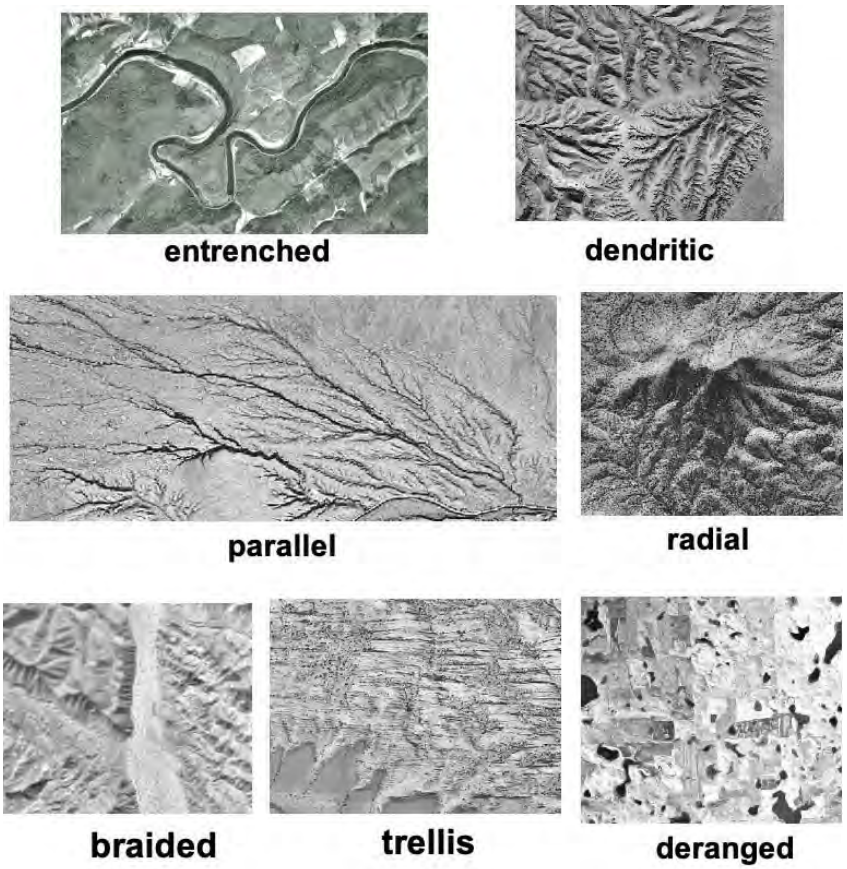


FIGURE 19.6 Aerial images of varied drainage patterns. Note that each pattern reveals the underlying topography. From USGS.

of deep, uniform, silt deposits, especially the wind-deposited sediments known as *loess*. Poorly drained, fine-textured, surface materials usually form shallow drainage ways with shallow, rounded saucer-like profiles.

On black-and-white aerial photography, tones of surface materials can indicate the character of the drainage. Soft, light tones generally indicate pervious, well-drained, soils, often characterized by light gray, uniform color tones, a flat appearance, and a lack of conformity; such patterns indicate a natural surface drainage. Clays and organic soils frequently appear as dark gray to black. In general, a sharp change in colors or tones signals a change in soil texture.

19.5 LINEAMENTS

Lineaments is the name geologists give to lines or edges of presumed geologic origin, visible on remotely sensed images. Such features have also been referred to as *linears* or *lineations,* although O'Leary, Friedman, and Pohn (1976) establish *lineaments* as the preferred term. Use of the term *lineament* in a geological context dates to 1904 and apparently has even earlier analogs in other languages. These early uses, prior to the availability of aerial images, applied to specific geologic or geomorphic features, such as topographic features (ridgelines, drainage systems, or coastlines), lithologic contacts, or zones of fracture.

As early as the 1930s, photogeologists studied fracture patterns visible on aerial photographs as a means of inferring geologic structure. These photo features apparently corresponded rather closely to faults and fractures defined in the field. More recently, in the context of geological remote sensing, the term has assumed a broader meaning. Any linear feature visible on an aerial image can be referred to as a *lineament*. A problem arises because it is sometimes difficult to establish a clear link between the features on images and corresponding features, if any, on the ground.

The strength of this link depends in part on the nature of the imagery. In the instance of interpretations from aerial photographs and photomosaics, scales may have fine detail, and linear features are likely to match to features that can be confirmed in the field. However, more extensive and subtle features cannot be easily detected on broad-scale imagery. Each photograph shows only a small area; if mosaics are formed to show larger regions, variations in illumination and shadowing obscure more subtle lineaments that might extend over many photographs. Furthermore, in the era prior to routine availability of the multispectral imagery, high-altitude photographs (which *can* show large areas under uniform illumination) were not of good visual quality due to the effects of atmospheric scattering.

The advent of nonphotographic sensors and the availability of the broad-scale view of satellite images changed this situation because it then became possible to view large areas with images of good visual clarity. Early radar images depicted large areas illuminated from a single direction at rather low-illumination angles—conditions that tended to increase topographic shadowing in a manner that enhanced the visibility of linear features (Figure 19.7). The locations of these features did not always match previously known faults or fractures, so the term *lineament* was evoked to avoid an explicit statement of a geological origin. Later, similar features were again detected on Landsat Multispectral Scanner System (MSS) imagery. Initially, the detection of lineaments and the interpretation of their meaning generated some controversy; the broad scale and coarse

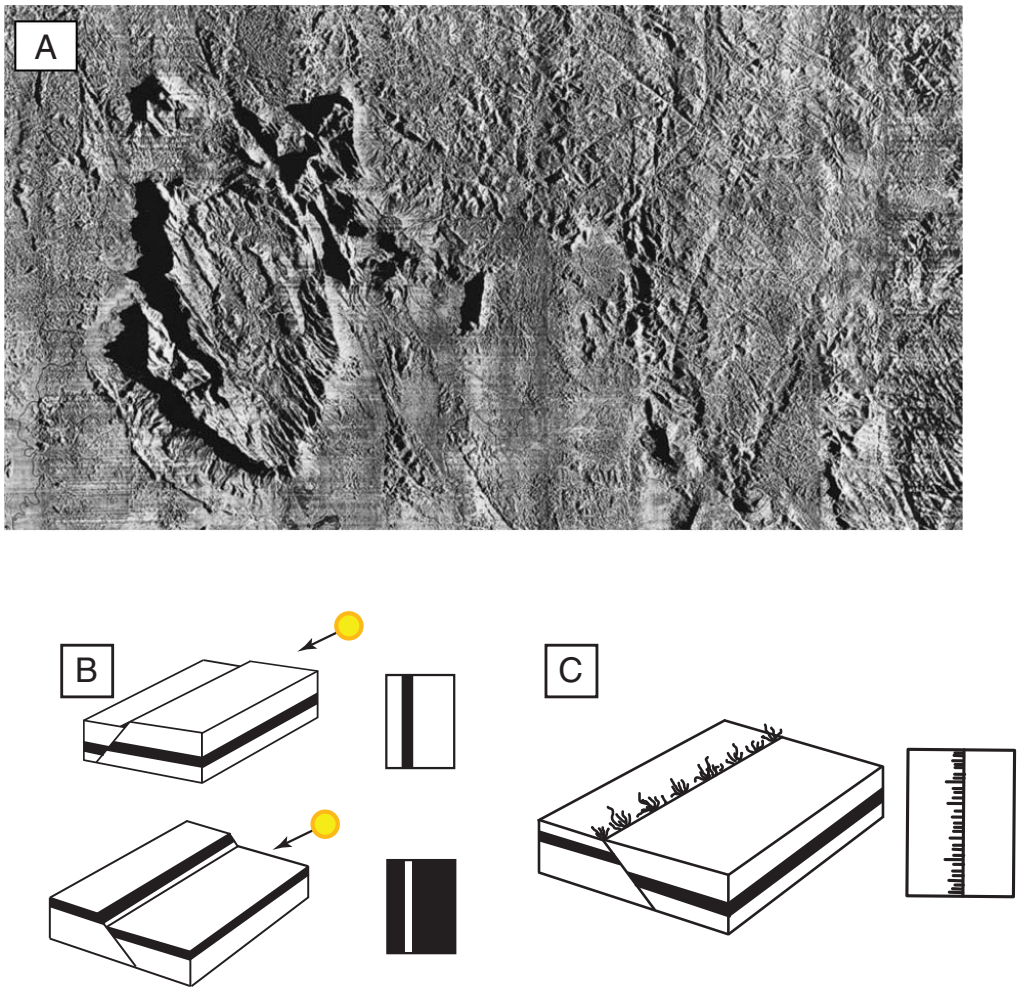


FIGURE 19.7 Synthetic aperture radar (SAR) mosaic of southern Venezuela, illustrating clear patterns of topography and drainage. (A) Prominent linear features include lineaments, often related to tectonic forces. From Goodyear Aerospace. (B) Linear features alternatively highlighted by shadowing and by solar illumination. Image by Susmita Sen. (C) Vegetation aligned with subsurface structure features can create linear surface features. Image by Susmita Sen.

resolution of the early satellite imagery caused lineaments to be subtly expressed and difficult to relate to previously known geological features. During the intervening years, investigation and debate have confirmed the value of lineaments as guides to understanding geologic phenomena.

Lineaments are, of course, "real" features if we consider them simply to be linear features detected on aerial imagery. The uncertainty arises when we attempt to judge their geological significance. There are sound reasons for assigning a geological meaning to some lineaments, even if they do not always correspond to clearly observable physical features at a specific point. In the simplest instances, a dip-slip fault may leave a subtle

topographic feature that is visible as a linear shadow on aerial photography when illuminated obliquely from the elevated side (Figure 19.7). The axis of displacement for a dip-slip fault is along the dip of the fault plane, to create a difference in relief, as depicted in Figure 19.7. In the alternative instance, a strike-slip fault, the axis of the fault displacement follows the strike of the fault (along the trace of the fault on the surface), creating fewer opportunities for surface expression. Of course, a fault that is oriented parallel to the direction of illumination may be indistinct or invisible. Not all faults are expressed topographically, but the fault plane may offer preferred avenues of movement for moisture and for the growth of plant roots (Figure 19.7c). Therefore, the trace of the fault may be revealed by vegetation patterns, even though it may have no obvious topographic expression. Faults of any form can alter drainage in a manner that creates linear drainage segments, which are then visible on aerial images. Such features, when clearly expressed, may have structural origins, and it may be possible to verify their existence through field observations. However, more subtle features may be observable only as broad-scale features, not easily verifiable at a given location. Such lineaments may be genuine structural features but not easily confirmable as such.

Linear features that are clearly of structural origin are significant because they indicate zones of fracturing and faulting. It is often assumed that regions of intersection of lineaments of differing orientation are of special significance, as in theory they might identify zones of mineralization, stratigraphic traps, regions of abundant groundwater, or zones of structural compression. For this reason, much is often made of the orientation and angles of intersection of lineaments.

Other lineaments may not have structural origins. It is conceivable that some may be artifacts of the imaging system or of preprocessing algorithms. In Figure 19.7, note the numerous human-made linear features. Some linear features may be purely surface features, such as wind-blown sediments, that do not reflect subsurface structure (Figure 19.8). Cultural patterns, including land cover boundaries, edges of land ownership parcels, and political borders, can all have linear form and can be aligned in a manner that creates the linear features observed on aerial images (Figure 19.7). Because these features may not be readily distinguished from those of geologic origin, and because of the inconsistency of individual delineations of lineaments from the same image, lineament analysis has been regarded with skepticism by a significant proportion of the geological community (Wise, 1982).

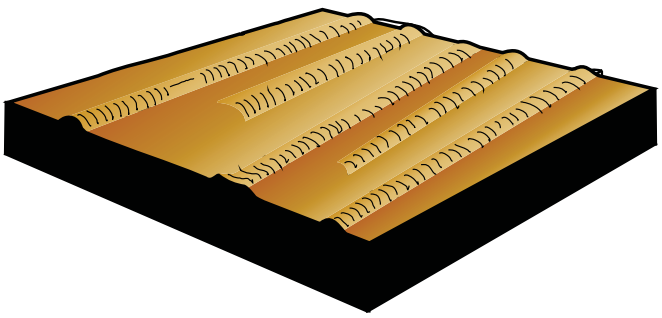


FIGURE 19.8 Linear surface features unrelated to geologic structure (in this example—sand dunes). Image by Susmita Sen.

Because the human visual system tends to generalize, manual interpretations may identify continuous linear features that are later found to be formed of separate segments of unrelated origin. Lineaments that are found to correspond to geologic features often are extensions of known fault systems rather than completely new systems. Results of manual interpretations vary greatly from interpreter to interpreter (Podwysocki, Moik, and Shoup, 1975). As a result, other research has attempted to automate the interpretation of lineaments. Because detection of edges and lines has long been an important task in the fields of image analysis and pattern recognition, the fields of image processing and machine vision were well equipped with techniques to apply to this inquiry. There is a rich literature that has investigated procedures for enhancing images to highlight linear features and then has analyzed patterns to extract linear features that might be present.

Fracture Density Studies

Often, images and field data can be analyzed to assess the abundance of lineaments by plotting their occurrence within cells used to partition the image. Analysis of the distribution and orientation of lineaments and fractures, in the context of field studies and other databases, can lead to assessment of groundwater or mineral potential.

Abrams et al. (1983) studied lineaments interpreted from a Landsat MSS color composite of a region in southwestern Arizona. Their analysis focused on the Helvetia-Rosemont area, known to be rich in porphyry copper. They attempted to determine whether their techniques would locate previously discovered mineralized zones. Paleozoic limestones, quartzites, shales, and dolomites cover Precambrian granites and schists in the study area. The Paleozoic deposits themselves were altered by Mesozoic uplift and erosion, and were then covered by further deposition. The Laramide orogeny during the late Cretaceous produced intrusion of granitic rocks with thrust faulting along a predominant northeast–southwest trend. This tectonic activity, together with later Paleocene intrusions, caused mineralization along some of the faults. Tectonic activity continued into the late Tertiary, producing extensive faulting and a complex pattern of lineaments.

Abrams et al. examined orientations of lineaments they interpreted from the Landsat image, and they then plotted their results in the form of a rose diagram (**Figure 19.9**). Each wedge in the circular pattern represents the number of lineaments oriented along

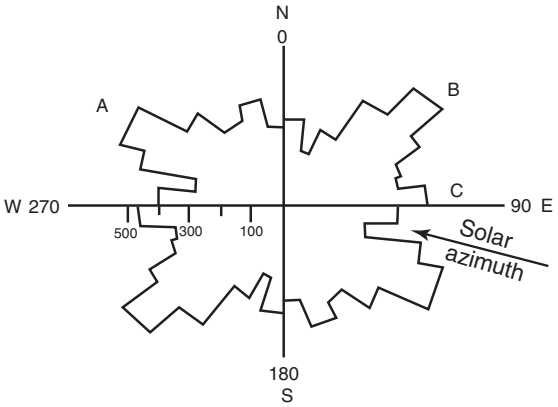


FIGURE 19.9 Strike-frequency diagram of lineaments for a region of southeastern Arizona. Features, such as discussed in the text, include (A) prominent northwest–southeast orientation; (B) prominent northeast–southwest orientation; and (C) low-frequency records of lineaments along the axis that parallels the solar beam at the time of image acquisition. Adapted from Abrams et al. (1983, p. 593) Image by Susmita Sen. Used by permission of the Society of Economic Geologists.

specific compass azimuths. Because each lineament has two azimuths (e.g., a "north–south" line is oriented as much to the south as it is to the north), the diagram is symmetric, and some investigators prefer to show only half of the diagram.

Several features are noteworthy. First, the low numbers of lineaments oriented at (approximately) 110 (and 290), roughly east/southeast to west/northwest, correspond to the axis that parallels the orientation of the solar beam at the time that the Landsat data were acquired. The lineaments that might have this orientation will not be easily detected because shadowing is minimized. The northeast–southwest trend (*B* in Figure 19.9) is said to represent the predominant trend of Precambrian faulting that is observed throughout Arizona. Superimposed over this pattern is the northwest–southeast trend (*A*) arising from faulting in the Mesozoic rocks mentioned above. Finally, the east–west trend (*C*) is interpreted as the result of the Laramide faulting known to be associated with the mineralization that produced the copper deposits present in this region. Abrams and his colleagues counted the frequencies of the lineaments in each cell of a 10-km grid superimposed over the image and concluded that the highest frequencies, which they interpreted to form favorable locations for intrusion of magmas and for mineralization, correspond to zones of known mineral deposits.

19.6 LIDAR'S CONTRIBUTIONS AND GEOSCIENCE INFORMATION

As described in Chapter 9, lidar (light detection and ranging) systems are based on technologies that transmit laser pulses, chiefly from airborne systems (also spaceborne, or ground-based vehicles) to measure distances to features on the Earth's surface to map terrain, vegetation, and structures. In modified form, lidar systems can be applied for bathymetric mapping to examine shallow coastal waters. Generally, use of lidar systems is relatively expensive but can provide valuable data.

Lidar instruments can acquire coordinates of ground features with submeter accuracy. In geoscience applications, the principal applications of lidar include mapping of geomorphic surfaces, including, for example, topographic features, debris flows, glacial features, alluvial fans, dune fields, and lava fields, especially when they are mobile, hazardous, or require frequent monitoring. Such features include flood mapping, landslides, debris flows, and earthquake damage.

Bathymetric lidars (designed for mapping shallow waters) are especially valuable for mapping submerged or emergent features. Although bathymetric lidars are ostensibly similar to conventional instruments, the marine environment presents numerous challenges that are not present in the conventional settings, including changing weather conditions, sea state, turbidity, subsurface vegetation, and varied water clarity.

Gypsum dunes. Figure 19.10 illustrates lidar data for New Mexico's gypsum dunes at White Sands National Monument, as observed by aerial photography and lidar imagery. The existing dunes sediments were formed by the shallow seas of the Permian Period, which covered the area now recognized as White Sands. These sediments were uplifted to form deposits in the nearby San Andreas and Sacramento Mountains. Rain dissolved the gypsum deposits, which were transported into the Tularosa Basin, including shallow gypsum deposits now covering over 275 square miles. The prevailing southwest winds, combined with the processes of weathering and erosion, eventually formed the white dunes. Lidar and aerial photography allow for measurements of the dune structure and their changing shape over time.

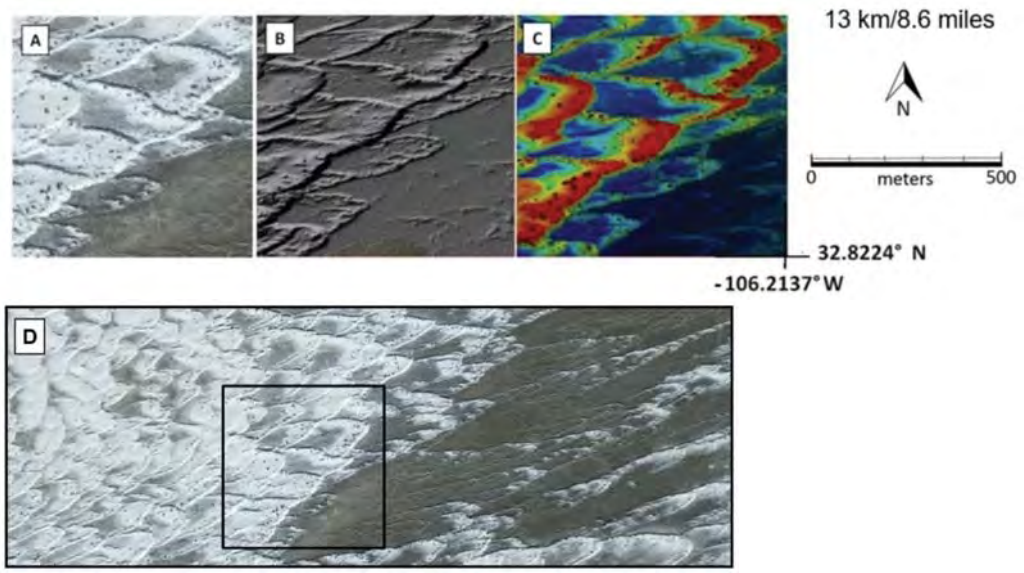


FIGURE 19.10 Gypsum dunes, White Sands National Monument (New Mexico), as observed by aerial photography and lidar. (A) Aerial image. (B) Hillshaded relief display. (C) Color-coded relief display (red represents highest elevations). (D) Broad-scale view of the region shown in the detailed insets. Lidar's detail and accuracy permit calculation of sand volume and, in the case of sequential imagery, measurement of rates of dune movement. From Google Earth, Open Topography. Used by permission.

Lidar data depicted in Figure 19.11 displays three-dimensional information acquired near the town of John Day, Oregon. The vegetation canopy conceals the underlying topography. The lidar data reveal the underlying debris flow, and its crescent-shaped cavity apparently reveals its origins as a rotational slide. The two large lobes at the base suggest a single sudden event, probably creating the two branches in the same event.

19.7 MASS WASTING AND DEBRIS FLOWS

Mass wasting (also sometime known as *mass movement*) refers to geomorphic processes that displace downslope, by gravitation force, materials such as unconsolidated rock, soil, sand, and regolith. A *landslide* generally refers to slow to rapid displacement of rock or soil. ***Debris flows*** are distinctive forms of mass wasting. Whereas many of the most commonly recognized mass wasting events displace (for example) coherent, rigid masses of solid or fractured rock, debris flows displace a mixture of saturated soil and fragmented rock, often channelized and effective in entraining nearby unconsolidated sediments. These events often share characteristics of landslides and the flow-like movements associated with sudden movement of water, snow and ice, and fine-grained sediments.

Debris flows typically occur without warning and are often associated with combinations of rainfall, snowmelt, landslides, or tectonic events (Ritter, Kochel, and Miller, 2011; Summerfield, 1991). Because of their sudden occurrence and unexpected power,

debris flows are especially hazardous. Here we present contrasting examples of debris flows that differ greatly with respect to size, age, location, and character.

The Landslide at Oso, Washington

In March 2014, the region at the south side of the Stillaguamish River in western Washington was flooded by the flow of its North Fork following prolonged rainfall. The steep terrain to the north of the channel failed, releasing mud and debris across the river channel to cover an area about one square mile in size. This event has been said to be the deadliest of its kind in U.S. history. It is often designated as the Oso Landslide in some official reports, or as the SR530 landslide by local jurisdictions in Snohomish County and Washington State. It is notable because of the amount of debris released by the event and the unusual speed of the failure. Forty-five days of heavy rain preceded the event, which damaged or destroyed 49 homes and related structures. The USGS report, *Revisiting the Oso Landslide* (2017) provides a review of the earlier history of events at this site, including instability.

Figure 19.12 is an oblique aerial photograph of the Oso Landslide in northwestern Washington. It depicts the event from an eastern perspective, showing the entire extent of the landslide source area and its path shortly after the event. The landslide and sizeable areas of open water are flooded by river water blocked by the landslide debris, having destroyed numerous dwellings and supporting structures.

Remote sensing has been used to study the impact of this event, with data from before and after the landslide. For example, scientists from the USGS used photogrammetry from aerial imagery to create high-resolution multiple digital elevation models (a

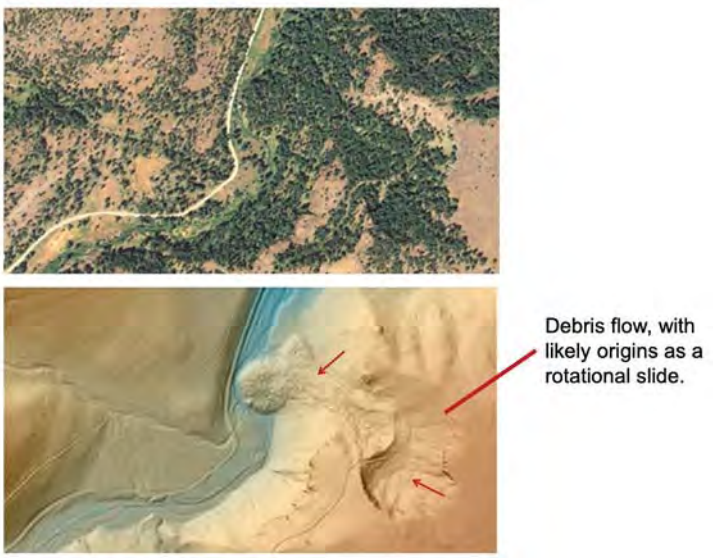


FIGURE 19.11 Lidar data of this region near John Day, Oregon, reveals a debris flow below a forest canopy. Photograph by the Oregon Department of Geology and Mineral Industries. Used by permission.

FIGURE 19.12 Oblique aerial photograph showing the impact of the Oso Landslide on the surrounding landscape (Snohomish County, northwest Washington, 2014). Note the areas of open water in the foreground, caused by flooding of the river water due to debris blocking the channel. From USGS. Photograph by Mark Reid.

technique referred to as *structure from motion*) over the flooding season after the event in order to quantify the erosion and sediment and assess changes in the channel over time.

Grand Mesa Debris Flow

Another major mass wasting event that can be clearly seen and monitored with remote sensing is the Grand Mesa debris flow, which occurred on the north-facing slope of Grand Mesa, Colorado, on May 25, 2014 (Figure 19.13). This was one of the largest recent debris flows in the United States, much larger than the Oso Landslide described above. Grand Mesa, at about 10,000 ft (3,048 m) above sea level and 6,000 ft (1,800 m) above Grand Junction, is often recognized as the world's largest mesa. Figure 19.13 provides a good example of the types of before-and-after remote sensing analysis possible when a natural disturbance has a large impact on the landscape. The area impacted can be clearly seen and measured/monitored over time.

The left-hand image in Figure 19.13 shows the same feature as it appeared in 2011. Here, vegetation (and coarser-resolution photography) have concealed details of earlier slides, but vegetation patterns appear to suggest the presence of earlier slides at the same location. This area has a known history of instability, due to the underlying geology and the accumulation of precipitation. Google Earth historical imagery of the same region

appears to suggest the presence of slides (although possibly not as large as those recorded on more recent imagery) as early as 1993, though not of the same magnitude.

Unseasonably heavy rains in early May 2014 melted local snowpack and saturated underlying soils. By late May, minor slope displacements were evident, increasing in magnitude. By May 25 a major failure of the headwall at the crest of the Mesa was triggered by snowmelt and heavy rains (Figure 19.14). These conditions caused a succession of failures that created a series of "rock avalanches," creating catastrophic failures over the 2.8-mile length of the valley (Figure 19.15). This event created a slide reaching 2,300 ft (700 m) below the crest of Grand Mesa. The depth of the debris was reported to be 123 ft (37 m). A series of smaller events occurred in the interval after the principal failure.

19.8 STREAM DIVERSION

This section describes a region that exhibits systematic "diversion" of stream channels over time. The formal meaning of "stream displacement" or "stream divergence" signifies intentional human rerouting of stream channels, often to open channels for navigation, permit passage of large vessels, or control flooding. Here, the discussion refers to natural processes that have diverted/displaced the courses of stream systems in rather distinctive patterns.

The northwestern region of Virginia is characterized by rough, uneven topography and narrow ridges. Here, streams follow valleys with floodplains, generally supporting pasture, cropland, and woodlands, bordered by steep, wooded topography (Figure 19.16). Valley floors are often positioned about 2200 ft above sea level, with neighbor-

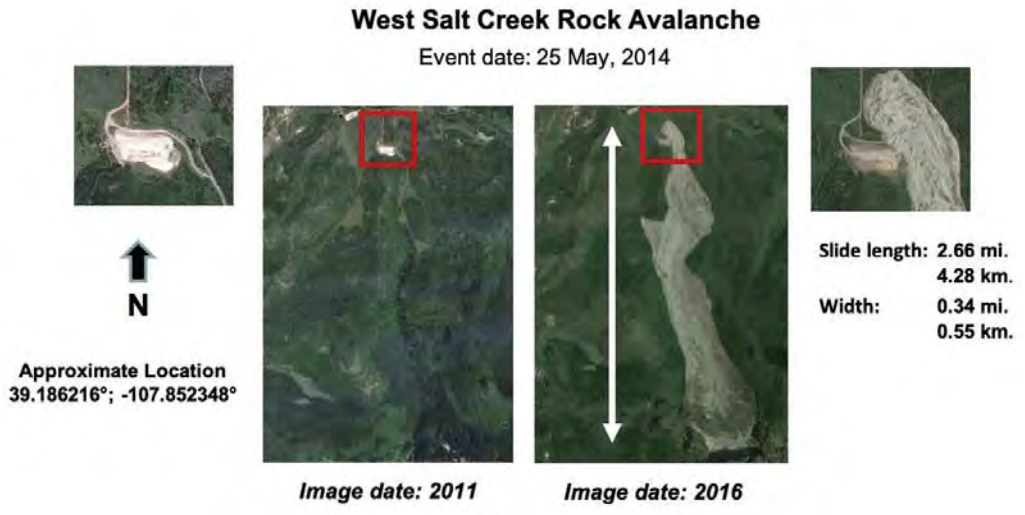


FIGURE 19.13 Before/after images of the West Salt Creek Rock Avalanche, May 25, 2014. A sudden release of saturated debris high at the Grand Mesa rim created this 2.6-mile flow. The left-hand Google Earth Pro image shows the area in August 2011 before the slide, with vegetation cover. The right-hand image is about 2 years after the 2014 slide. From Landsat/Copernicus.

FIGURE 19.14 Oblique, southward-facing aerial photograph of the crest of the Grand Mesa, which failed due to snowmelt and heavy rain (see the arrow). To appreciate the scale of the failure, compare the gray, cliff-like failure slope with the trees visible at the base of the cliff surface. The trees have been backrotated, tilted toward the cliff surface. The distance from the trees to the crest of the cliff is estimated to be about 100 m. From Getty Images, no. 493951413. Used by permission.

ing ridgelines about 3200 ft in elevation (creating an elevation difference of about 100 ft from valley floor to ridgelines). Ridges and steep terrain are generally characterized by complex geology, as well as a history of high-intensity cyclonic storms that often linger over tectonic events and human disturbance on mountain slopes and modifications of vegetation cover.

Many of the valleys in this region have distinctive features caused by the Appalachian Mountain chain. Valley edges are defined by clear-cut breaks in edges characterized by steep terrain that rises from valley borders to ridge crests. Within the valleys, streams switch sharply from one side of the valley to the other, as is evident in **Figure 19.17**.

Inspection of local terrain indicates something of the nature of topography and hydrologic processes that have been driving the overall topography of this area. **Figure 19.17** (A and B) shows the abrupt traverse of stream channel from one valley side to the opposite side. Inspection of the channel using the terrain tools in Google Earth Pro permits definition of the catchment area (C) and the deposition zone (D). Trial-and-error use of the Google Earth Pro terrain tools permits definition of sediment deposits at the western edge of the valley. **Figure 19.18** provides contours showing that there is about 100 ft of relief from the mouth of the catchment to the shallow end of the diversion of

the stream course at the eastern bend of the stream. Within these valleys, the distinctive patterns of natural stream diversions are characteristic of stream patterns throughout much of this region.

The long ridges characteristic of this region are supported by erosion-resistant sandstones, whereas valleys are characterized by more easily eroded rocks and sediments. Drainage is channeled along the narrow valleys parallel to the long ridges. Ridges are underlain by resistant rocks, while softer rocks of the valleys to the east and west are underlain by more easily weathered shales and limestone. The ridges are subject to mass wasting similar to the Grand Mesa events discussed above, but they are more the result of severe storms that displace large, fragmented strata of local ridges and, in effect, dam the narrow valleys (Godt and Coe, 2007, and Wooten et al., 2016). Wooten et al. (2016, pp. 203–204) outline the overall characteristics of mass wasting events in the Southern Appalachian Highlands (SAH):

> Debris flows, dominant among landslide processes in the SAH, are triggered when rainfall increases pore-water pressures in steep, soil-mantled slopes. Storms that trigger hundreds of debris flows occur about every 9 years and those that generate thousands occur about every 25 years. Rainfall from cyclonic storms triggered hundreds to thousands of

FIGURE 19.15 Oblique, northward-facing aerial photograph of the debris flow caused by the Grand Mesa failure on May 25, 2015. The cliff in this figure is off-camera at the lower right-hand corner of the image, with the debris flow having cascaded downslope from the lower right corner toward the distant center. The photograph depicts the terminus of the slide just as it vanishes into the distant topography at the top of the image. From Getty Images, no. 493951373. Used by permission.

FIGURE 19.16 Stream and topographic patterns, Spring Run, northwestern Virginia (38° 8.5938′ N, 79° 50.938′ W). This Google Earth Pro image illustrates the distinctive, irregular channel patterns within this narrow valley. From Landsat/Copernicus.

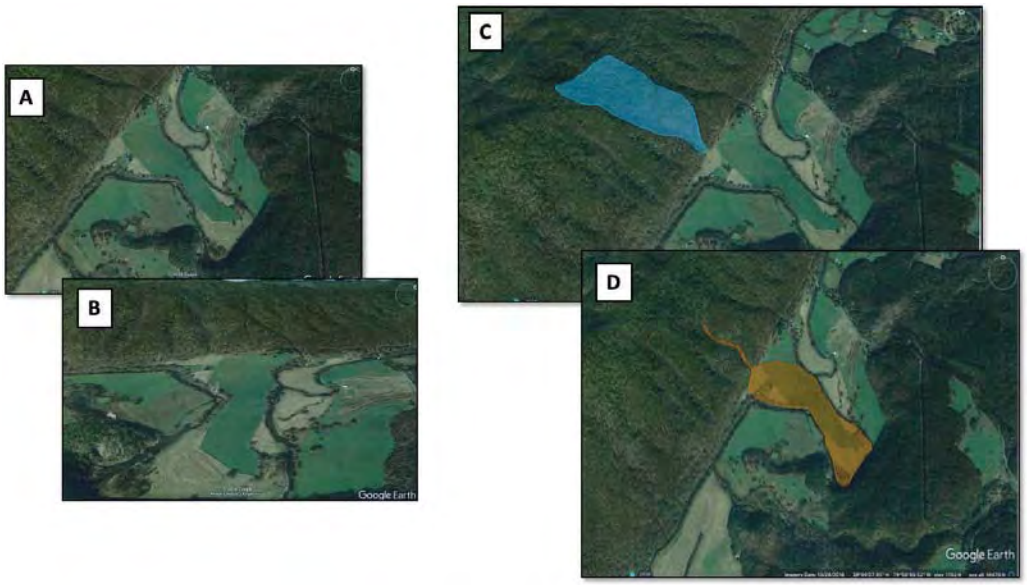


FIGURE 19.17 Historic debris flows in northwestern Virginia (38° 3.772′ N, 79° 50.143′ W). A and B show distinctive shapes of stream patterns after diversion by debris flows. C and D show estimates of the catchment area (C) and the deposition zone (D) after the debris flow. Graphics by Robert Severynse. Used by permission. Google Earth Pro image from Landsat/Copernicus.

FIGURE 19.18 Debris flow and stream diversion, northwestern Virginia, as shown previously in **Figure 19.17**. Here, contours map the topography to better illustrate elevations. The red arrows identify discharge sites from the catchment area, perhaps about 100 ft above the current stream elevation. This difference would appear sufficient for debris deposits to divert streamflow to form the distinctive stream channel visible today. Graphics by Robert Severynse. Used by permission. Topographic map from USGS. Google Earth Pro image from Landsat/Copernicus.

debris flows in 1916, 1940, 1969, 1977, 1985, and 2004. Debris flows have caused loss of life and property, and severely affected forest lands by altering forest structure and disrupting aquatic ecosystems.

Although there are fewer documented landslide events for the Blue Ridge Mountains, and Ridge and Valley of Virginia and West Virginia, rainfall events there have triggered the greatest numbers of documented landslides. The remnants of Hurricane Camille in 1969 generated a total of 5,377 documented landslides (mainly debris flows) in Virginia and West Virginia making it the largest magnitude, well-documented landslide event in the SAH.

19.9 GEOBOTANY

Geobotany is the study of relationships between plants and the underlying geologic substratum (Ustin et al., 1999). Such relationships include variations in the presence, abundance, vigor, and appearance of plants, indicator species, and distinctive phonological behavior. Often it can refer more specifically to examination of the relationships between plants and their response to unusual levels of nutrients or toxins released to the soil by weathering of geologic materials. *Regional geobotany* focuses on the study of vegetation patterns as indicators of lithologic variations. Specific plant species may have local significance as indicator species, for they tend to avoid or favor certain lithologic units. In other instances, variations in abundance or vigor may signal the occurrence of certain lithologies.

Geologic processes may concentrate trace elements in specific strata or regions. Geologic weathering may release trace elements in a form that can be absorbed by vegetation. Plants that absorb these trace elements at higher than normal levels may display abnormal spectral signatures, thereby signaling the existence and location of the anomalous concentration of elements.

Geobotany depends on knowledge of how geologic materials release elements into the nutrient pool, how these elements are absorbed by the soil and concentrated in plant tissues, and how they can alter the spectral signatures of plants. Because specific kinds of plants may thrive under certain unusual conditions of soil fertility, detailed knowledge of the spectral characteristics of such plants can be very valuable. Geobotanical studies are especially valuable in heavily vegetated regions where soil and rock are not exposed to the direct view of the sensor, but they may also be useful in sparsely vegetated regions where it may be possible to identify individual plants or where vegetation patterns may be especially sensitive to subtle environmental variations.

The practice of geobotanical reconnaissance is restricted by several factors. First, geobotanical studies may depend on observations of subtle distinctions in vegetation vigor and pattern, so successful application may require data of very fine spatial, spectral, and radiometric resolution—resolution much finer than that of Landsat or Sentinel data, for example. Thus, some of the concepts of geobotanical exploration cannot be routinely applied with the imagery and data most commonly available. Hyperspectral remote sensing (Chapter 14) provides capabilities for developing geobotanical knowledge. Second, geology is only one of the many factors that influence plant growth. Site, aspect, and disturbance (both by humankind and by nature) influence vegetation distributions in very complex patterns. The effects of these several causes cannot always be clearly separated.

The timing of image acquisition is critical. Some geobotanical influences may be detectable only at specific seasons or may be observable as, for example, advances in the timing of otherwise normal seasonal changes in vegetation coloring. The issue of timing may be especially important because the remote sensing analyst may not have control over the timing of image acquisition or may have imagery only for a single date, whereas observation at several dates may be necessary to observe the critical changes. Finally, many geobotanical anomalies may have a distinctively local character. Specific indicators may have meaning only for restricted regions, so intimate knowledge of regional vegetation and geology may be necessary to fully exploit knowledge of geobotany.

Geobotanical knowledge can be considered first in the context of the individual plant and its response to the geological substratum. The growth of most plants is sensitive not only to the availability of the *major nutrients* (phosphorus [P], potassium [K], and nitrogen [N]), but also to the *micronutrients,* those elements (including barium [Ba], magnesium [Mg], sulfur [S], and calcium [Ca]) that are required in very small amounts. It is well established that the growth and health of plants depend on the availability of elements present in very low concentrations and that sensitivity may be high in some plant species. In contrast, other elements are known to have toxic effects, even at very low concentrations, if they are present in the soil in soluble form. Heavy metals, including nickel (Ni), copper (Cu), chromium (Cr), and lead (Pb), may be present at sufficient levels to reveal their presence through stunted plant growth or through the localized absence of specific species that are especially sensitive to such elements. From such evidence, it may be possible to indirectly use the concentrations of these elements to reveal the locations of specific geologic formations or zones of mineralization worthy of further investigation.

Hydrocarbons may significantly influence plant growth if they are present in the root zone. Hydrocarbon gases may migrate from subsurface locations; at the surface, concentrations may be locally sufficient to influence plant growth. Hydrocarbons may be present in greater concentrations at the surface as petroleum seeps or coal seams. The presence of hydrocarbons in the soil may favor plant growth by increasing soil organic

matter but may also alter soil atmosphere and soil biology in a manner that is toxic to some plants.

Specific geobotanical indicators include the following, which of course may have varied geologic interpretation depending on the local setting. Variations in *biomass,* either significantly higher or lower than expected, may indicate geobotanical anomalies. Or *the coloring* of vegetation may be significant. The term *chlorosis* refers to a general yellow discoloration of leaves due to a deficiency of chlorophyll. Chlorosis can be caused by many of the geologic factors mentioned above, but it is not uniquely of geologic origin, so its interpretation requires knowledge of the local geologic and biologic setting.

Observation of such indicators may be facilitated by use of the vegetation indices and ratios mentioned in Chapter 16. Manual interpretation of color-infrared model imagery often provides especially good information concerning the vigor and distribution of vegetation. Of special significance in the context of geobotany is the *blue shift* in the chlorophyll absorption spectra. The characteristic spectral response for living vegetation was discussed in Chapter 16. In the visible, peak reflectance is in the green region; the absorption spectra of chlorophyll decreases reflectance in both the blue and the red regions. However, in the near infrared, reflectance increases markedly (Chapter 16).

Collins, Chang, and Kuo (1981) and others have observed that geochemical stress is most easily observed in the spectral region from about 0.55 to 0.75 μm (which includes portions of the green, red, and infrared regions). Most notably, the position and slope of the line that portrays the increase in reflectance at the edge of the visible region, although constant for healthy green plants, are especially sensitive to geochemical influences. As geochemical stress occurs, the position of this line shifts toward shorter wavelengths (i.e., toward the blue end of the spectrum; hence the term *blue shift*), and its slope becomes steeper (Figure 19.19).

It must be emphasized that, by everyday standards, the change is very subtle (0.007–0.010 μm) and has been observed only in data recorded by high-resolution sensors processed to filter out background reflectance. Such data are not acquired by remote sensing instruments available for routine use, although increased spectral, spatial, and radiometric resolution provided by hyperspectral remote sensing (Chapter 14) opens up opportunities to exploit this kind of knowledge. Nonetheless, Collins et al. (1983) reported that the blue shift had been observed in all plant species studied thus far; they also reported

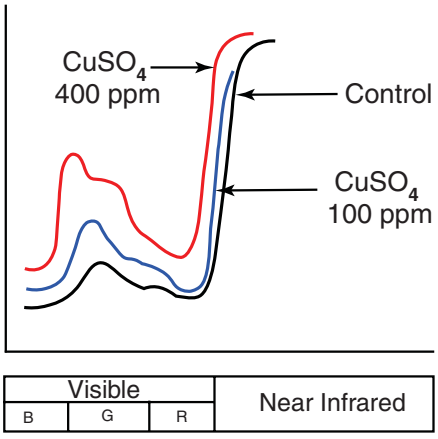


FIGURE 19.19 Blue shift at the edge of the chlorophyll absorption band. As the heavy mineral concentrations in plant tissue increase, the edge of the absorption band shifts toward shorter wavelengths (i.e., toward the blue end of the spectrum). The solid black line shows the spectrum for a control plant. Dotted and dashed lines show how magnitudes of the shift increase as concentration of the heavy minerals increases. Adapted from Chang and Collins (1983, p. 727). Image by Susmita Sen. Used by permission of the Society of Economic Geologists.

that they believed this effect to be "universal for most green plants" (p. 739), so it is not confined to a few species.

The cause of the blue shift is not clearly understood at present. Unusual concentrations of heavy metals in the soil are absorbed by plants and apparently tend to be transported toward the actively growing portions of the plant, including the leaves. Concentrations of such metals can cause chlorosis, which is often observed by more conventional methods. The relationship between heavy metals and the blue shift is clear, but work by Chang and Collins (1983) indicates that the presence of heavy metals in the plant tissues does not alter the chlorophyll itself: the metallic ions apparently do not enter the structure of the chlorophyll. The heavy metals appear to retard the development of chlorophyll, thereby influencing the abundance, but not the quality, of the chlorophyll in the plant tissue.

Chang and Collins (1983) conducted laboratory and greenhouse experiments to confirm the results of spectral measurements in the field and to study the effects of different metals and their concentrations. Their experiments used oxides, sulfates, sulfides, carbonates, and chlorides of Cu, zinc (Zn), iron (Fe), Ni, manganese (Mn), molybdenum (Mo), and vanadium (V). Selenium (Se), Ni, Cu, and Zn produced stress at concentrations as low as 100 ppm. Fe and Pb appeared to have beneficial effects on plant growth; Mo and V produced little effect at the concentrations studied. In combination, some elements counteracted each other, and some varied in effect as concentrations changed.

19.10 DIRECT MULTISPECTRAL OBSERVATION OF ROCKS AND MINERALS

A second broad strategy for remote sensing of lithology depends on accurate observation of spectra from areas of soil and rock exposed to observation. Color and, by extension, the spectral characteristics of rocks and minerals can be distinctive identifying features in the direct examination of geologic samples. Geologists use color for identification of samples in the field, and in the laboratory sensitive instruments are used to measure spectral properties across a range of wavelengths in the ultraviolet, visible, and infrared regions. Spectral emittance and reflectance of rocks and minerals are important properties that are often closely related to their physical and chemical properties. In the laboratory, spectral characteristics can be observed in such sufficient detail that they can sometimes form diagnostic tests for the presence of specific minerals. In addition, small radiometers can be taken to the field to permit *in situ* observation of rocks and minerals in nonvisible portions of the spectrum. As a result, geologists have considerable experience in observing the spectral properties of geologic materials, and they have developed an extensive body of knowledge concerning the spectral properties of rocks and minerals.

The application of this knowledge in the context of remote sensing can be very difficult. In the laboratory, spectra can be observed without the contributions of atmospheric attenuation, vegetation, or shadowing to the observed spectra. In the laboratory setting, scientists can hold secondary properties (such as moisture content) constant from one sample to the next or can compensate for their effects. In remotely sensed data, the effects of such variables often cannot be assessed. The usual remote sensing instruments do not have the fine spectral and radiometric resolutions required to make many of the subtle distinctions this strategy requires. Furthermore, the relatively coarse spatial resolution of many remote sensing systems means that the analyst must consider composite signatures

formed by the interaction of many landscape variables rather than the pure signatures that can be observed in the laboratory. Whereas photogeology distinguishes between units without attempting to make fine lithologic distinctions, much of the research in multispectral remote sensing in geology is devoted to more precise identification of specific minerals thought to be both spectrally distinctive and especially valuable in mineral exploration.

Locations of zones of hydrothermal alteration may be revealed by the presence of limonite at the surface. *Limonite* refers broadly to minerals bearing oxides and oxyhydrites of ferric iron, including goethite and hematite. Such minerals tend to exhibit broad absorption bands in the near-infrared, visible, and ultraviolet regions. Typically, their spectra decline below 0.5 μm, producing a decrease in the ultraviolet that is not normally observed in other minerals (**Figure 19.20**). A broad, shallow absorption region from 0.85 μm to 0.95 μm is observed in the near infrared. The presence of limonite, either as a primary mineral or secondarily as the product of geologic weathering, may identify the location of a zone of hydrothermal alteration, thereby suggesting the possibility of mineralized zones. However, the identification of limonite is not definitive evidence of hydrothermal activity, as limonite can be present without hydrothermal alteration and hydrothermal activity may occur without the presence of limonite. Thus, the observation of these spectra, or any spectra, is not uniquely specific, and the analyst must always consider such evidence in the context of other information and his or her knowledge of the local geologic setting.

Clay minerals often decrease in reflectance beyond 1.6 μm; this behavior is common to enough minerals and occurs over a spectral region that is broad enough that it has been used to locate zones of hydrothermal alteration by means of the surface materials rich in clay minerals. At finer spectral resolution, these clay minerals and carbonates character-

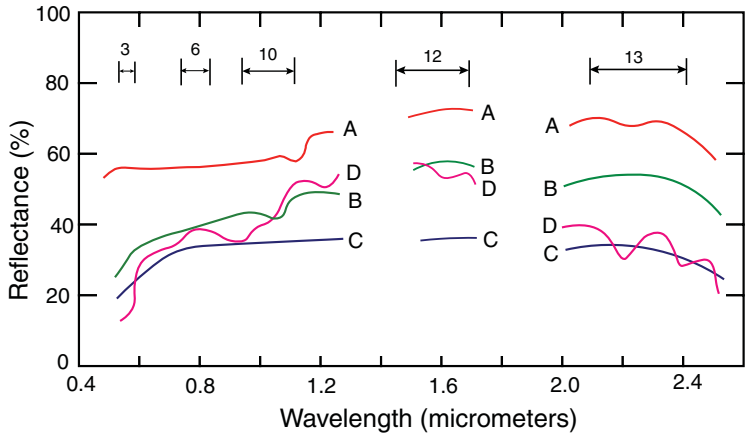


FIGURE 19.20 Spectra of some natural geologic surfaces: (A) White Joe Lott Tuff member of Mt. Belknap volcanics; (B) limonitic Joe Lott Tuff member of Mt. Belknap volcanics; (C) tan soil on rhyolite; (D) orange-layered laterite. Data were collected in the field near the surface using portable spectroradiometers. Limonitic minerals (B, C, D) show low reflectances at wavelengths below 0.5 m, relative to nonlimonitic materials (A). Based on Goetz et al. (1975); adapted from Williams (1983, p. 1748). Image by Susmita Sen. Used by permission.

ized by aluminum hydroxide and magnesium hydroxide structures display spectra with a narrow but distinctive absorption band from 2.1 to 2.4 μm (Figure 19.20).

Emittance in the spectral region from 8–12 μm (the mid-infrared region) permits the identification of some silicate minerals and the separation of silicate minerals from nonsilicate minerals. Silicate minerals typically exhibit emittance minima in this region, whereas nonsilicate minerals have minima at longer wavelengths.

19.11 PHOTOCLINOMETRY

Photoclinometry, loosely defined, is the understanding of the relationship between image brightness and the orientation of the surface that generated the brightness. If an irregular surface of uniform reflectance is illuminated at an angle, variations in image brightness carry information concerning the slopes of individual facets on the ground. Therefore, the full image, composed of many such pixels, depicts the shape of the terrain (Eliason, Soderblom, and Chavez, 1981). The Earth's surface does not reflect uniformly, so the brightness caused by irregular terrain is mixed with the brightnesses caused by different surface materials. Remote sensing is usually concerned with the spectral differences that we observe at different places on the Earth's surfaces as clues to the local abundance of minerals and other resources. Therefore, the brightnesses caused by surface orientation are often regarded as complicating factors in our effort to extract and interpret spectral information.

The field of photoclinometry contributes to remote sensing by providing tools that permit analysis to separate the spectral information from the brightness that arises from the orientation of the surface. A diffuse reflector (Chapter 2) will have a brightness that is predictable from the angle of illumination. It is intuitively obvious that as the angle of illumination changes, so also will the brightness of the surface. In the context of remote sensing, the surface and the source of illumination have a fixed geometric relationship, but the scene itself is composed of many individual facets, each with a specific orientation with respect to the solar beam. That is, the solar beam illuminates the Earth at a fixed angle at any given instant, but topographic irregularities cause the image to be formed of many varied brightnesses. Thus, for a uniform surface of irregular topography, variations in image gray tone portray variations in surface slope and orientation.

Although few surfaces on the Earth meet the assumption of uniform reflectance, some extraterrestrial surfaces display only small variations in reflectivity, primarily because of the absence of vegetative cover and differences in color caused by the weathering of geologic materials. Photogeologists who have studied the Moon's surface and those of the planets have formalized the methods necessary to derive topographic information from images (Wildey and Pohn, 1964). Photoclinometry encompasses elements of remote sensing, photogrammetry, photometry, and radiometry. Although both the theory and practice of photoclinometry must still deal with many unsolved issues, it has important applications in remote sensing.

For the present discussion, let us note simply that it is possible to approximate the brightnesses of the terrain, measured in a specific pair of bands, by a ratio between them:

$$\frac{G_{\text{band 1}}}{G_{\text{band 2}}} = C\frac{R_1}{R_2} \qquad \text{(EQ. 19.1)}$$

where G represents the brightness of the gray tone on an image of spectral band selected from a multiband image, and C represents an arbitrary constant. This ratio image (R_1/R_2) has several useful properties.

First, the relationship holds both for shadowed pixels and for directly illuminated pixels. Therefore, a ratio image shows pure reflectance information without the effects of topography. This result permits geologists to examine the spectral properties of surfaces without the confusing effects of mixed brightnesses of topography and material reflectances. In the raw image, a difference in pixel brightness can be caused by a difference in slope, by shadowing, or by differences in the color of the surface material. In a ratio image, the geologist knows that differences in brightness are caused by differences in reflectance only.

In the context of photoclinometry, ratio images have a special application. An unsupervised classification of the several ratio images derived from a multispectral image permits identification of regions on the Earth's surface that have uniform spectral behavior. Remember that the ratio images convey only spectral, not topographic, information, so a classification performed on ratio images, not on the original digital values, will produce classes based on uniform spectral properties regardless of slope.

19.12 BAND RATIOS

Band ratios have further significance in geologic remote sensing. By removing the effects of shadowing, which otherwise are mixed with the spectral information necessary to make lithologic discriminations, they can convey purely spectral information. Furthermore, ratios may minimize differences in brightness between lithologic units (i.e., ratios tend to emphasize color information and to deemphasize absolute brightness) and may facilitate comparisons of data collected on different dates, which will differ in solar angle.

The use of ratio images carries certain risks, as mentioned in Chapter 16. In a geologic context, it is important that data be free of atmospheric effects or that such effects be removed by preprocessing. The significance of the atmosphere can be appreciated by reexamining the earlier discussion, in which the diffuse light values in the two band values were estimated to have predictable values that would cancel each other out. If severe atmospheric effects are present, they will differ from one band to the other. The value of the ratio will no longer portray only spectral properties of the ground surface but instead will have values greatly altered by the varied atmospheric contributions to the separate bands.

19.13 SOIL AND LANDSCAPE MAPPING

Soil scientists study the mantle of soil at the Earth's surface in an effort to understand soil formation and to map patterns of soil variation. *Soil* is a complex mixture of inorganic material weathered from the geologic substratum and mixed with decayed organic matter from tissues of plants and animals. Typically, a "soil" consists of three layers (known as *soil horizons*) of varying thickness and composition (Figure 19.21). Nearest the surface is the A horizon, which is usually dark in color and rich in decayed organic matter. The A horizon is sometimes known popularly as *topsoil*—the kind of soil one would like to have for a lawn or garden. In this layer, plant roots are abundant, as are microorgan-

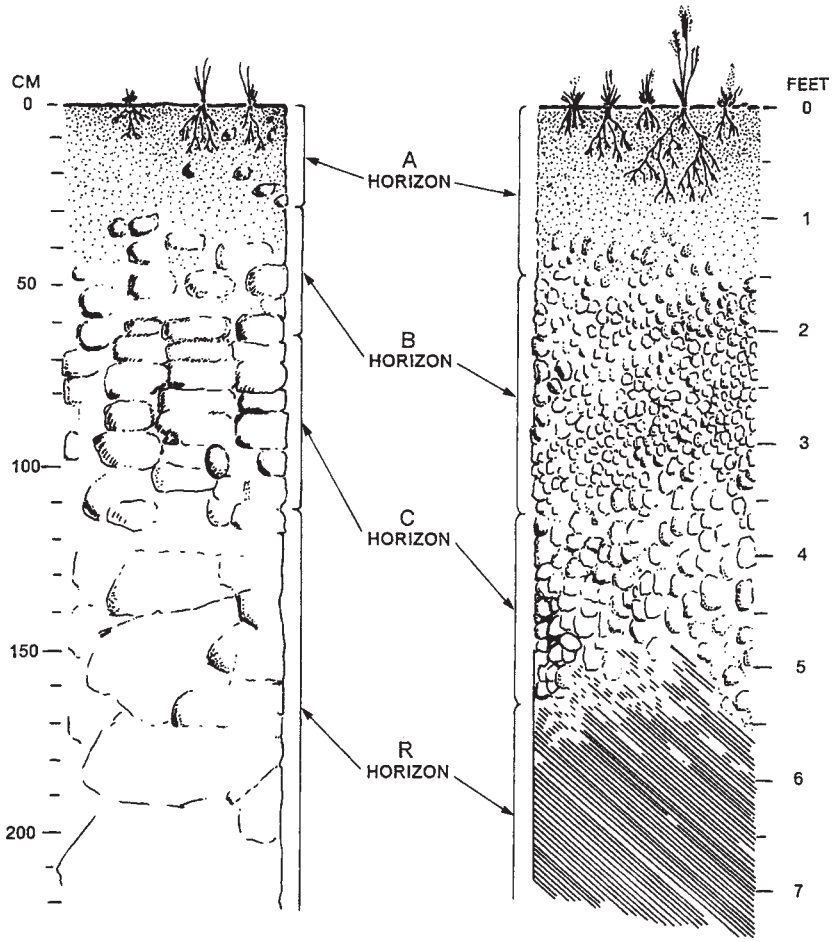


FIGURE 19.21 Soil profiles, presented here as cartoon-like sketches to illustrate varying structures and forms of contrasting soils. Detailed assessment of soil profiles requires firsthand observation (such as **Figure 19.23**).

isms, insects, and other animals. Below the A horizon is the second layer, the B horizon, which is usually lighter in color and more compact and where plant roots and biologic activity are less abundant. Sometimes the B horizon is known as the *subsoil*—the kind of hard, infertile soil one would prefer *not* to have at the surface of a lawn or garden. At even greater depth is the C horizon, the deepest pedologic horizon, which consists of weathered geologic material, decayed or fractured into material that is softer and looser than the unweathered geologic strata below. This is the material that usually forms the raw material for the A and B horizons. Finally, below the C horizon is the R horizon, consisting of unaltered bedrock.

The exact nature of the horizons at a given place is determined by the interaction between local climatic, topographic, geologic, and biologic elements as they act over time. Because varied combinations of climate, vegetation, and topography produce dif-

ferent soils from similar geologic materials, soil science is distinct from, though closely related to, geology and geomorphology.

The landscape is covered by a mosaic of patches of different kinds of soil, each distinct from its neighbors with respect to character and thickness (Figure 19.22). Soil surveyors outline on maps those areas covered by specific kinds of soil as a means of showing the variation of soils on the landscape. Each symbol represents a specific kind of soil, or, when the pattern is very complex, sometimes two or three kinds of soil that occur in an intimate pattern (Figure 19.23). For example, in Figure 19.23, near the center of the image is a large patch of a relatively flat, non-treed land with type 16B soil. This is defined as Groseclose and Poplimento soil, with 2–7% slopes (Table 19.1). Soil maps portray distributions of pedologic units, which, together with other maps and data, convey valuable information concerning topography, geology, geomorphology, hydrology, and other landscape elements. In the hands of a knowledgeable reader, they convey a comprehensive picture of the physical landscape. As a result, they can be considered to be among the most practical of all forms of landscape maps and are used by farmers, planners, and others who must judge the best locations for specific agricultural activities, for community facilities, or for construction of buildings and highways.

Soil mapping is conducted routinely by national soil surveys (or their equivalents) in most countries and by their regional counterparts at lower administrative levels. Details of mapping technique vary from one organization to another, but the main outlines of the procedure are common to most soil survey organizations. A soil map is formed by subdividing the landscape into a mosaic of discrete parcels. Each parcel is assigned to a class of soil—a *mapping unit*—that is characterized by specific kinds of soil horizons. Each mapping unit is represented on the map by a specific symbol—usually a color or

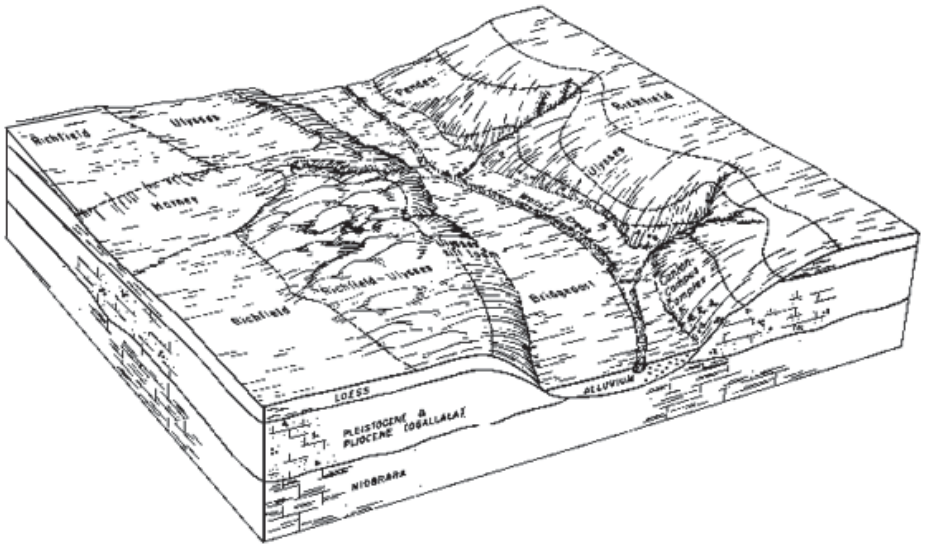


FIGURE 19.22 The soil landscape, as represented by a schematic sketch depicting principal landscape features and dominant soil units, northeastern Kansas. Such diagrams present a broad perspective of the overall pattern of principal soils of the region. From U.S. Department of Agriculture, Natural Resources Conservation Service.

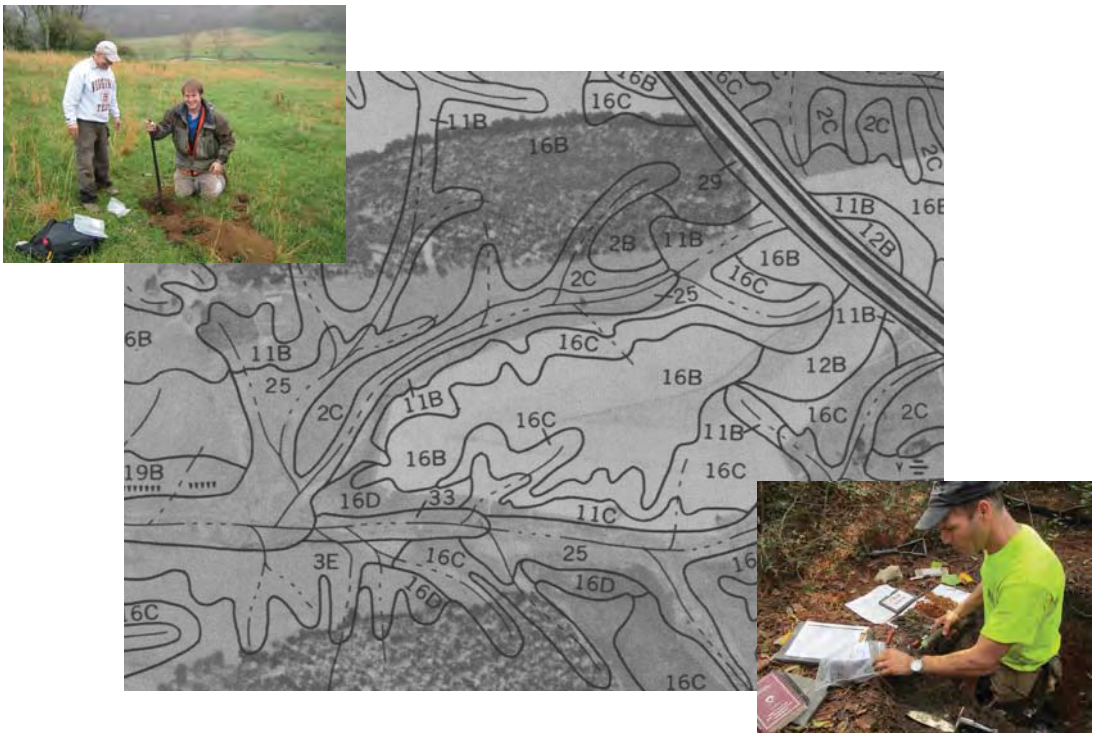


FIGURE 19.23 Field collection of soil data as a component of soil mapping. The map overlay (middle) shows soil surveyors' delineations of distinctive soil units in Montgomery County, Virginia (subset from Sheet Number 19, in Porter et al., 1985, pp. 192). The upper left and lower right images show USGS scientists collecting soil samples in the field. The soil codes shown in this image are explained in **Table 19.1**. From USGS. Upper left photograph by Nora Foley.

an alphanumeric designation. In theory, mapping units are homogeneous with respect to pedologic properties, although experienced users of soil maps acknowledge the existence of internal variation, as well as the presence of foreign inclusions. Mapping units are usually defined with links to a broader system of soil classification defined by the soil survey organization, so that mapping units are consistently defined across, for example, county and state borders. For each region, a mapping unit can be evaluated with respect to the kind of agriculture that is important locally, so the map can serve as a guide to select the best uses for each soil. The following paragraphs outline an example of the soil survey process, although details may vary from one organization to the next.

The first step in the actual preparation of the draft map is the *mapping* process (Figure 19.23). The term *mapping* in this context means specifically to designate the delineation of landscape units on an aerial photograph or map base, rather than in the broader sense that refers to the entire set of activities necessary to generate the map. In the mapping step, the soil scientist examines aerial photographs, often using a stereoscope, to define the major boundaries between classes of soil. The soil scientist has already learned much about the region's soils through field observations, so he or she has knowledge of the kinds of soils present and their approximate locations. The soil scientist can use this

field knowledge to interpret the photograph, using breaks in slope, boundaries between vegetation classes, and drainage patterns to define boundaries between soil mapping units. Seldom if ever can the soil classes actually be identified from the photograph; identification must be based on field observations. But in the mapping step, the scientist can subdivide the landscape into a mosaic of soil parcels, each of which is then treated as an independent unit in the later steps.

Completion of the mapping step requires considerable field experience within the region to be mapped, as the surveyor must acquire knowledge of the numbers and kinds of soils present within a region, their properties and occurrence, and their uses and limitations. As a result, completion of this step requires a much longer time than that actually devoted to marking the aerial photographs if we include the time required to learn about the local landscape. Morphological descriptions of each mapping unit are made in the field, and mapping units are sampled for later analysis.

The second step is *characterization*. Samples are collected from each prospective mapping unit and then subjected to laboratory analysis for physical, chemical, and mineralogical properties.

These measurements form the basis for *classification,* the third step. In the United States, classification is the implementation of the classification criteria specified by Soil Taxonomy, the official classification system of the Soil Conservation Service, U.S. Department of Agriculture. In other countries, classification is conducted by applying the classification criteria established by each national soil survey organization.

Correlation, the fourth and possibly the most difficult step, matches mapping units within the mapped region to those in adjacent regions and to those in ecologically analogous areas. Whereas the other steps in a soil survey may be conducted largely on a local basis, correlation requires the participation of experienced scientists from adjacent regions, and even those from national or international levels in the organization, to provide the broader experience and perspective often required for successful correlation.

TABLE 19.1 **Soil Types Present in the Map from Montgomery County, Virginia, Depicted in Figure 19.23**

Symbol	Type
2B	Berks–Groseclose Clymer complex, 2–7% slopes
2C	Berks–Groseclose Clymer complex, 7–15% slopes
3E	Berks–Lowell–Rayne complex, 25–65% slopes
11B	Duffield–Ernest complex, 2–7% slopes
12B	Frederick and Vertrees silt loams, 2–7% slopes
16B	Groseclose and Poplimento soils, 2–7% slopes
16C	Groseclose and Poplimento soils, 7–15% slopes
16D	Groseclose and Poplimento soils, 15–25% slopes
25	McGary and Purdy soils
29	Udorthents and urban land
33	Weaver soils

The fifth and final step is *interpretation,* in which each mapping unit is evaluated with respect to prospective agricultural and engineering uses. The interpretation step provides the user of the map with information concerning the likely suitability of each mapping unit to the land uses most common in the region.

Within this rather broad framework there are, of course, many variations in details of technique and in overall philosophy and strategy. The result is a map that shows the pattern of soils in a region and a report that describes the kind of soil that is encountered in each mapping unit. Each mapping unit is evaluated with regard to the kinds of uses that might be possible for the region, so that the map serves as a guide to wise use of soil resources.

19.14 INTEGRATED TERRAIN UNITS

Definition of *integrated terrain units* is based on the concept that the varied and complex assemblages of soil, terrain, vegetation, and so on form distinctive spectral responses that can be recognized and mapped. This general approach has long been applied, under many different names, to the interpretation of resource information from aerial photographs and other forms of imagery. The best known, and oldest, of these methods is known as *land system mapping.*

Land system mapping subdivides a region into sets of recurring landscape elements based on comprehensive examination of distributions of soils, vegetation, hydrology, and physiography. The formalization of land system mapping procedures dates from the period just after World War II, although some of the fundamental concepts had been defined earlier. Today a family of similar systems is in use, all of which are based on similar principles and methods, although the terminology and details vary considerably. One of the oldest, best known, and most formally defined systems is that developed by the Australian Commonwealth Scientific and Industrial Research Organisation, Division of Land Research and Regional Survey (Christian, 1959); variations have been developed and applied throughout the world.

The method is based on a hierarchical subdivision of landscapes. *Land systems* are recurring, contiguous associations of landforms, soils, and vegetation, composed of component *land units*. The basic units of the system (which are assigned varied definitions and names in alternative versions of this basic strategy) are areas of uniform lithology with relatively uniform soil and drainage. A characteristic feature of all versions of the land system method is a hierarchical spatial organization, so that subcategories are nested within the broader categories defined at higher levels. Designation of separate levels within the hierarchy differs among alternative versions. For example, Thomas (1969) defines the following sequence: *site* (the smallest units), *facet, unit landform, landform complex, landform system,* and *landform region*. Wendt, Thompson, and Larson (1975) define a system using *landtypes* (the smallest regions), *landtype associations, landtype phases, subsections, sections,* and *provinces*. Although many versions imply that both land systems and their component elements recur in widely separated geographic regions (though under analogous ecological circumstances), closer examination of the method (Perrin and Mitchell, 1969) reveals that land systems are most effective when defined to be essentially local units.

Implicit, if not explicit, in most applications is the use of aerial photography as a primary tool for land system mapping, together with direct observation in the field. Aerial

photography provides the broad overview and the map-like perspective favoring convenient definition and delineation of land systems. It portrays the complex spatial patterns of topography, vegetation, and drainage in an integrated form that is compatible with the assumptions, methods, and objectives of this approach to terrain analysis. "There would be no point in defining any terrain class if its chances of being recognized from the air photographs and background information were small" (Webster and Beckett, 1970, p. 54).

An assumption of the procedure is that easily identified landscape features (e.g., vegetation and physiography) form surrogates for more subtle soil features not as easily defined from analysis of aerial photographs. Although the procedure has been applied at a range of scales from 1:1,000,000 to 1:25,000, most applications are probably at fairly broad scales, for reconnaissance surveys, or mapping of rather simple landscape divisions under circumstances in which the ultimate use will be rather extensive.

The method and its many variations have been criticized for their subjectivity and the variability of the results obtained by different analysts. Compared with most large-scale soil surveys, land system mapping presents a rather rough subdivision of the landscape, as the mapping units display much greater internal variability and are not as carefully defined and correlated as one would expect in more intensive surveys. As a result, the procedure may be best suited for application in reconnaissance mapping, for which broad-scale, low-resolution mapping is required as the basis for planning more detailed surveys.

The method of integrated terrain units has also been applied to analysis of digital data. At the coarse resolution of satellite sensors, individual spectral responses from many landscape components are combined to form composite responses. In this context, the usual methods of image classification for mapping individual classes of soil, vegetation, or geology may be extraordinarily difficult. Therefore, many analysts have applied the integrated terrain unit strategy to digital classification in an effort to define image classes in a more realistic manner. In the ideal, we might prefer to define pure categories that each show only a single thematic class. But given the complexity of the natural landscape, composite categories are often well suited to representation of the gradations and mixtures that characterize many environments (Green, 1986; Robinove, 1981).

19.15 WETLANDS INVENTORY

Wetlands are areas of land characterized by saturated or inundated soil. In the United States, wetlands are legally defined by the presence of (1) hydrophytic vegetation (plants that occur predominantly in wetlands environments); (2) a wetlands hydrologic regimen, including evidence of inundation or saturation; and (3) diagnostic pedologic features caused by low oxygen levels, most notably gray or black colors, often with distinctive mottling indicative of prolonged saturation. Wetlands are not defined merely by proximity to bodies of open water but are encountered in a wide variety of ecological and physiographic settings, including uplands.

Understanding the occurrence of wetlands is significant for a broad spectrum of policy issues, including land-use regulation, water quality, flood abatement, agriculture, and carbon sequestration. As a result, mapping their occurrence and extent is an important responsibility of local, state, and federal governments. For some purposes, field studies are necessary to identify and delineate the extent of wetlands.

For other purposes, however, remotely sensed imagery forms an important tool for inventory and monitoring of wetlands. At a national level in the United States, the National Wetlands Inventory conducted by the U.S. Fish and Wildlife Service prepares a national inventory based on photointerpretation of aerial imagery such as that from the National High-Altitude Aerial Photography Program (NHAP) and National Agricultural Imagery Program (NAIP). Although wetlands vary greatly in their occurrence and distinctive properties (Lyon, 2001), optical, thermal, and SAR data all have distinctive capabilities that can detect some of the distinctive characteristics of wetlands. Because wetlands vary so greatly with respect to their distinct properties and vary seasonally, it may not be feasible to propose a single uniformly applicable approach to remote sensing of wetlands. It seems more likely that remotely sensed imagery may form the central resource for a suite of data that might be useful in delineating and monitoring wetlands.

19.16 RADAR IMAGERY FOR EXPLORATION

As a conclusion to this chapter, we note the application of SAR data to offshore petroleum exploration. Natural seepage of petroleum from the ocean floor causes a thin film of petroleum floating on the ocean surface (a "slick") that creates a locally smooth water surface, which in turn induces conditions that promote specular reflection over that region of the ocean. Oil slicks occur when molecules of oil reach the sea surface to form a thin layer of petroleum that dampens the ocean-surface capillary waves. SAR technology is sensitive to differences in surface roughness, so it can easily discriminate between the smoother oil slick and the surrounding rough water. Thus, oil slicks are regions of little to no backscatter, characterized by distinct areas of darkness on the radar image. Therefore, the dark returns assist in identifying regions where such slicks might originate. Because SAR has the ability to observe such regions day and night, and over a wide range of meteorological conditions, SAR data assist in identifying regions where such slicks form on a regular basis. Critical exploration maps can be produced by using a geographic information system to overlay other information such as coastline, bathymetry, gravimetric data, shipping lanes, and existing oil rigs on the SAR image. These maps allow decision makers to reduce magnetic field exploration risk and increase the cost-effectiveness of offshore drilling efforts.

19.17 SUMMARY

Aerial photography and remote sensing have long been applied to problems in the earth sciences. Geology, topographic mapping, and related topics formed one of the earliest routine applications of aerial photography, and today remote sensing, used with other techniques, continues to form one of the most important tools for geologic mapping, exploration, and research.

Research for geologic remote sensing spans the full range of subjects within the field of remote sensing, including use of additional data in the thermal, near-infrared, visible, and microwave regions of the spectrum. More than most other subjects, efforts such as geobotanical research are especially interesting because they bring knowledge of so many different disciplines together to bear on a single problem. In addition, geological

investigations have formed important facets of research and applications programs for optical (i.e., Landsat, SPOT, ERS-2, and others) and for SAR (i.e., RADARSAT, ERS-1, and others) systems. Geological studies are said to form a major economic component of the practice of remote sensing. Certainly, they form one of the most important elements in both the theoretical and practical advances in remote sensing.

SOME TEACHING AND LEARNING RESOURCES

- Geologic Fieldwork with ASD: part 1 of 3
  www.youtube.com/watch?v=cjuitl_-_LU
- Geologic Fieldwork with ASD: part 2 of 3
  www.youtube.com/watch?v=xBknHsZdG2E&feature=related
- Geologic Fieldwork with ASD: part 3 of 3
  www.youtube.com/watch?v=2-AkC8jcL08&feature=related

REVIEW QUESTIONS

1. Why is the timing of overpasses of Earth observation satellites such as Landsat or SPOT likely to be of special significance for geologic remote sensing?
2. Explain why the ability to monitor the presence of moisture (both open water bodies and soil moisture at the ground surface) might be of special significance in geologic, geomorphic, and pedologic studies.
3. In the past, there has been discussion of conflicts between earth scientists and plant scientists concerning the design of spectral sensitivity of sensors for the Multispectral Scanner System (MSS) and Thematic Mapper (TM), based on the notion that the two fields have quite different requirements for spectral information. Based on information in this chapter, explain why such a distinct separation of information requirements may not be sensible.
4. Why should Landsat imagery have been such an important innovation in studies of geologic structure?
5. Summarize the significance of lineaments in geologic studies and the significance of remote sensing to the study of lineaments.
6. Compare the relative advantages and disadvantages of photogeology and geologic remote sensing, as outlined in this chapter.
7. Some might consider lineament analysis, application of geobotany, and other elements of geologic remote sensing to be modern, state-of-the-art techniques. If so, why should photogeology, which is certainly not as technologically sophisticated, be so widely practiced today?
8. Full application of the principles and techniques of geologic remote sensing requires much finer spectral, radiometric, and spatial resolution than most operational sensors have at present. What difficulties can you envision in attempts to use such sensors for routine geologic studies?

9. Refer back to Chapter 12; compare and contrast the concept of image classification with that of integrated terrain units (Section 19.14). Write two or three paragraphs that summarize the major differences between the two strategies.
10. In what ways might radar imagery (Chapter 8) be especially useful for studies of geology and other earth sciences?
11. Write a short description of a design for a multispectral remote sensing system tailored specifically for recording geologic information. Disregard all other applications. Suggest how you would select the optimum spatial and radiometric resolution and the most useful spectral regions.

REFERENCES

Abrams, M. J., D. Brown, L. Lepley, and R. Sadowski. 1983. Remote Sensing for Porphyry Copper Deposits in Southern Arizona. *Economic Geology,* Vol. 78, pp. 591–604.

Agbu, P. A., and E. Nizeyimana. 1991. Comparisons between Spectral Mapping Units Derived from SPOT Image Texture and Field Mapping Units. *Photogrammetric Engineering and Remote Sensing,* Vol. 57, pp. 397–405.

Bauer, B. O. 2004. Geomorphology. In *Encyclopedia of Geomorphology, Vol. 1* (A. S. Goudie, ed.). London: Routledge, pp. 428–435.

Berlin, G. L., P. S. Chavez, T. E. Grow, and L. A. Soderblom. 1976. Preliminary Geologic Analysis of Southwest Jordan from Computer Enhanced Landsat I Image Data. In *Proceedings, American Society of Photogrammetry.* Bethesda, MD: American Society of Photogrammetry, pp. 543–563.

Bishop, J. L., and E. Murad. 2005. The Visible and Infrared Spectral Properties of Jarosite and Alunite. *American Mineralogist,* Vol. 90, pp. 1100–1107.

Casas, A. M., A. L. Cortés, A. Maestro, M. Asunción Soriano, A. Riaguas, and J. Bernal. 2000. LINDENS: A Program for Lineament Length and Density Analysis. *Computers and Geosciences,* Vol. 26, pp. 1011–1022.

Chang, S., and W. Collins. 1983. Confirmation of the Airborne Biogeophysical Mineral Exploration Technique Using Laboratory Methods. *Economic Geology,* Vol. 78, pp. 723–726.

Chavez, P. S., G. L., Berlin, and A. V. Costa. 1977. Computer Processing of Landsat MSS Digital Data for Linear Enhancements. In *Proceedings of the Second Annual William T. Pecora Memorial Symposium.* Falls Church, VA: American Society for Photogrammetry, pp. 235–251.

Christian, C. S. 1959. The Eco-Complex and Its Importance for Agricultural Assessment. Chapter 36 in *Biogeography and Ecology in Australia. Monographiae Biologicae,* Vol. 8, pp. 587–605.

Clark, C. D., and C. Watson. 1994. Spatial Analysis of Lineaments. *Computers and Geosciences,* Vol. 20, pp. 1237–1258.

Coe, J. A., R. L. Baum, K. E. Allstadt, B. F. Kochevar, Jr., R. G. Schmitt, M. L. Morgan, . . . and J. W. Kean. 2016, April. Rock-Avalanche Dynamics Revealed by Large-Scale Field Mapping and Seismic Signals at a Highly Mobile Avalanche in the West Salt Creek Valley, Western Colorado. Research Article. *Geosphere,* Vol. 12, No. 2, pp. 607–631.

Cole, M. M. 1982. Integrated Use of Remote Sensing Imagery in Mineral Exploration. *Geological Society of America Transactions,* Vol. 85, Part 1, pp. 13–28.

Collins, W., S. H. Chang, and J. T. Kuo. 1981. *Detection of Hidden Mineral Deposits by Airborne Spectral Analysis of Forest Canopies.* NASA Contract NSG-5222. Final Report. 61 pp.

Collins, W., S. Chang, G. Raines, F. Channey, and R. Ashley. 1983. Airborne Biogeochemical Mapping of Hidden Mineral Deposits. *Economic Geology,* Vol. 78, pp. 737–749.

Crippen, R. E., and R. G. Blom. 2001. Unveiling the Lithology of Vegetated Terrains in Remotely Sensed Imagery. *Photogrammetric Engineering and Remote Sensing,* Vol. 67, pp. 935–943.

Dorn, R. I., 2012. Do Debris Flows Pose a Hazard to Mountain-Front Property in Metropolitan Phoenix, Arizona? *The Professional Geographer,* Vol. 64, No. 2, pp. 197–210.

Eaton, L. S., B. A. Morgan, R. C. Kochel, and A. D. Howard. 2003. Role of Debris Flows in Long-Term Landslide Denudation in the Central Appalachians of Virginia. *Geology,* Vol. 31, pp. 339–342.

Ehrich, R. W. 1977. Detection of Global Edges in Textured Images. *IEEE Transactions on Computers,* Vol. C–26, pp. 589–603.

Ehrich, R. W. 1979. Detection of Global Lines and Edges in Heavily Textured Images. In *Proceedings, Second International Symposium on Basement Tectonics*. Denver, CO: Basement Tectonics Committee Conference held in Delaware, pp. 508–513.

Eliason, P. T., L. A. Soderblom, and P. S. Chavez. 1981. Extraction of Topographic and Spectral Albedo Information from Multispectral Images. *Photogrammetric Engineering and Remote Sensing,* Vol. 48, pp. 1571–1579.

Franklin, S. E., M. B. Lavigne, E. R. Hunt, Jr., B. A. Wilson, D. R. Peddle, G. J. McDermid, and P. T. Giles. 1995. Topographic Dependence of Synthetic Aperture Radar Imagery. *Computers and Geosciences,* Vol. 21, pp. 521–532.

Gathright, W. S., W. S. Henika, and J. L. Sullivan. 1977. *Geology of the Waynesboro East and Waynesboro West Quadrangles, Virginia*. Charlottesville, VA: Commonwealth of Virginia, Department of Conservation and Economic Development, Division of Mineral Resources.

Glenn, N. F., and J. R. Carr. 2003. The Use of Geostatistics in Relating Soil Moisture to RADARSAT-1 SAR Data Obtained over the Great Basin, Nevada, USA. *Computers and Geosciences,* Vol. 29, pp. 577–586.

Godt, J. W., and J. A. Coe. 2007. Alpine Debris Flows Triggered by a 28 July, 1999 Thunderstorm in the Central Front Range, Colorado. *Geomorphology*, Vol. 84, No. 1–2, pp. 90–97.

Goetz, A. F. H. 1984. High Spectral Resolution Remote Sensing of the Land. In P. N. Slater (ed.), *Remote Sensing.* Bellingham, WA: SPIE, pp. 56–68.

Goetz, A. F. H., F. C. Billingsley, A. R. Gillespie, M. J. Abrams, R. L. Squires, E. M. Shoemaker, . . . and D. P. Elston. 1975. *Application of ERTS Images and Image Processing to Regional Geological Problems and Geological Problems in Northern Arizona.* U.S. National Aeronautics and Space Administration, Jet Propulsion Laboratory, Technical Report 32–1597, 188 pp.

Goetz, A. F. H., B. N. Rock, and L. C. Rowan. 1983. Remote Sensing for Exploration: An Overview. *Economic Geology,* Vol. 78, pp. 573–590.

Goetz, A. F. H., G. Vane, J. E. Solomon, and B. N. Rock. 1985. Imaging Spectrometry for Earth Remote Sensing. *Science,* Vol. 228, pp. 1147–1153.

Green, G. M. 1986. Use of SIR-A and Landsat MSS Data in Mapping Shrub and Intershrub Vegetation at Koonamore, South Australia. *Photogrammetric Engineering and Remote Sensing,* Vol. 52, pp. 659–670.

Hewitt, K. 2010. Gifts and Perils of Landslides. *American Scientist,* Vol. 95, pp. 410–419.

Jezek, K. C. 1999. Glaciological Properties of the Antarctic Ice Sheet from RADARSAT-1 Synthetic Aperture Radar Imagery. *Annals of Glaciology,* Vol. 29, pp. 286–290.

Kaab, A. 2002. Monitoring High-Mountain Terrain Deformation from Repeated Air- and Spaceborne Optical Data: Examples Using Digital Aerial Imagery and ASTER Data. *ISPRS Journal of Photogrammetry and Remote Sensing,* Vol. 57, pp. 39–52.

Kervyn, F. 2001. Modeling Topography with SAR Interferometry: Illustrations of a Favorable and Less Favorable Environment. *Computers and Geosciences,* Vol. 27, pp. 1039–1050.

King, R. B. 1970. A Parametric Approach to Land System Classification. *Geoderma,* Vol. 4, pp. 37–46.

Lee, W. T. 1922. The Face of the Earth as Seen from the Air. *American Geographical Society.* Special Publication No. 4, 110 pp.

Lueder, D. R. 1959. *Aerial Photographic Interpretation: Principles and Applications*. New York: McGraw-Hill, 462 pp.
Lyon, J. G. 2001. *Wetland Landscape Characterization: GIS, Remote Sensing, and Image Analysis*. Boca Raton, FL: Lewis, 160 pp.
Lyon, J. G., and J. McCarthy. 1995. *Wetland and Environmental Applications of GIS*. Boca Raton, FL: Lewis Publishers, 373 pp.
Massonnet, D., and K. L. Feigl. 1998. Radar Interferometry and Application to Changes in the Earth's Surface. *Reviews of Geophysics,* Vol. 36, pp. 441–500.
McBratney, A. B., M. L. M. Santos, and B. Minasny. 2003. On Digital Soil Mapping. *Geoderma,* Vol. 117, pp. 3–52.
Metternicht, G. I., and J. A. Zinck. 2003. Remote Sensing of Soil Salinity: Potentials and Constraints. *Remote Sensing of Environment,* Vol. 85, pp. 1–20.
Meyer, B. S., D. B. Anderson, R. H. Bohning, and D. G. Fratianne. 1973. *Introduction to Plant Physiology*. New York: Van Norstrand, 565 pp.
Miller, V. C. 1961. *Photogeology*. New York: McGraw-Hill, 248 pp.
Millington, A. C., and Townshend, J. R. G. 1987. The Potential of Satellite Remote Sensing for Geomorphological Investigations: An Overview. In V. Gardiner (ed.), *International Geomorphology*. Chichester, UK: John Wiley, pp. 331–342.
Moore, G. K., and F. A. Waltz. 1983. Objective Procedures for Lineament Enhancement and Extraction. *Photogrammetric Engineering and Remote Sensing,* Vol. 49, pp. 641–647.
O'Leary, D. W., J. D. Friedman, and H. A. Pohn. 1976. Lineament, Linear, Lineation: Some Proposed New Standards for Old Terms. *Geological Society of America Bulletin,* Vol. 87, pp. 1463–1469.
Pantaleoni, E., R. H. Wynne, J. M. Galbraith, and J. B. Campbell. 2009. A Logit Model for Predicting Wetland Location Using ASTER and GIS. *International Journal of Remote Sensing,* Vol. 30, pp. 2215–2236.
Pantaleoni, E., R. H. Wynne, J. M. Galbraith, and J. B. Campbell. 2009. Mapping Wetlands Using ASTER Data: A Comparison between Classification Trees and Logistic Regression. *International Journal of Remote Sensing,* Vol. 30, pp. 3423–3440.
Perrin, R. M. S., and C. W. Mitchell. 1969. *An Appraisal of Physiographic Units for Predicting Site Conditions in Arid Areas*. Military Engineering Experimental Establishment (MEXE) Report 1111, Vols. I and II, Christchurch, Hants, UK.
Podwysocki, M. H., J. G. Moik, and W. D. Shoup. 1975. Quantification of Geologic Lineaments by Manual and Machine Processing Techniques. In *Proceedings, NASA Earth Resources Survey Symposium*. Houston, TX: NASA, pp. 885–903.
Porter, H. C., W. H. Creggar, and H. C. Hudson. 1985. *Soil Survey of Montgomery County, Virginia*. Washington, DC: U.S. Department of Agriculture, Soil Conservation Service, 158 pp.
Post, D. F., E. H. Horvath, W. M. Lucas, S. A. White, M. J. Ehasz, and A. K. Batchily. 1994. Relations between Soil Color and Landsat Reflectance on Semiarid Rangelands. *Soil Science Society of America Journal,* Vol. 58, pp. 1809–1816.
Prospero, J. M., P. Ginoux, O. Torres, S. E. Nicholson, and T. E. Gill. 2002. Environmental Characterization of Global Sources of Atmospheric Soil Dust Identified with the Nimbus 7 Total Ozone Mapping Spectrometer (TOMS) Absorbing Aerosol Product. *Reviews of Geophysics,* Vol. 40, pp. 1–31.
Raghavan, V., K. Wadatsumi, and S. Masummoto. 1994. SMILES: A FORTRAN–77 Program for Sequential Machine Interpreted Lineament Extraction Using Digital Images. *Computers and Geosciences,* Vol. 20, pp. 121–159.
Rast, M., J. S. Hook, C. D. Elvidge, and R. E. Alley. 1991. An Evaluation of Techniques for the Extraction of Mineral Absorption Features from High Spectral Resolution Remote Sensing Data. *Photogrammetric Engineering and Remote Sensing*, Vol. 57, pp. 1303–1309.
Ray, R. G. 1960. *Aerial Photographs in Geologic Interpretation and Mapping* (Geological Survey Professional Paper 373). Washington, DC: U.S. Government Printing Office, 230 pp.

Rignot, E., J. L. Bamber, M. R. Van Den Broeke, C. Davis, Y. H. Li, W. J. Van De Berg, and E. van Meijaard. 2008. Recent Antarctic Ice Mass Loss from Radar Interferometry and Regional Climate Modeling. *Nature Geoscience,* Vol. 1, pp. 106–110.

Rignot, E., and P. Kanagaratnam. 2006. Changes in the Velocity Structure of the Greenland Ice Sheet. *Science,* Vol. 311, pp. 986–990.

Rignot, E., and R. H. Thomas. 2002. Mass Balance of Polar Ice Sheets. *Science,* Vol. 297, pp. 1502–1506.

Ritter, D. F. R., C. Kochel, and J. R. Miller. 2011. *Process Geomorphology* (5th ed.). Long Grove, IL: Waveland Press, 652 pp.

Robinove, C. J. 1981. The Logic of Multispectral Classification and Mapping of Land. *Remote Sensing of Environment,* Vol. 11, pp. 231–244.

Rowan, L. C., A. F. H. Goetz, and R. P. Ashley. 1977. Discrimination of Hydrothermally Altered and Unaltered Rocks in Visible and Near Infrared Multispectral Images. *Geophysics,* Vol. 42, pp. 522–535.

Simpson, R. B. 1966. Radar, Geographic Tool. *Annals of the Association of American Geographers*, Vol. 56, pp. 80–96.

Smith, M. J., and C. F. Pain. 2009. Applications of Remote Sensing in Geomorphology. *Progress in Physical Geography,* Vol. 33, No. 4, pp. 568–582.

Soil Survey Staff. 1966. *Aerial Photointerpretation in Classifying and Mapping Soils* (Agricultural Handbook 294). Washington, DC: U.S. Department of Agriculture, 89 pp.

Strozzi, T., U. Wegmüller, L. Tosi, G. Bitelli, and V. Spreckels. 2001. Land Subsidence Monitoring with Differential SAR Interferometry. *Photogrammetric Engineering and Remote Sensing,* Vol. 67, pp. 1261–1270.

Summerfield, M. 1991. *Global Geomorphology.* Essex, UK: Longman, 537 pp.

Thomas, M. F. 1969. Geomorphology and Land Classification in Tropical Africa. In *Environment and Land Use in Africa.* (M. F. Thomas and G. W. Whittington, eds.). London: Methuen, pp. 103–105.

U.S. Geological Survey. 2017, November. *Revisiting the Oso Landslide.* Available at *https://usgs.gov/news/featured-story/revisiting-oso-landslide.*

Ustin, S. L., M. D. Smith, S. Jacquemond, M. Verstrate, and Y. Govaerts. 1999. Geobotany Vegetation Mapping for the Earth Sciences. In A. N. Rencz and R. A. Ryerson (eds.), *Manual of Remote Sensing* (Remote Sensing for the Earth Sciences), 3rd ed., Vol. 3. New York: John Wiley, pp. 198–248.

Vanderbrug, G. J. 1976. Line Detection in Satellite Imagery. *IEEE Transactions on Geoscience Electronics,* Vol. GE–14, pp. 37–44.

Vaughan, R. W. 1983. *A Topographic Approach for Lineament Recognition in Satellite Images.* Unpublished masters thesis, Department of Computer Science, Virginia Polytechnic Institute, Blacksburg, 70 pp.

Vincent, R. K., P. K. Pleitner, and M. L. Wilson. 1984. Integration of Airborne Thematic Mapper and Thermal Infrared Multispectral Scanner Data for Lithologic and Hydrothermal Alteration Mapping. In *Proceedings, International Symposium on Remote Sensing of Environment, Third Thematic Conference, Remote Sensing for Exploration Geology.* Ann Arbor, MI: Environmental Research Institute of Michigan, pp. 219–226.

Watershed Sciences. 2006. *Lidar Remote Sensing Data Collection: Desolation Creek, Middle Fork John Day River, and John Day River, Oregon.* Department of Fisheries and Wildlife, Portland, OR: Prepared for Puget Sound Lidar Consortium, U.S. Bureau of Reclamation, and Oregon Trout, 30 pp.

Webster, R., and P. H. T. Beckett. 1970. Terrain Classification and Evaluation by Air Photography: A Review of Recent Work at Oxford. *Photogrammetria,* Vol. 26, pp. 51–75.

Weismiller, R. A., and S. A. Kaminsky. 1978. An Overview of Remote Sensing as Related to Soil Survey Research. *Journal of Soil and Water Conservation,* Vol. 33, pp. 287–289.

Wendt, G. E., R. A. Thompson, and K. N. Larson. 1975. *Land Systems Inventory: Boise National*

Forest, Idaho. A Basic Inventory for Planning and Management. Ogden, UT: U.S. Forest Service Intermountain Region, 49 pp.

Wheeler, R. L., and D. U. Wise. 1983. Linesmanship and the Practice of Linear Geo-Art: Discussion and Reply. *Geological Society of America Bulletin,* Vol. 94, pp. 1377–1379.

Wildey, R. L., and H. A. Pohn. 1964. Detailed Photoelectric Photometry of the Moon. *Astronomical Journal,* Vol. 69, pp. 619–634.

Williams, R. S. (ed.). 1983. Geological Applications. *Manual of Remote Sensing* (2nd ed.). (R. N. Colwell, ed.). Vol. II. Falls Church, VA: American Society for Photogrammetry, pp. 1667–1916.

Wise, D. U. 1982. Linesmanship and the Practice of Linear Geo-Art. *Geological Society of America Bulletin,* Vol. 93, pp. 886–888.

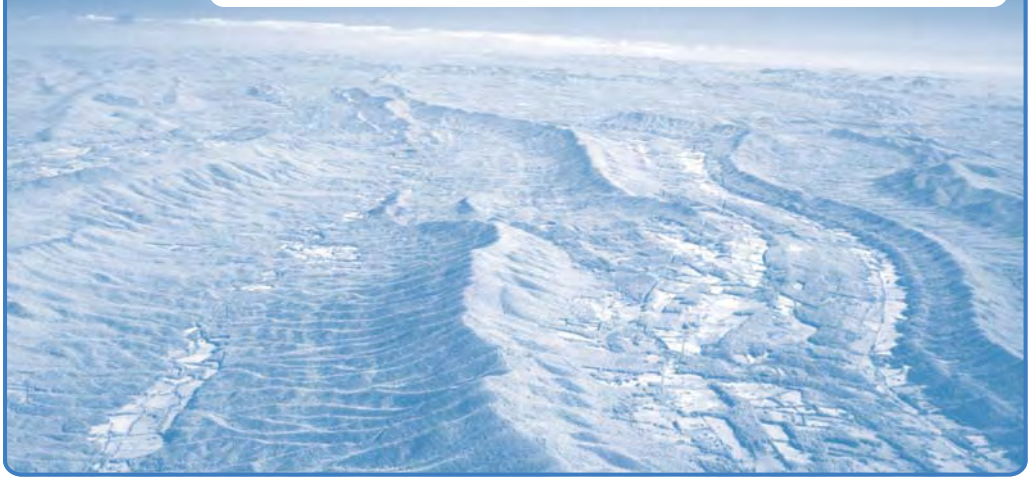
Wooten, R. M., A. C. Witt, C. F. Miniat, T. C. Hales, and J. L. Aldred. 2016. Frequency and Magnitude of Selected Historical Landslide Events in the Southern Appalachian Highlands of North Carolina and Virginia: Relationships to Rainfall. Geological and Ecohydrological Controls, and Effects. In *Natural Disturbances and Historic Range of Variation. Managing Forest Ecosystems* (C. Greenberg and B. Collins, eds.), Vol. 32. Cham, Switzerland: Springer, pp. 203–262.

Wright, J. S., T. C. Vogel, A. R. Pearson, and J. A. Messmore. 1981. *Terrain Analysis Procedural Guide for Soil* (ETL-0254). Ft. Belvoir, VA: U.S. Army Corps of Engineers, Geographic Services Laboratory, Engineering Topographic Laboratories, 89 pp.

Yamaguchi, Y. 1985. Image-Scale and Look-Direction Effects on the Detectability of Lineaments in Radar Images. *Remote Sensing of Environment,* Vol. 17, pp. 117–127.

Zwally, H. J., B. Schutz, W. Abdalati, J. Abshire, C. Bentley, A. Brenner, . . . R. Thomas. 2002. ICESat's Laser Measurements of Polar Ice, Atmosphere, Ocean, and Land. *Journal of Geodynamics,* Vol. 34, pp. 405–445.

20 Coastal Processes and Landforms



MAJOR TOPICS TO UNDERSTAND

- Remote Sensing of Water Characteristics
- Bathymetry
- Coastal Processes and Landforms
- Impact of Hurricane Sandy on Mantoloking, New Jersey
- Lidar for Coastal Erosion of North Carolina Beaches
- Challenges in Coastal Communities—Example of Miami Beach

20.1 INTRODUCTION

Open water covers about 74% of our planet's surface. Oceans account for about 95% of the surface area of open water, but freshwater lakes and rivers (about 0.4%) have a significance that exceeds their small areal extent (Table 20.1).

In addition, soil and rock near the Earth's surface hold significant quantities of freshwater (but only about 0.01% by volume of the Earth's total water), as do the ice and snow of polar regions. Moisture in the form of "permanent" ice (about 5% of the Earth's surface) is largely beyond our reach, although the seasonal accumulation and melt of snow-

TABLE 20.1 Water on the Earth

	Percentage	
	By surface area[a]	By volume[b]
Oceans	94.90	97.1
Rivers and lakes	0.40	0.20
Groundwater	—	0.60
Permanent ice cap	4.69	2.20
Earth's atmosphere	—	0.001

Note: Data from Nace (1967).
[a]Percentage by area of the Earth's total water surface.
[b]Percentage by volume of the Earth's water.

pack in temperate mountains is an important source of moisture for some agricultural regions in otherwise arid zones, such as the Great Basin of North America. Hydrologists and meteorologists monitor water as it occurs in all these forms, as it changes from liquid to vapor, condenses to rain and snow, and moves on and under the Earth's surface (Figure 20.1). In addition, studies of sea ice, movement of pollutants, and ocean currents are among the many important subjects that attract the attention of scientists who study hydrology, oceanography, and related subjects.

Most traditional means of monitoring the Earth's water depend largely on measurements made at specific points or collections of samples from discrete locations. The nature of ocean, lake, and river water can be studied by samples collected at the surface

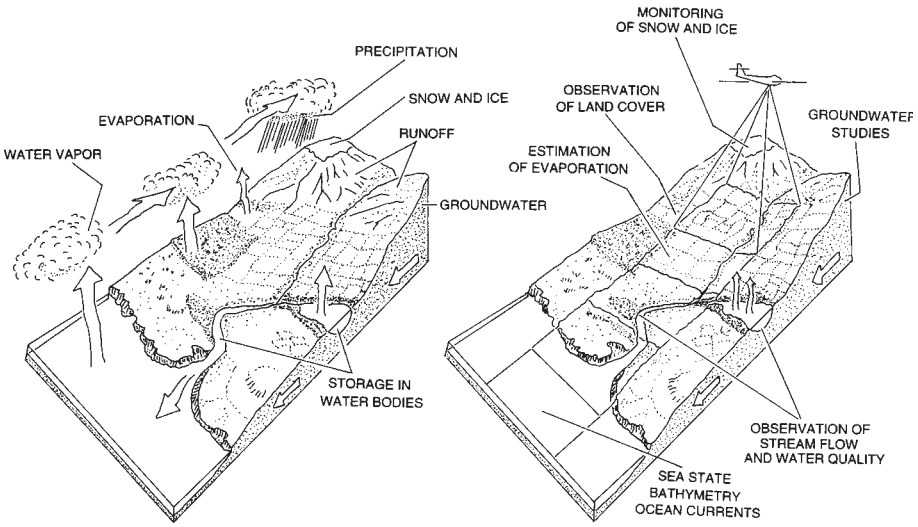


FIGURE 20.1 Schematic diagrams, hydrologic cycle. Left: Idealized representation of key components of the hydrologic cycle, including, in part, evaporation, precipitation, snowfields, groundwater, and runoff. Right: View of application of remotely sensed imagery to acquire information concerning components of the hydrologic cycle.

or by devices designed to gather water at specified depths. Groundwater can be studied by the collection of samples from wells or boreholes. Water samples can be subjected to chemical and physical tests to measure levels of pollutants, to detect bacteria and other biological phenomena, and to examine oxygen levels, sediment content, and many other qualities. Such measurements or samples, of course, provide information about discrete points within the water body, whereas the analyst usually is interested in examining entire water bodies or the regions that contribute to water sources. Although measurements can be collected at several locations to build a record of place-to-place variation within the water body, such efforts are at best a piecemeal approach to studying the complex and dynamic characteristics of water bodies.

Therefore, remote sensing provides a valuable resource to understand broad-scale dynamic patterns that can be difficult to examine in detail using point measurements only. Careful coordination and placement of surface samples permit establishment of relationships between sample data and those collected by remote sensing. Remotely sensed data can be especially valuable in studying phenomena over large areas. Satellite sensors provide the opportunity for regular observation of even very remote regions. Although remotely sensed images seldom replace usual sources of information concerning water resources, they can provide valuable supplements to field data by revealing broad-scale patterns not recognizable at the surface, by recording changes over time, and by providing data for inaccessible regions.

20.2 REMOTE SENSING OF WATER CHARACTERISTICS

Spectral Characteristics of Water Bodies

Spectral qualities of water bodies originate from the interaction of several factors, including radiation incident to the water surface, optical properties of water bodies, surface roughness, angles of observation and illumination, and in some instances, reflection of light from subsurface features. As incident light strikes the water surface, some is reflected back to the atmosphere; such reflected radiation carries little information about the water itself, although it may convey information about the roughness of the surface and therefore about wind and waves. Instead, spectral properties (i.e., "color") of water bodies originate largely by energy scattered and reflected within the water body itself, known as *volume reflection* because it occurs over a range of depths rather than at the surface. Some of this energy is directed back toward the surface, where it again passes through the atmosphere and then to the sensor. This light, sometimes known as *underlight*, forms the primary source of color of a water body.

Light that enters a water body is influenced by (1) absorption and scattering by pure water and (2) scattering, reflection, and diffraction by particles that may be suspended in the water. For pure water, some of the same principles described previously for atmospheric scattering apply. Scattering by particles that are small relative to wavelength (*Rayleigh scattering*) causes shorter wavelengths to be scattered the most. Thus, for deep-water bodies, we expect water (in the absence of impurities) to be blue or blue-green in color (Figure 20.2). Maximum transmittance of light by clear water occurs in the range of 0.44–0.54 μm, with peak transmittance at 0.48 μm. Because the color of water is determined by volume scattering rather than by surface reflection, spectral properties of water bodies (unlike those of land features) are determined by transmittance rather than

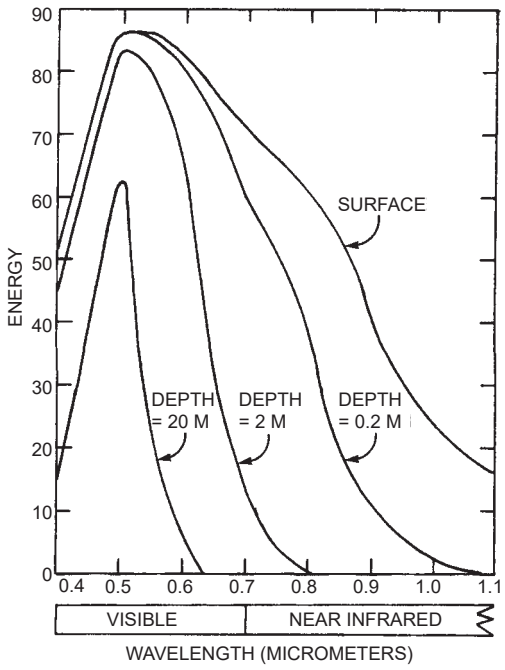


FIGURE 20.2 Light penetration within a clear water body. From Moore (1978, p. 458). Used by permission of the Environmental Research Institute of Michigan.

by surface characteristics alone. In the blue region, light penetration is not at its optimum, but at slightly longer wavelengths, in the blue-green region, penetration is greater. At these wavelengths, the opportunity for recording features on the bottom of the water body is greatest. At longer wavelengths, in the red region, absorption of sunlight is much greater, and only shallow features can be detected. Finally, in the near-infrared region, absorption is so great that only land–water distinctions can be made.

As impurities are added to a water body, its spectral properties change. Sediments are introduced both from natural sources and by human activities. Such sediments consist of fine-textured silts and clays eroded from stream banks, or in water running off disturbed land, which are fine enough to be carried in suspension by moving water. As moving water erodes the land surface or the shoreline, it carries small particles as suspended sediment; faster-flowing streams can erode and carry more and larger particles than can slower-moving streams. As a stream enters a lake or ocean, the decrease in velocity causes coarser materials to settle from the water body. Even slow-moving rivers and currents can carry large amounts of fine-textured sediments, such as clays and silts, which can be found even in calm water bodies. Sediment-laden water is often designated as *turbid water.* Scientists can measure *turbidity* by sampling the water body or by using devices that estimate turbidity from transparency of the water. One such device is the *Secchi disk,* a white disk of specified diameter that can be lowered on a line from the side of a small boat. Because turbidity decreases the transparency of the water body, the depth at which the disk is no longer visible can be related to sediment content. Another indication of turbidity, nephelometric turbidity units, are measured by the intensity of light that passes through a water sample. A special instrument uses a light beam and a sensor to detect differences in light intensity. Water of high turbidity decreases the intensity of the light in a manner that indicates sediment content.

Thus, as sediment concentration increases, the spectral properties of a water body change. First, overall brightness in the visible region increases, so the water body ceases to act as a "dark" object and becomes more and more like a "bright" object. Second, as sediment concentration increases, the wavelength of peak reflectance shifts from a maximum in the blue region toward the green. The presence of larger particles means that the wavelength of maximum scattering shifts toward the blue-green and green regions (Figure 20.3). Therefore, as sediment content increases, there tends to be a simultaneous increase in brightness and a shift in peak reflectance toward longer wavelengths. The peak itself becomes broader, so at high levels of turbidity, color becomes a less precise indicator of sediment content. As sediment content approaches very high levels, water color begins to approach that of the sediment itself. Green reflectance of turbid water appears as light blue on the color-infrared model image (Chapter 4), which provides a sharp contrast in color with the dark blue or black of clear water bodies.

Spectral Changes as Water Depth Increases

Figure 20.3 shows the spectral characteristics of sunlight as it penetrates a clear water body. Near the surface, the overall shape of the curve resembles the spectrum of solar radiation, but the water body increasingly influences the spectral composition of the light as depth increases. So, at a depth of 20 m, little or no infrared radiation is present because the water body is an effective absorber of these longer wavelengths. At this depth, only blue-green wavelengths remain; these wavelengths are therefore available for scattering back to the surface, from the water itself and from the bottom of the water body.

The attenuation coefficient (k) describes the rate at which light becomes dimmer as depth increases. If E_0 is the brightness at the surface, then the brightness at depth z, or E_z, is given by:

$$E_z = E_0 e^{-kz} \qquad \text{(EQ. 20.1)}$$

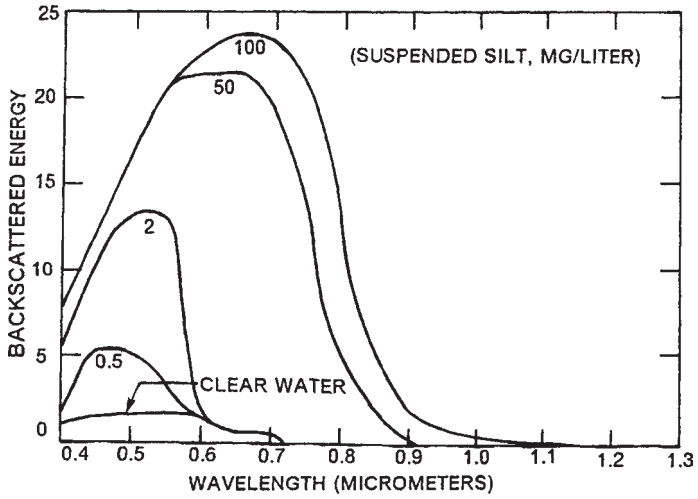


FIGURE 20.3 Effects of turbidity on spectral properties of water. From Moore (1978, p. 460). Used by permission of the Environmental Research Institute of Michigan.

In hydrologic studies, the influence of atmosphere can be especially important. The atmosphere, of course, alters the spectral properties of incident radiation and also influences the characteristics of the reflected signal. Although these influences are also present in remote sensing of land surfaces, they assume special significance in hydrologic studies, in part because such studies often depend on subtle spectral differences (easily lost in atmospheric haze) and also perhaps because much of the hydrologic information is carried by short wavelengths that are easily scattered by the atmosphere.

Water bodies are typically dark, so the analyst must work with a rather restricted range of brightnesses relative to those available for study of land surfaces. As a result, analysts specializing in remote sensing of water bodies must devote special attention to the radiometric qualities of the remotely sensed data. Typically, data to be analyzed for hydrologic information are examined carefully to assess quality, Sun angle, and the effects of the atmosphere. Geometric preprocessing is used with caution, to avoid unnecessary alteration of radiometric qualities of the data. In some instances, analysts calculate average brightness over blocks of contiguous pixels to reduce transient noisy effects of clouds and whitecaps. Sometimes, several scenes of the same area, acquired at different times, can be used to isolate permanent features (e.g., shallows and shoals) from temporary features (including waves and atmospheric effects). Also, it is often advantageous to estimate original radiometric brightnesses from digital values as a means of accurately assessing differences in color and brightness.

Lidar Penetration at Different Depths

Under favorable circumstances, lidar (Chapter 9) has the ability to assess water depth and to differentiate among varied classes of subsurface vegetation. Lidar returns from water bodies have characteristic form if the water surface is calm, the water is clear, and depths are within limits of penetration of the signal. Lidar penetration in clear water is often described as reaching within two to three Secchi depths—perhaps as deep as 30 m in some instances. Wang and Philpot (2007), found that their analysis could correct for scan angle effects, assess slope of subsurface topography, and separate among classes of bottom vegetation. Figure 20.4 depicts an idealized lidar waveform (i.e., the accumulated history of the returned signal) from a water body. The base of the waveform is formed by backscattering from the water column, with distinct peaks formed by returns from the water surface and from the subsurface topography.

Roughness of the Water Surface

Figure 20.5 shows spectra for calm and wave-roughened surfaces observed in the visible and the near infrared. Wave-roughened surfaces are brighter than smoother surfaces. Calm, smooth, water surfaces direct only volume-reflected radiation to the sensor, but rough, wavy water surfaces direct a portion of the solar beam directly to the sensor. As a result, wavy surfaces are much brighter, especially in the visible portion of the spectrum. McKim, Merry, and Layman (1984) described a procedure that uses polarizing filters to separate surface brightnesses from those of the water body itself, using high-resolution data.

The most intensive studies of ocean roughness have used active microwave sensors (Chapter 8). *Sea state* refers to the roughness of the ocean surface, determined by wind

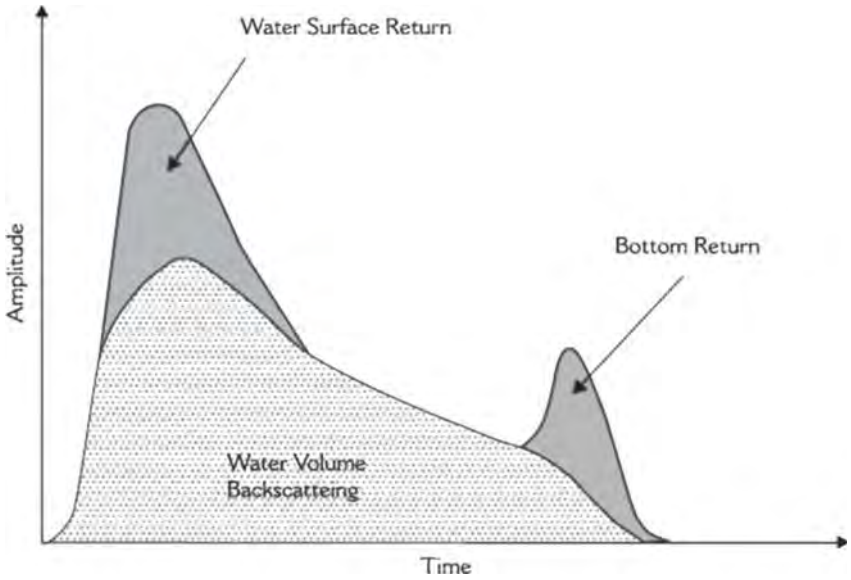


FIGURE 20.4 Example of a generic bathymetric lidar waveform, illustrating the contributions of lidar returns contributed by the water surface, water column, and shallow subsurface features. From Wang and Philpot (2007, p. 124). Copyright © 2007. Used by permission of Elsevier.

speed and direction as they interact locally with currents and tides. Sea state is an important oceanographic and meteorological quality because, if studied over large areas and over time, it permits inference of wind speed and direction—valuable information for research and forecasts. Radars provide data concerning ocean roughness; the backscattering coefficient increases as wave height increases. Radars therefore provide a means of indirectly observing sea state over large areas and a basis for inferring wind speed and direction at the water's surface. If the imaging radar is carried by a satellite, as was the case with Seasat (the first satellite designed for oceans, a synthetic aperture radar [SAR] system) and the shuttle imaging radar, then the analyst has the opportunity to observe

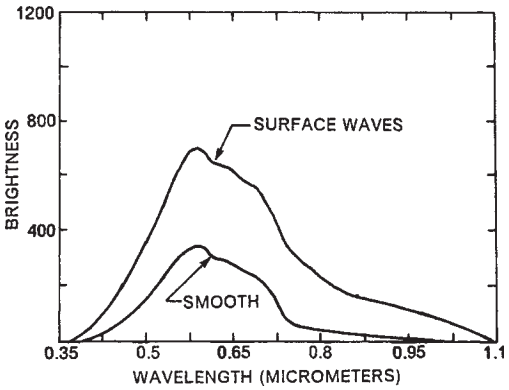


FIGURE 20.5 Spectra of calm and wind-roughened water surfaces (from laboratory experiments). Adapted from McKim et al. (1984, p. 358). Copyright © 1984 the American Society for Photogrammetry and Remote Sensing. Used by permission.

sea state over very large areas at regular intervals. These observations permit oceanographic studies that were impractical or very difficult using conventional data. Furthermore, timely information regarding sea state has obvious benefits for marine navigation and can, in principle, form the basis for inferences of wind speed and direction that can contribute to meteorological data. Direct observation of sea state is relatively straightforward, but because ocean areas are so very large and conditions are so rapidly changeable, usual observations from ships in transit are much too sparse to provide reliable data.

Rigorous study of sea state by radar started soon after World War II, continuing into the 1950s and 1960s, culminating with the radars carried by Seasat and the U.S. space shuttle. Although many experiments have been conducted using a variety of microwave instruments, one of the most important broad-scale sea state experiments was based on the Seasat A project, which used several instruments to monitor the Earth's oceans. The Seasat SAR and RADARSAT (a Canadian SAR satellite program designed for ocean monitoring) have provided an opportunity to observe large ocean areas on a repetitive basis. Calm ocean surfaces, with waves that are small relative to the radar wavelength, act as specular reflectors and appear as dark regions on the image, as energy is reflected away from the antenna. As wind speed increases, the ocean surface becomes rougher and acts more like a diffuse reflector, causing bright regions on the imagery (**Figure 20.6**). Because radar wavelength and system geometry are known, the received signal can form the basis for estimates of wave height and velocity.

FIGURE 20.6 Seasat SAR image illustrating a rough ocean surface, influenced by winds and currents. Nantucket Shoals, August 27, 1978. From NASA, Jet Propulsion Laboratory.

20.3 BATHYMETRY

Bathymetry refers to water depth and configuration of the ocean floor. Bathymetry is one of the most basic forms of hydrographic data, especially significant near coastlines, harbors, shoals, and banks. Accurate bathymetric data are essential where shallow water is common, presenting hazards to navigation, and where sedimentation, erosion, and scouring of channels alters underwater topography. Subsurface features are often subject to notable changes that are not observable at the surface, so bathymetric observations are often a vital component of marine navigation and mapping.

Water depth can be measured by instruments carried on vessels, especially by acoustic (sonar) instruments that measure depths directly below the vessel. Ideally, bathymetric maps should be compiled from a more or less uniform network of depth measurements rather than from a limited set of data from traverses. Although modern side-scanning sonar can yield accurate depth information, aerial imagery is an important means of mapping subsurface topography in shallow water.

Photogrammetry (Chapter 4) can be applied to bathymetric measurement if high-quality, large-scale photography is available and if the water is clear. Filters have been used to separate radiation in the spectral region 0.44–0.54 μm, where solar energy is most easily transmitted by clear water. As Figure 20.2 indicates, in principle, sunlight can penetrate to depths of about 20 m. Aerial photographs record information only from relatively shallow depths, although depths up to 16 m have been mapped using aerial photographs.

Special problems in applications of photogrammetry to bathymetry include estimation of differences in refraction between air and water and the difficulty in acquiring a reliable underwater control point. Furthermore, mapping is difficult in regions far removed from the shoreline because of the difficulty in extending control across zones of deep water.

Multispectral Bathymetry

Among the many capabilities offered by Landsat and related technologies is the ability to examine imagery to derive bathymetric information from an overhead perspective. Polcyn and Lyzenga (1979) reported that, for their study region in the Bahamas, under optimum conditions, maximum penetration for Landsat Multispectral Scanner System (MSS) band 1 might be as deep as 15–20 m, and about 4–5 m for MSS band 2 (see Figure 20.7).

The ability to observe subsurface features decreases exponentially with depth (i.e., brightness decreases as depth increases). Thus, MSS band 1 (positioned at or near the spectral region of maximum penetration) is useful for estimating depth. Observed dark values suggest that subsurface features are deep, beyond the 20-m subsurface range represented in Table 20.2. Bright values suggest that the subsurface is near the surface (Figure 20.8). Polcyn and Lyzenga (1979) have contributed to bathymetric analysis by defining an effective depth-finding algorithm. Jie et al. (1992), Gao (2009), Jawak et al. (2015a, 2015b), are among those many researchers who have investigated this topic. And in 1985, Jupp et al. summarized published values as estimates of maximum penetration of solar radiation within clear, calm ocean water under a clear sky, often applied for examination of Landsat imagery:

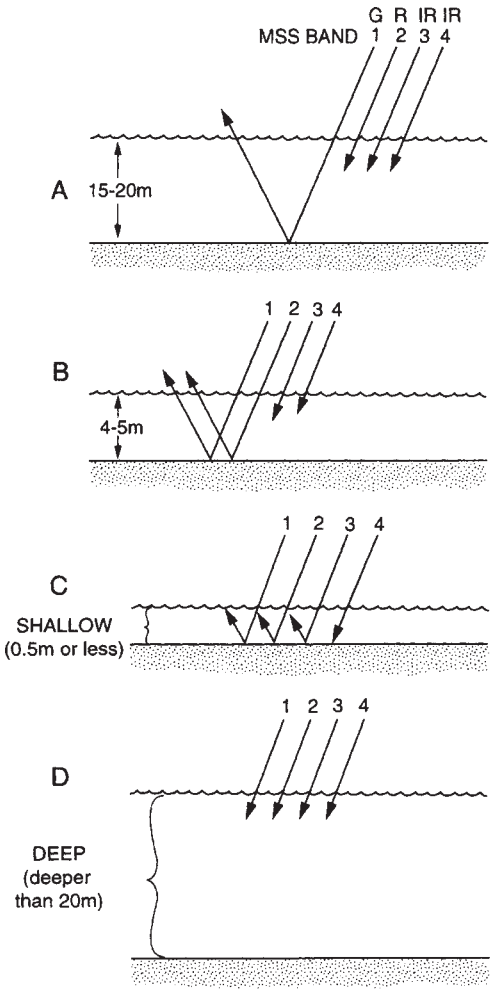


FIGURE 20.7 Multispectral bathymetry (illustrated here using Landsat MSS green, red, and infrared bands). Relative brightnesses of the several spectral bands indicate depth classes, as illustrated.

TABLE 20.2 Logarithmic Transformation of Brightness

	Ln(*x*)	Brightness	Depth
2	0.69	Dim	Deep
6	1.79		
10	2.30		
14	2.64		
18	2.89		
22	3.09		
26	3.26	Bright	Shallow

- MSS band 1: 15–20 m (50–60 ft)
- MSS band 2: 4–5 m (13–17 ft)
- MSS band 3: 0.5 m (1.5 ft)
- MSS band 4: 0

Thus, MSS4, positioned at or near the spectral region of maximum penetration, is useful for estimating depth. Dark values suggest that the bottom is deep, beyond the 20-m range represented in Table 20.2. Many depth extraction algorithms use logarithmic transformations of MSS 4 brightness as a means of applying our knowledge that brightness decreases exponentially with depth.

More recently, Jagalingam, Akshaya, and Hegde (2015) applied Landsat 8 data to investigate use of bathymetry to examine its effectiveness using Landsat 8 (Figure 20.9). Their study examined the effectiveness of bathymetry mapping of the southwest coast of India (13° 0′ 0′ N and 74°50′ 0′ E) by applying the ratio transform algorithm on the blue and green bands of Landsat 8 satellite imagery. Statistical metrics such as the coefficient of determination (R^2), root mean square error, and mean absolute error are computed between the algorithm's derived value and the hydrographic chart sounding value. Their result reveals a good correlation between the algorithm's derived value and hydrographic chart sounding values.

These procedures seem to be effective if the water body is clear (otherwise, turbidity contributes to brightness) and bottom reflectance is uniform. When attenuation of water and bottom reflectivity are known, accuracies as high as 2.5% have been achieved,

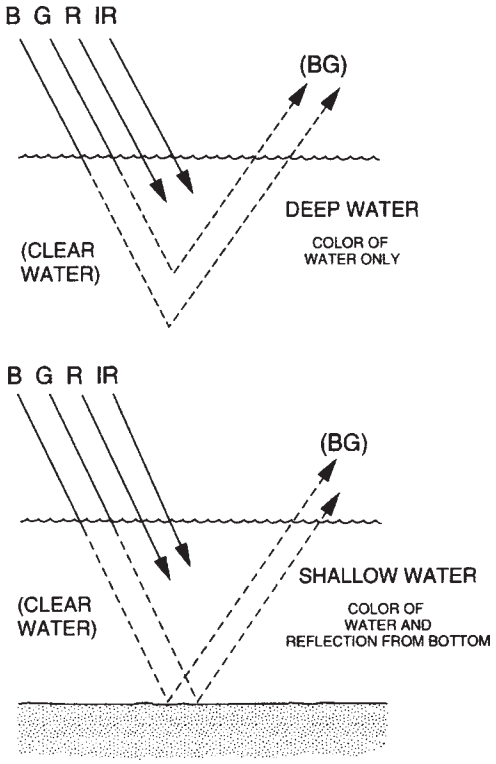


FIGURE 20.8 Multispectral bathymetry. Deep water (below about 20 m) does not reveal subsurface features because light reaching such depths will not return to the surface. Shallow water (see Figure 20.7) will reflect light from the ocean floor back to the water surface, revealing the presence of subsurface features. Radiation reflected from the ocean floor may convey information concerning the ocean floor and the colors of shallow water.

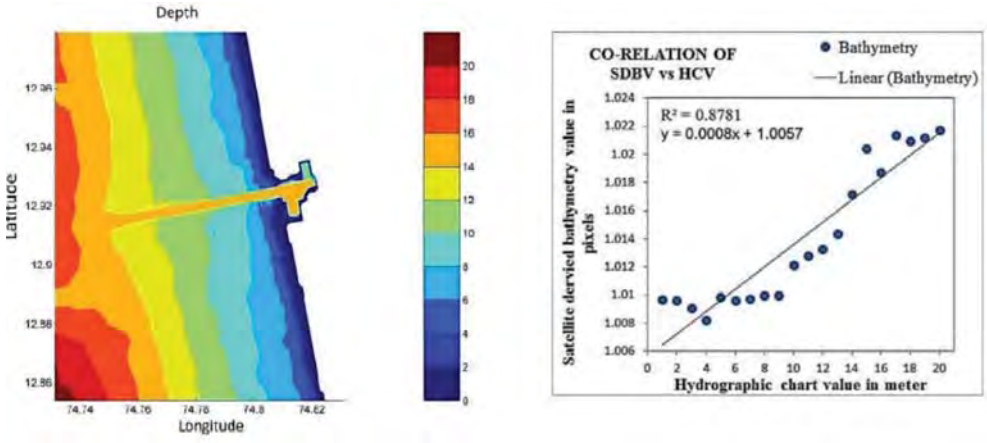


FIGURE 20.9 Multispectral bathymetry using Landsat 8 image data. Relative brightnesses of the several spectral bands indicate depth classes, as depicted by the color scale at the right of the map. The chart at the right indicates the depths of bathymetric readings plotted to indicate depths as assessed. From Jagalingam et al. (2015). Copyright © 2015. Used by permission of Elsevier.

although typical accuracies are lower. If the bottom reflectance is not uniform, then differences in brightness will be caused by differences not only in depth, but also in reflectivity from the subsurface. Because such differences are commonly present due to contrasting reflectances of differing sediments and vegetative cover, it is often necessary to adjust for differing subsurface reflectance. The effectiveness of this procedure varies with Sun angle (due to variations in intensity of illumination) and is most effective when data of high radiometric resolution are available.

In some instances, the ratios of two bands may remove differences in bottom reflectivity. Lyzenga (1979) defines an index that separates differing reflectivities using differing spectral bands. His index, calculated for each pixel, can be applied to classify pixels into differing reflectivity classes, each of which can then be examined to estimate water depth.

20.4 COASTAL PROCESSES AND LANDFORMS

Worldwide, coastal zones form regions significant for settlement, marine shipping, manufacturing, fishing, and more generally, interfaces between land and sea. This section discusses some of the processes that shape the nature of coastal landforms and processes that guide their form and character. Our discussion refers mainly to unconsolidated marine sediments and forces that position and shape coastal features.

Origins and Nature of Oceanic Waves

Although ocean waves can originate from storms, the majority of ocean waves originate far offshore in regions characterized by currents, unsettled weather, and prevailing winds (Figure 20.10). Surface waves are typically generated far from coastline zones, by prevailing winds that intersect the water surface ("ocean surface waves"). The greater the *wind*

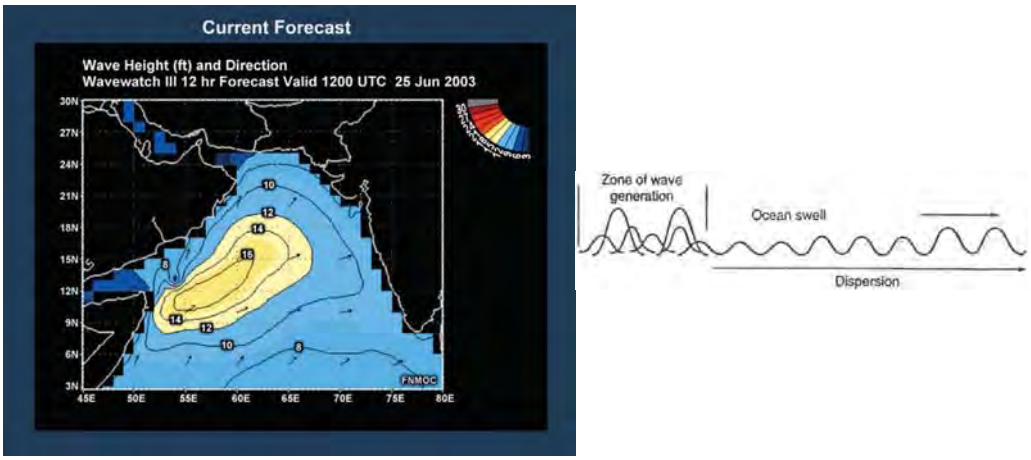


FIGURE 20.10 Left: U.S. Navy wave forecast for the northern Indian Ocean (the pastel region) and dispersal of waves to the broader oceanic zone (blue). From U.S. Navy. Right: Schematic sketch depicting local generation of ocean waves and their dispersal as they reach the broader ocean area.

speed, the larger are the waves and the greater the *duration of the wind*. The broader the area influenced by wind (known as *fetch*), the larger and stronger the waves. The most dramatic waves (such as those of hurricanes, *tsunamis*, or earthquakes) have significant effects but are less common.

Often, rough waves are buffered by prevailing conditions, thereby reducing their size and motion (Figure 20.11) as they reach broader oceanic regions.

- *Wavelength* is the lateral distance between successive wave crests.
- *Wave base* is the subsurface depth at which wave motion ceases, assessed as half of wavelength.

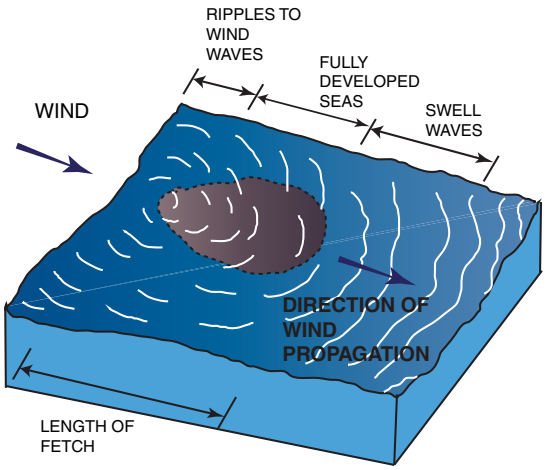


FIGURE 20.11 Steady winds generating ocean waves as they blow over the ocean's fluid surface, far from coastlines. These *wind waves*, known as *swells*, travel for long distances before reaching coastlines. The shaded section in the diagram highlights the focus of wind effects. The term *fetch* signifies the distance traveled by wind, generated without obstruction. From Hoa Tran. Used by permission.

At the wave base and below, wave motion is minimal. Above the wave base, water droplets move in circular patterns within the water column as illustrated in Figure 20.11—larger near the surface, smaller near the wave base. Despite appearances, wave motion transports the energy of wind and waves but does not transport water, which has a vertical motion within the wave.

Figures 20.10–20.12 illustrate several additional features and processes related to ocean waves that can be characterized with remote sensing, as follows.

- *Waves of oscillation:* Within the open ocean, ocean swell tends to suppress larger waves, often distant from coastlines, shallow water, and islands. Such waves are *waves of oscillation,* characteristic of the open ocean (Figure 20.10, right).
- *Waves of translation:* As oceanic waves approach a coastline, the wave base will intersect shallow waters of the coastline, causing vertical motion of oceanic waves to cascade toward the beach, tipping forward as they approach the coastline. Such waves are *waves of translation,* characteristic of the waves we see at the beach (Figure 20.11).
- *Beach dynamics:* Typically, beach waves approach a coastline at an angle, as a function of prevailing winds or local topography (Figure 20.12). Because breaking waves systematically carry sediment to the beach, tipping forward as indicated in Figure 20.11, wave action maintains the local form of the beach.
- *Swash* is the forward cascade of a wave toward the beach, casting the wave several feet forward on the beach, to flow back to the breaking surf in a zig-zag pattern, as illustrated in Figure 20.12. Although each wave transports a small amount of sediment, the net effect is to transport substantial amounts of sediment along to the coastline.
- *Backswash* (or backwash) is the return flow of a wave's water back toward the edge of the shore (typically a few feet or more distant from the wave itself). Back-

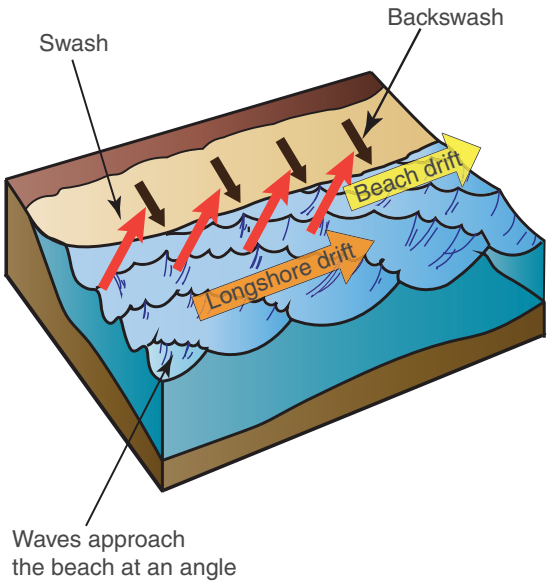


FIGURE 20.12 Swash, backswash, and longshore drift. *Swash* is the forward spill of waves as they reach the beach. *Backswash* is the return flow of water toward incoming waves, to be transported laterally along the coastline, as represented in the diagram. *Beach drift* is the lateral transport of sediment along the coastline at the edges of breaking waves. *Longshore drift* is the lateral flow of currents and sediment, offshore, parallel to the beach. From Hoa Tran. Used by permission.

swash does not flow directly back to the ocean, but rather laterally along the beach, transporting water, and beach sand suspended in the water, to contribute to the form of the beach (Figure 20.12).

- *Longshore drift* is sediment drift parallel to the coastline, as oblique waves channel water movement to flow parallel to the coastline, typically many meters offshore (Figure 20.12). Longshore drift contributes to the evolution of shorelines, as sediment, wind direction, and other influences shape the formation and evolution of beach systems.

These processes transport sediments parallel to the coastline, supporting the overall integrity of the coastal barrier that protects the coastline. Summer beach profiles, formed from offshore marine sediments, create sandy berms positioned above the shoreline. These form an elevated berm parallel to the beach, sometimes designated as "recreational beaches" (Figure 20.13). Equally significant are the effects of winter's rough weather, which can bring offshore sediments to the shore. In brief, oceanic processes can transport marine deposits closer to the shore as the summer season ends, and winter storms can disrupt the more stable nature of the summer season.

As summer weather diminishes, winter weather can bring waves strong enough to bring offshore deposits close enough to the coastline to erode the summer's beach profile (Figure 20.14, top). Alternatively, in some situations, winter weather can accumulate ample offshore sediments to form buffers large enough to shield the shoreline by intercepting winter waves offshore, forcing them to break before they reach the shoreline (Figure 20.14, middle). Or sometimes, very shallow, gentle slopes are established far enough offshore to protect beaches from winter storms by intercepting large waves before they reach the coastline. Later, as winter weather declines, milder summer weather will begin to restore the beach profiles described previously, restoring berms (Figure 20.14, bottom).

Barrier Islands

Barrier islands are crucial for the resilience of our coastlines and are dynamic systems heavily influenced by ocean activity and waves. Remote sensing allows us to monitor and study these landforms, their changes over time, and their influence on coastal processes.

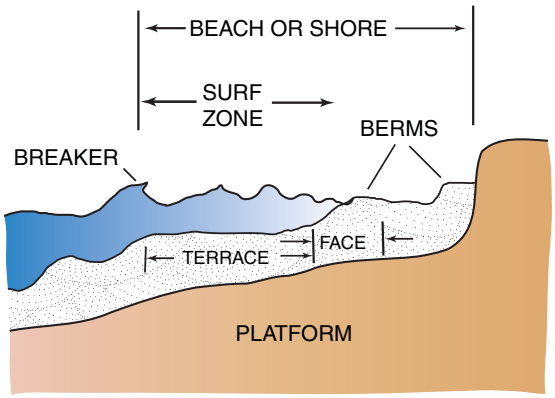


FIGURE 20.13 Surf zone. Cross-sectional sketch of the recreational beach, as might be observed in summer. Summer weather will typically allow calm waves to build coastal terraces, berms, and beaches that will endure during the summer season. From Hoa Tran. Used by permission.

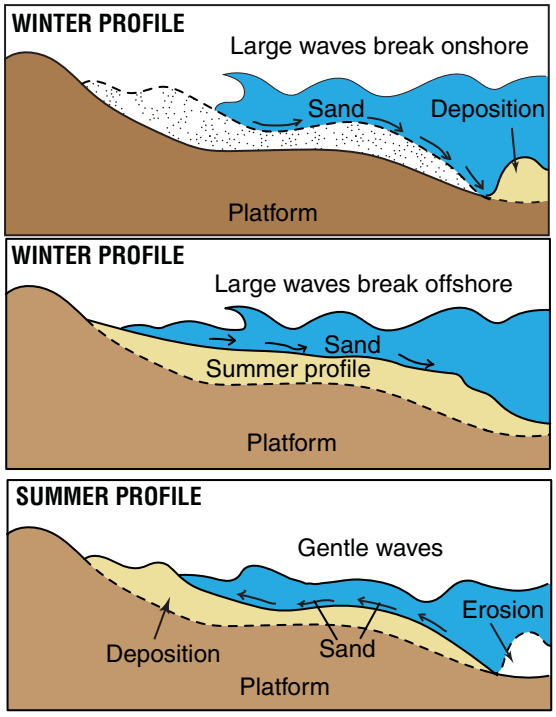


FIGURE 20.14 Winter and summer beach profiles. Top: If winter storms break near the beach, large waves may destroy the beach profile formed during summer months. Middle: In some situations, especially if the offshore coast is shallow, waves may create offshore deposits that are large enough to cause winter waves to break offshore before they reach the summer beach, thereby retaining portions of the summer beach. Bottom: The summer profile is formed by the persistent, but calm, water of the summer season. Although summer storms can occur in spring and into the summer season, the overall summer profile tends to form beaches and berms. Image by Susmita Sen.

In this section, we describe the barrier islands of Virginia and provide some imagery and illustrations. Note that barrier islands exist in many other places along the coast.

The barrier islands of Virginia form a continuous chain of elongated low-lying sand and scrub islands, separated by narrow inlets. Shallow marshes and tidal bays are positioned along the entire coast of Virginia's end of the Delmarva Peninsula. The islands have been uninhabited since 1936, when the last remaining town (Broadwater, Virginia) was evacuated due to a hurricane. Several of the islands are part of the Virginia coast reserve, and overall the region is very important for coastal ecology. They offer Virginia's coastline an important natural resilience to sea-level rise and thus are frequently monitored and studied. Remote sensing offers synoptic and detailed views of their changes over time.

Virginia's barrier islands terminate at their southern extent at the mouth of the Chesapeake Bay and are marked at the north by a barrier spit (Fenwick Island) that crosses the Maryland and Delaware border. In Figure 20.15 (right), land areas appear in silhouette due to underexposure to record hydrographic features. As a result, the land–water interface is very sharply delineated. The Chesapeake Bay occupies the western portion of the image. Virginia's coastal waters are visually distinct from those of the bay along the curved front visible at center right of the image, due to currents, surface oils, temperature, and wind and wave patterns.

The islands themselves are long, low strips of sand shaped by the action of both water and wind (Figure 20.16). At the edge of the ocean, a long beach slopes toward the water; the lowest sections of the beach are influenced each day by the effects of tides

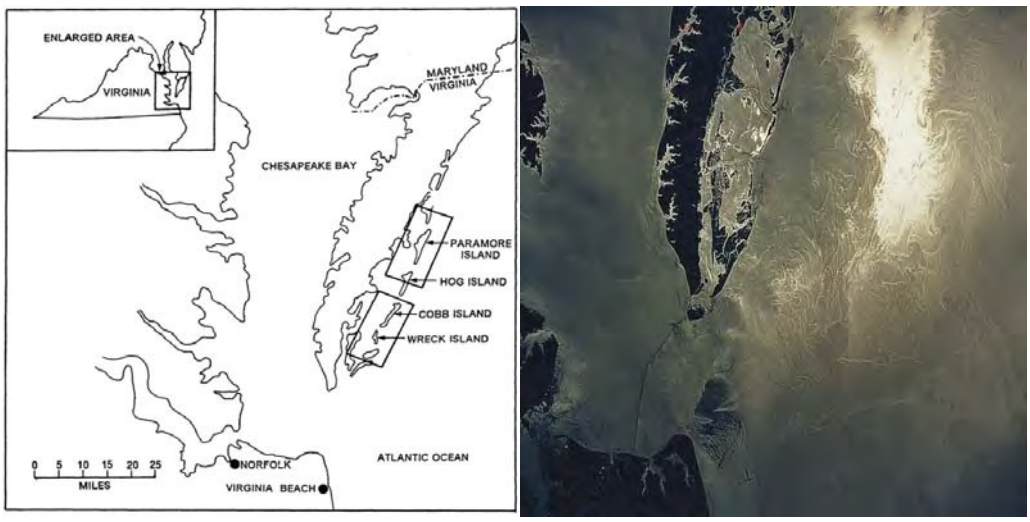


FIGURE 20.15 Virginia's barrier islands. Left: Map of the eastern shore of Virginia. Right: Photograph taken on June 8, 1991, from the space shuttle. The Chesapeake Bay is the water body on the left of the image, with the Atlantic Ocean on the right. Norfolk and Virginia Beach, Virginia, are on the land mass in the lower left of the image. The land mass in the center of the image is most of the Virginia portion of the Delmarva Peninsula and the barrier islands. From NASA, no. STS-40-614-047 (right).

and currents. The upper, higher sections are influenced only by the highest tides and strongest storms, so these regions may experience major changes only once or twice a year, or even less often. Inland from the upper beach ridges is windblown sand formed into dunes that are generally above the reach of waves, although very strong storms may alter these regions. Dunes are reshaped by winds, but their general configuration is often stabilized by grasses and small shrubs that cover much of this zone. Inland from the dunes, elevations decrease, and water again assumes a dominant influence in shaping ecosystems. Water from tidal marshes between the islands and the mainland rises and falls with the tides but is without the strong waves that characterize the seaward side of the islands.

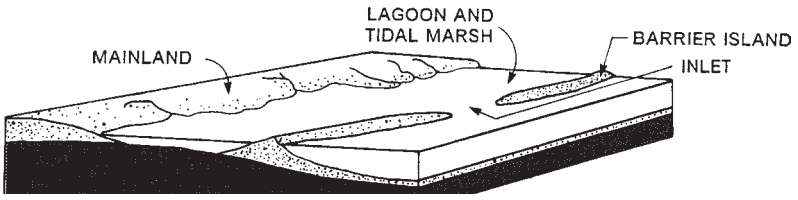


FIGURE 20.16 Schematic sketch of Virginia's barrier islands, formed from unconsolidated sediments transported southward by currents discharged by the Delaware River and other rivers to the north. Currents following the Atlantic shoreline carry sediments to the south, creating offshore sandbars.

Thermal images of the Atlantic coast were acquired shortly after noon on July 23, 1972, using an infrared scanner sensitive in the thermal region. The setting is depicted in a broader context in Figures 20.17 and 20.18; the coastline is protected by a series of barrier islands formed by sediment transported from the north and periodically reshaped by currents and storms. These tidal flats are covered with vegetation adapted to the brackish water; they experience tidal fluctuations and have poorly drained soils. Many of these features are clearly visible on the infrared images of this region (see the tops of Figures 20.17 and 20.18). Bright white strips along the seaward sides of the islands contrast sharply with the dark, cool water of the open ocean. This bright strip is, of course, the warm surface of the open sandy beach, which has received direct energy of the solar beam for several hours and now, at midday, is very hot. Inland from the beach, dunes are visible as slightly darker regions, with darker areas caused by shadowing formed (Figure 20.19). The topographic structures of beach ridges are clearly visible in several regions. In tidal marshes, image appearance is controlled largely by vegetative cover, which provides a clear delineation of edges of the open channel not usually visible on other images. The open water in tidal lagoons has lighter tones than that of the open ocean, indicating its warmer temperatures. Shallow water and restricted circulation enabled water in these areas to absorb the solar radiation it had been receiving for several hours by the time the

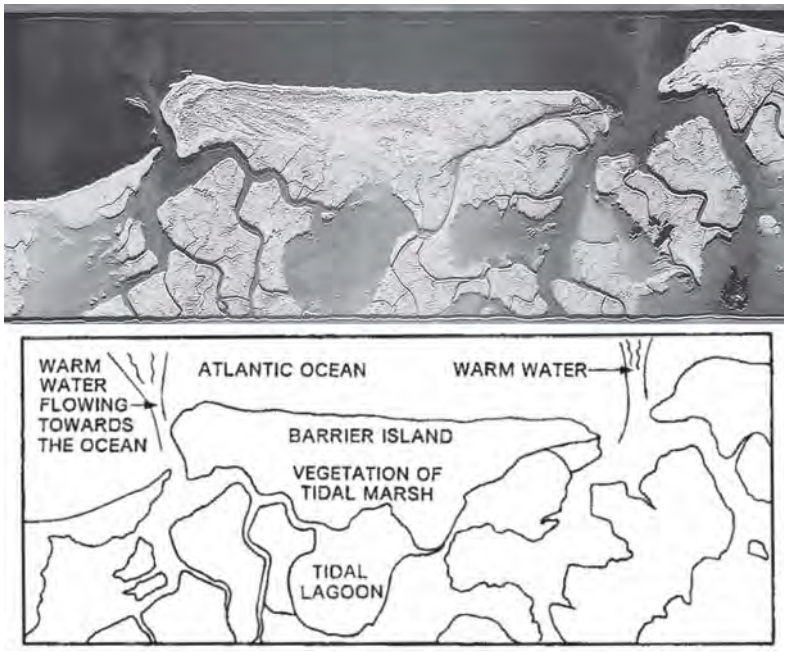


FIGURE 20.17 Top: Thermal infrared image of the eastern shore of Virginia, July 23, 1972, approximately 1:50 P.M. This area is north of the image in **Figure 20.19**. A portion of Hog Island is visible on the top right of the image, overlapping with **Figure 20.18**. Swash Bay, a tidal lagoon, can be seen in the bottom center of the image. Thermal differences are seen in the varying brightnesses of the interior water bodies compared to the cool ocean on the top of the image. From NASA. Bottom: Schematic of the same area, with some areas of warmer water indicated.

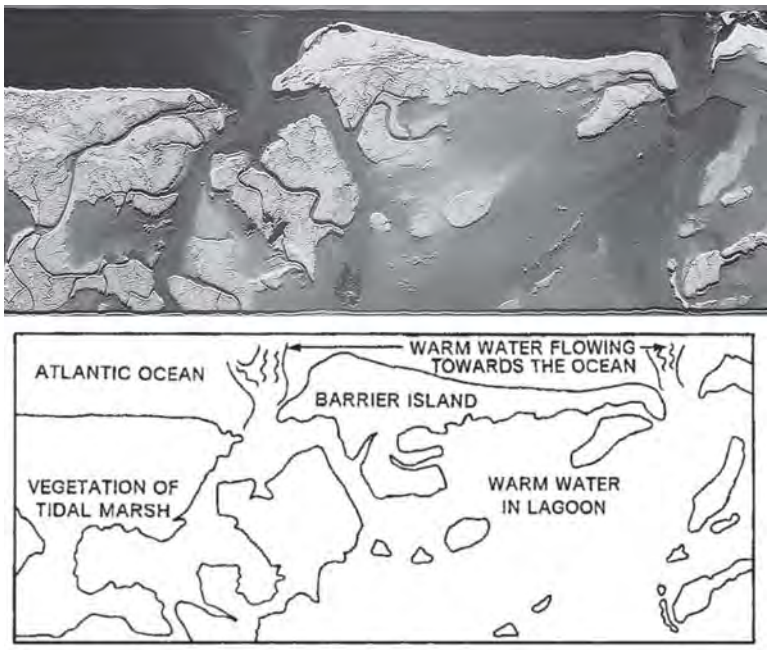


FIGURE 20.18 Top: Thermal infrared image of the eastern shore of Virginia, July 23, 1972, approximately 1:50 P.M. Hog Island can be seen in the center near the top of the image. From NASA. Bottom: Schematic of the same area highlighting thermal differences in the water bodies.

images were acquired. Nonetheless, this water is still cooler than water flowing from the land surface; it is possible to see bright (warm) plumes of water from streams that flow into tidal marshes from the mainland. Finally, tidal currents are clearly visible as bright (warm) streams of water passing between barrier islands to enter the darker (cooler) water of the open ocean. This image captures the tide flowing outward toward the open ocean (ebb tide).

River Systems

Remote sensing is also used to monitor our river systems, including how rivers interact with other rivers and with coastlines, oceans, and lakes; significant impacts are often hidden, ignored, or neglected. Rivers, through their geographic, chemical, and geologic dimensions, contribute to many unrecognized physical, chemical, and biologic dimensions. The flow of minor streams is likely to reach ponds or small lakes without pronounced impacts, although such flow likely includes overland flow, subsurface transport, wetlands, and vernal pools. (See, for example, Williams [2005], *The Biology of Temporary Waters*.) As river flows reach large bodies of open water, they can form currents directed offshore, then disperse into broad plumes, settle into offshore currents, or reside at the subsurface, sometimes far offshore. Large water bodies, if characterized by currents, can capture river flow, or some rivers may have flow powerful enough to dominate

FIGURE 20.19 The beach at Hog Island. The sandy shore is warmer than the surrounding water, causing the thermal images in **Figure 20.18** (top) to appear brighter on land. The cooler waves and larger waves of the Atlantic Ocean can be seen on the left side of the image. Photograph by Gordon Campbell, At Altitude Gallery. Used by permission.

flow within a lake. (See the U.S. Geological Survey's [USGS's] Water Science School at *www.usgs.gov/special-topics/water-science-school*.) A detailed description of all of these processes is beyond the scope of this book, but readers are encouraged to view the following works for further background.

- Borneman, 2014. How Rivers Change the Landscape
- Bianchi, 2013. Where the River Meets the Sea
- Alizadeh et al., 2018. Effect of River Flow on the Quality of Estuarine Coastal Waters Using Machine Learning Models
- Pilkey and Cooper, 2014. *The Last Beach*
- Williams, 2006. *The Biology of Temporary Waters*

Jetties

Jetties occur in various designs and sizes but are basically barriers constructed to maintain river flow as it reaches the coastline of a water body such as a lake or ocean. A jetty is constructed as a barrier, extending along at least one side of the river channel as it leaves the coastline and into a water body (Figure 20.20). Construction often maintains clear

channels through the gap between jetties through "tidal scour" (allowing tidal motion to carry sediment to open water), or through dredging with sand pumps, augers, or vacuum technologies. The jetty maintains river flow on its existing trajectory, preventing lake or ocean currents from blocking or redirecting the river channel as it leaves the coastline.

Figure 20.21A depicts an example of a jetty directing the flow of a river into a lake near Kalispell, Montana. Here, the Flathead River, originating near Glacier National Park, winds through uplands and then travels through lower terrain near Kalispell before reaching lower landscapes. The area shown in Figure 20.21A illustrates the river flowing through agricultural landscapes, transporting sediment downstream before discharging into Flathead Lake. As the river enters the lake, its velocity declines, forming plumes as it enters the lake. Note that the jetty captures sediments on its upstream side as the lake currents approach the river channel, thus requiring occasional dredging or amendment. Without jetties, large rivers are free to change course as they reach the coastline, often accumulating deposits at the mouth of the river.

Numerous examples of jetties can be found along the coastline of the United States directing rivers into the ocean. A few examples on the east coast of the United States include the jetties at Virginia Beach, the Indian River Inlet in Delaware, and the Masonboro Inlet in North Carolina. Jetties into the oceans or very large lakes, such as the Great Lakes, are often significant structures with large channels and clearly visible barriers (Figure 20.21B, C). In some cases, the structures are huge waterways with multiple jetties, canals, and rivers, and significant boat activity. A good example is the shipping channel in Port Arthur, Texas (the Sabine-Neches Waterway and Sabine Pass Ship Channel) (Figure 20.22). The Neches River, which passes through 400 miles in Texas, discharges into the Gulf of Mexico, following the border between Louisiana and Texas. Long barriers on either side of the shipping channel extend far into the Gulf of Mexico, to prevent northbound currents from traversing across Sabine Pass and diverting the path of the channel.

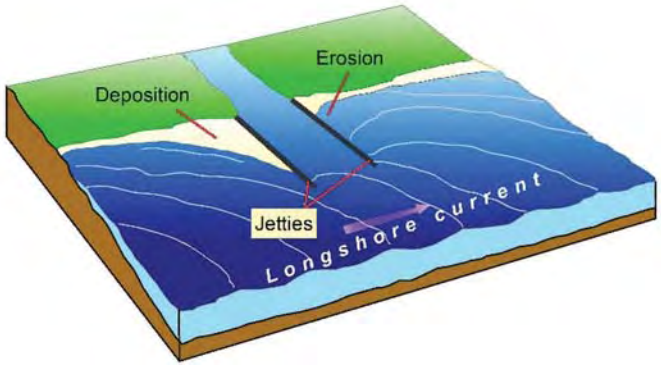


FIGURE 20.20 Direct channel flow at the coastline. Jetties direct river flow into a large lake without blocking flow as it reaches the coastline. The diagram shows the jetty at the left as it accumulates sediment from the current moving from left to right along the shoreline. The two jetties confine the river as it enters the lake and capture sediment. Note that the right-hand side is starved of sands because currents on the left are diverted into deeper waters by the jetty. From Hoa Tran. Used by permission.

FIGURE 20.21 Examples of jetties. (A) The Flathead River near Kalispell, Montana, as it flows through agricultural landscapes to enter Flathead Lake. As the river enters the lake, its velocity declines as its course levels to create the two arrow-shaped features at either side of the river. Note the plumes formed as the river enters the lake, and the other lake features that are formed by other streams and currents. Approximate location: 48°03′30.94″N, 114°08′04.22″W. Google Earth Pro image from Landsat/Copernicus. (B) Jetty at Corpus Christi, Texas, using visible colors, with river flow directed offshore and currents originating from the east and south. Approximate location: 27°49′20.24″N, 97°01′31.19″W. Google Earth Pro image from Data SIO (Scripps Institution of Oceanography), the U.S. National Oceanic and Atmospheric Administration (NOAA), U.S. Navy, National Geospatial-Intelligence Agency (NGA), and the General Bathymetric Chart of the Oceans (GEBCO). (C) The same jetty at Corpus Christi on a different date as a color-infrared image in a different orientation and showing a narrower section, providing a view of sediment suspended in currents and clearly showing the barriers extending into the Gulf of Mexico. From USGS.

20.5 IMPACT OF HURRICANE SANDY ON MANTOLOKING, NEW JERSEY

In this section, we examine the impact of Hurricane Sandy on Mantoloking, New Jersey, after it made landfall on October 29, 2012. Mantoloking is a borough in Ocean County occupying a portion of the Barnegat Peninsula, an elongated 20-mi (3.2-km) island situated at the New Jersey coastline. Its western shore forms a narrow bay facing the New Jersey coast; its eastern shore faces the Atlantic Ocean. The USGS Program had previously selected this region as a research site focused on examining responses of this site to marine effects. They had installed instruments in the Mantoloking barrier island to monitor local responses to tides, storms, currents, and related events. The site therefore was positioned to capture, in detail, the effects of the 2012 hurricane event in this region and to understand some of the dynamics of the storm and how the island responded to Hurricane Sandy.

FIGURE 20.22 USGS coastal classification of Texas Point, Jefferson County, Texas, showing an overhead view of the channel at Sabine Pass. Texas is the land mass on the left and Louisiana is on the right. At the upper left, we can see Port Arthur docks, refineries, storage, and transport facilities. Port Arthur is the largest U.S. oil refinery network and an important focal point for fuel transport. The colored stripes at the lower left of the image indicate the nature of the sand dunes, beach, structures, and dwellings that border the Texas coastline. Note the differences in the terrain on either side of the pass, an effect of the coastal transport of sediment south to north along the coastline. From USGS.

The USGS examined details of the pre- and post-Sandy landscape using airborne lidar (see Chapter 9) surveys to record coastal elevations before and after Hurricane Sandy and to document the large barrier island breach at Mantoloking (Figure 20.23). The left-hand diagram depicts lidar-derived pre-storm elevations from October 26, 2012. The center diagram presents post-storm elevations from lidar acquired on November 1, 2012. The right-hand diagram shows the subtraction of the two lidar-derived elevation layers, which illustrates elevation changes caused by the storm. Here, red tones indicate elevation losses, which exceed 5 m (16 ft) in some places. The center diagram records the effects of the overwash, including cuts through existing structures, perhaps following the effects of previous storms.

Mantoloking was hit especially hard by Hurricane Sandy, with the storm effects damaging about 90% of the properties and completely breaching the barrier island (Figures 20.24 and 20.25). More than 50 homes were demolished, accounting for almost 10% of the housing units in the borough (as of the 2010 Census).

The instrumentation of the barrier island and the remote sensing analysis demonstrated its value after Hurricane Sandy, as it documented the effects of breaches in the

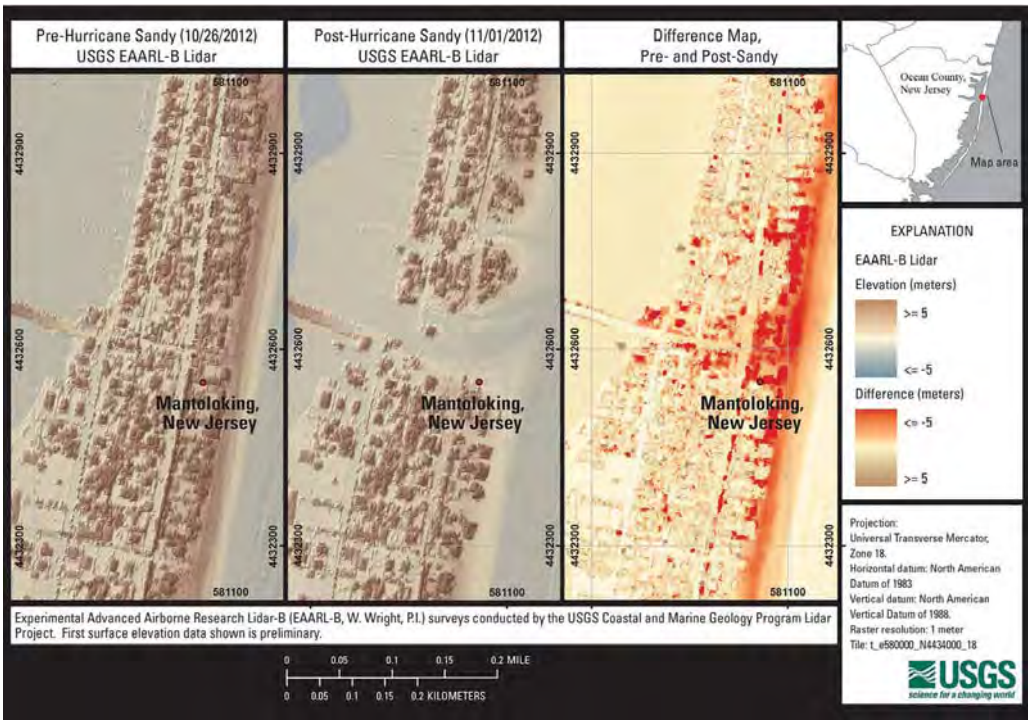


FIGURE 20.23 Pre– and post–Hurricane Sandy maps of Mantoloking, New Jersey. The left-hand diagram presents the pre-Sandy landscape, October 26, 2012. The center diagram displays the post-Sandy situation, November 1, 2012, depicting the effects of Sandy's storm surge waves in eroded beaches and dunes, and in forming a breach cut through the 250-m-wide island, destroying structures, roads, and depositing overwash over large areas. The right-hand diagram illustrates the difference in pre- and post-Sandy elevation, where red shows areas of most significant elevation loss. From USGS.

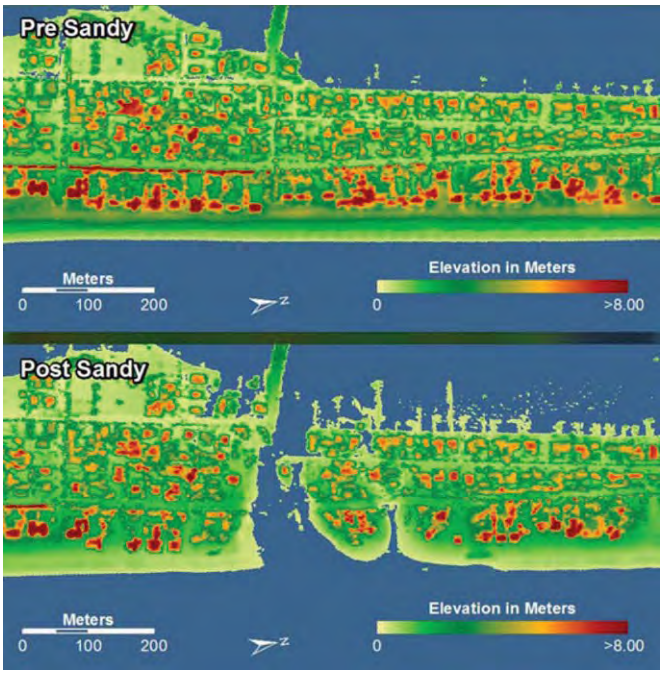


FIGURE 20.24 Elevation maps of building structural damage before and after Hurricane Sandy. Here we see a segment of the Mantoloking barrier island before (top) and after (bottom) Hurricane Sandy. Yellow and red areas identify structures before and after storm damage, marking damaged sites. Compare with **Figure 20.25** for an aerial photograph of the same region acquired at the same time. From USGS.

barrier islands on water quality and of tidal water levels in the estuary. Interestingly, analysis and models showed a decrease in residence time (i.e., the time required to completely flush the estuary) of estuarine water of 10 days because of the large breaches, which indicates an improvement in water quality. There was also little evidence of impact on the mainland shore from increased tidal waters due to the breach. This science has important implications for post-storm coastal resource management when considering whether and how to repair the breach.

20.6 LIDAR FOR COASTAL EROSION OF NORTH CAROLINA BEACHES

As discussed in Chapter 9, lidar contributes to several dimensions of hydrologic and hydrographic studies, especially for monitoring coastal features and bathymetry, and for examining subsurface topography. Here we examine White and Wang (2003), who interpreted sequential digital elevation data compiled from lidar imagery to examine changes in the form of North Carolina beaches over the interval from 1995 to 2005 due to erosion (Figure 20.26). These images show a small section of beach, with the open water of the Atlantic at the right, visible as the uniformly dark regions, and the white polygons rep-

FIGURE 20.25 Aerial view, storm damage, Mantoloking, New Jersey, October 2012. Barnegat Bay and the New Jersey mainland are visible in the central section of the image and in the background. The storm surge imaged here damaged about 90% of the properties, including the bridge, which was removed from service until it was repaired. Compare with **Figures 20.23 and 20.24**. From USGS.

resenting beach houses constructed immediately adjacent to the coastline edge. The gray areas that form the background for the left-hand regions represent the slightly elevated beaches and coastal topography. These sequential images show the varied impacts of beach erosion and beach nourishment, with the final images in the sequence showing the decline of ability to protect the structures. The economy, detail, and precision provided by lidar imagery offer advantages over previous methods, especially for dynamic landforms such as coastlines.

20.7 CHALLENGES IN COASTAL COMMUNITIES—EXAMPLE OF MIAMI BEACH

This discussion continues the topics presented in earlier sections, such as storm damage, coastal transport of sediments, and disturbance of beach features, at varied scales, using the example of Miami Beach, a coastal resort in Miami-Dade County, Florida. It is located on narrow barrier islands positioned between the Atlantic Ocean and Biscayne Bay, which separates Miami Beach from the mainland city of Miami. Miami Beach was formed in the early 1900s through the dredging of Biscayne Bay, which created an

improvised beach immediately offshore from mainland Miami. It eventually formed one of Miami's most popular tourist destinations, recognized for its beachfront location. It is characterized by dense anthropogenic structures immediately adjacent to the coastline, and so it is subject to many destructive events, including weather systems, tides, winds, and rising sea levels. In this section, we review a list of features that are often present in such coastline communities in the context of Miami Beach (Figure 20.27):

1. Use of groins
2. Shallow depths near the beach, allowing large storm waves to break near beaches and hotels
3. Narrow coastlines that allow waves to approach hotels and other building structures
4. Coastlines that allow lateral transport of sediment along the coastline

Remote sensing helps identify conditions, monitor change, and predict associated risks.

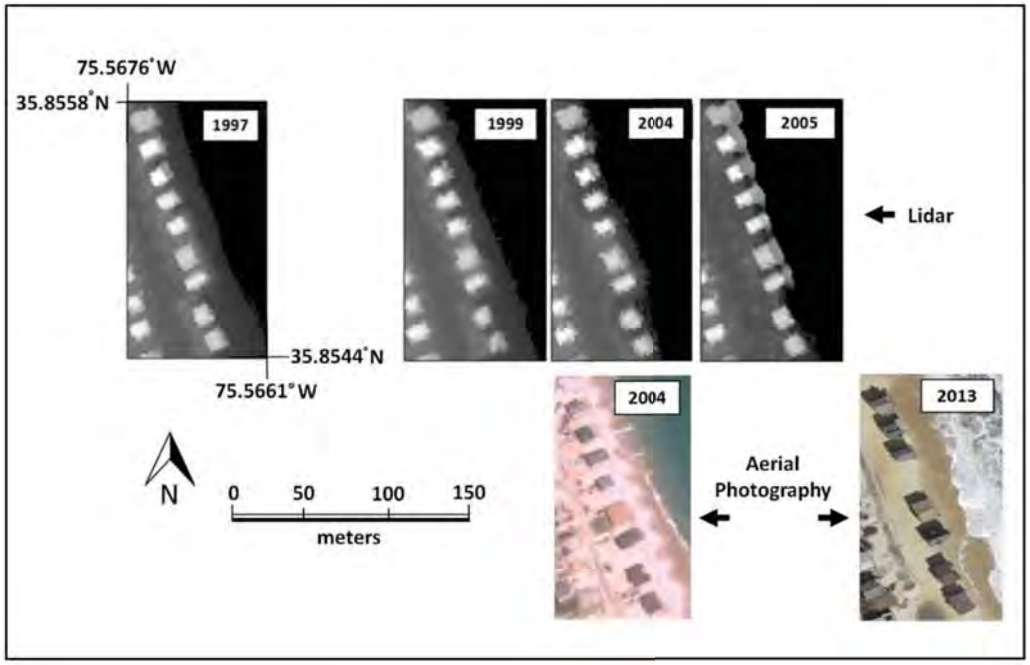


FIGURE 20.26 Lidar DEMs depicting beach erosion, Nags Head, North Carolina, 1997–2005. The bright rectangular shapes are beachfront homes as viewed from above; black areas to the right show open water of the Atlantic Ocean as observed by lidar (the top row). Lidar point data were collected in 1997, 1999, 2004, 2005, and later by NOAA Coastal Mapping programs hosted by the Coastal Service Center (CSC) of NOAA at Charleston, South Carolina. The aerial photographs show the same areas as do the lidar images, although specific dates and tidal cycles are different. Note that the 2013 aerial photo shows loss of structures at the water line. Data were derived online from the CSC web, downloaded and rasterized at East Carolina University. Ground reference data were collected between 2006 and 2008. See White and Wang (2003) and Allen and Wang (2010).

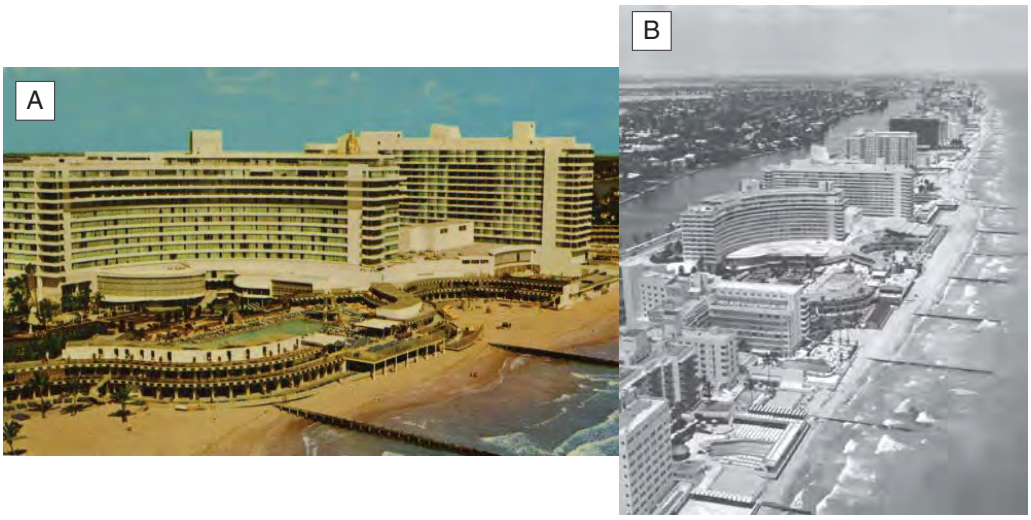


FIGURE 20.27 Two images showing Miami Beach over time. (A) Aerial image of Fontainbleau Hotel, constructed in 1954, which highlights the narrow beaches, shallow water, and groins. (B) The same area in 1970, with many more structures along the coastline. There has been subsequent modest expansion of beaches focused on selected hotels. Combined, these features depict characteristics regarded as unfavorable for large structures positioned near coastlines: (1) groins, now recognized as largely ineffective; (2) shallow waters close to coastlines, allowing storm waves to break close to shore; and (3) narrow beaches, allowing storms to reach coastlines near structures. From Florida Memory Library.

During the last decade, the rate of sea-level rise in southern Florida has tripled, including regular flooding of important avenues at high tide, with some regions building higher streets and seawalls and adding pumping stations. We have previously discussed how beaches are constantly changing environments, especially subject to seasonal fluctuations that influence the overall structures of shorelines and alter coastal features. Groins, found at Miami Beach, are another type of beach structure that extends perpendicularly from the shoreline, positioned across the surf zone. Their purpose is to capture and stabilize sands carried by wave action, as sketched in **Figure 20.28**. However, in some situations, groins can capture sands, creating cascades of sands that fill neighboring beach segments, fill adjacent groin areas, and distribute sand beyond the reach of groins. Thus, such groins can create difficulties by capturing sands that flow into neighboring groins, thereby contributing to downdrift beach losses.

Because Miami Beach is vulnerable to ocean risks, the city has implemented numerous programs to increase its resilience. They have elevated streets, created and elevated seawalls, and installed pump stations in lower areas to help control flooding, which is common at high tide. They are also attempting to restore beaches impacted by erosion and to reduce erosion in the long term (**Figure 20.29**). The restoration of beaches, known as beach renourishment, involves taking sand from other places ("borrow sites") and placing it on the damaged beaches. This allows for the installation of dune systems and a widening of the beach to create a larger buffer between hotels and other buildings and the coastline. The city of Miami is also attempting to preserve the coral reefs in the area, which provide additional fortification and protection from erosion.

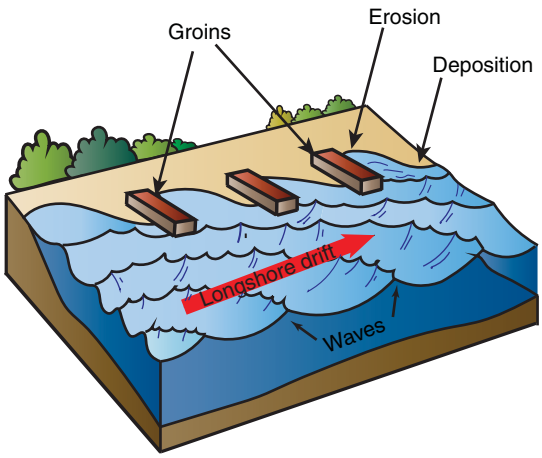


FIGURE 20.28 Groins positioned at coastlines to control longshore transport of sands. In concept, groins are intended to equalize the distribution of beach sands, although they may fall short of that objective. Here, the diagram shows erosion at the upstream sides of groins, deposition at the downstream sides. Sometimes, aggressive transport of sand can spill sand to cover neighboring groins. Image by Susmita Sen.

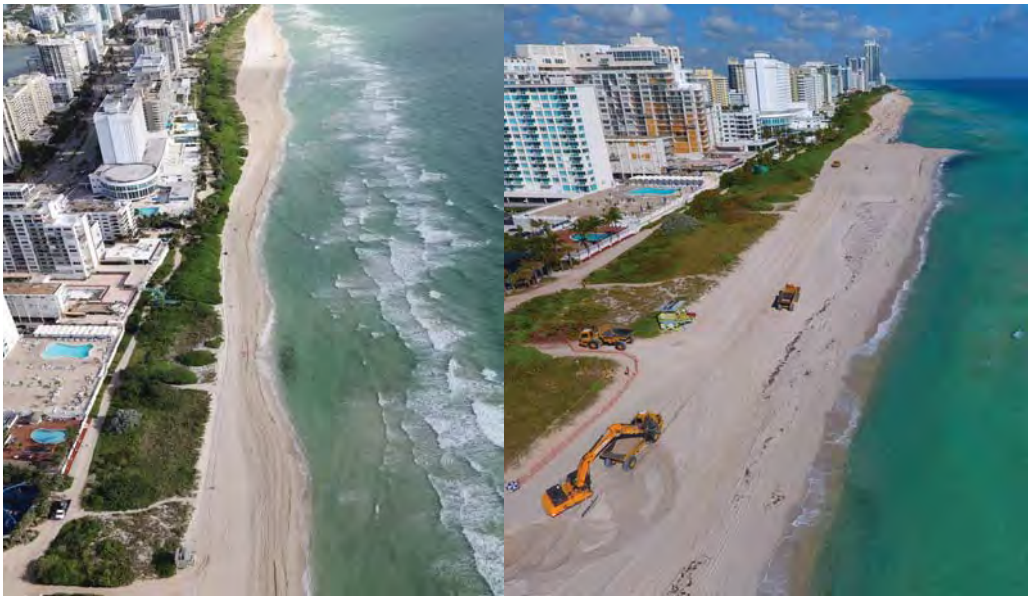


FIGURE 20.29 A beach renourishment project near Miami Beach. Left: Coastal features prior to the 54th Street renourishment project, later engineered for renewal and expansion. For this project, the U.S. Army Corps of Engineers transported beach-quality sands by truck from inland sites to renew and expand the beach, with costs shared by the federal government, the City of Miami, and Miami-Dade County. Right: A beach renourishment project in progress in the foreground. The completed coastline now provides a wider buffer between hotels and the coastline, forming a broader beach, and a wider extent of shallow waters. From U.S. Army Corps of Engineers.

In addition to shoreline erosion and sea level rise, the growth and development of Miami and Miami Beach have decreased the water quality of Biscayne Bay, primarily because of increased stormwater discharge and other pollutants. Remote sensing can be used to monitor these various dynamics both on land and in the water. It enables the assessment of sea surface temperature, turbidity, changes in photosynthetic material (algae), and other indicators of water quality. Figure 20.30 shows Miami Beach from the view of the 10-m Sentinel-2 bands (Chapter 7). The narrow beach and renourished area are visible in the natural-color composite (left) and can be monitored over time with multiple images. A version of the normalized difference turbidity index (NDTI) (i.e., [red – green]/[red + green]) calculated from the 10-m green and red Sentinel-2 band highlights the turbidity in Biscayne Bay and along the coast of the Atlantic (right). The influence of the two jetties connecting Biscayne Bay to the Atlantic Ocean can also be seen with respect to the turbidity and deposition on either side of the jetty.

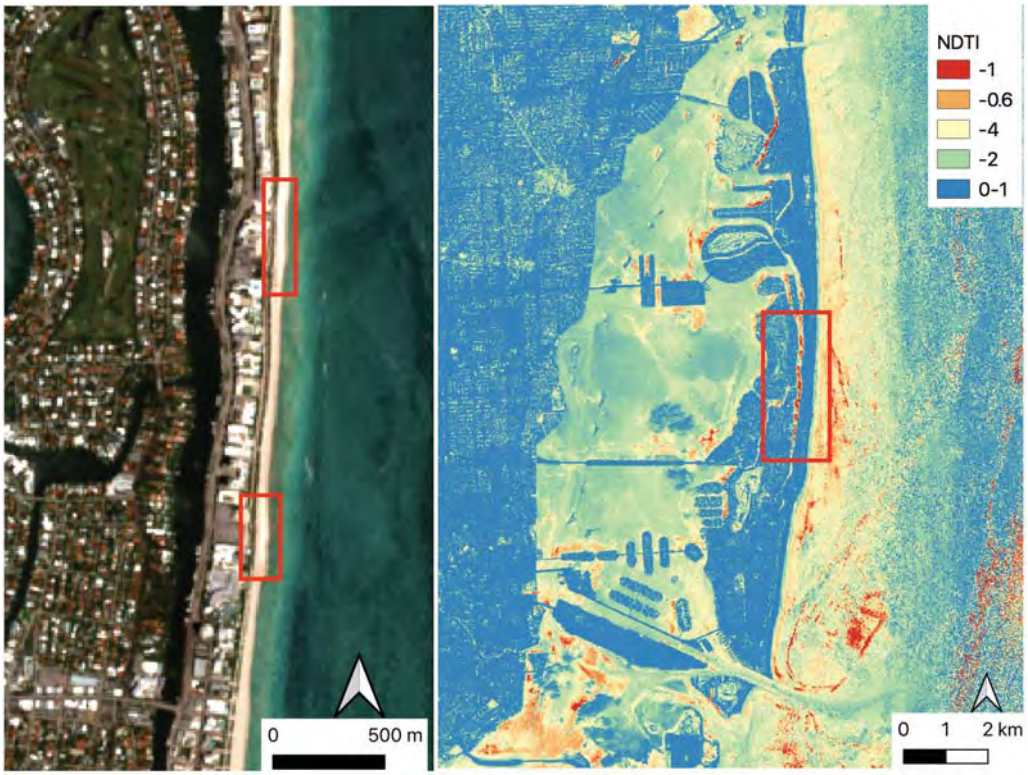


FIGURE 20.30 Miami Beach as seen from Sentinel-2 on May 5, 2020. Left: Natural-color composite (Sentinel-2 bands 4, 3, 2) with two examples of beach renourishment highlighted in red boxes. The top box shows the 54th Street renourishment area described in **Figure 20.29**. Right: A version of the normalized difference turbidity index (red – green)/(red + green) calculated from Sentinel-2 bands 4 and 3 showing turbidity in Biscayne Bay and along the Atlantic coast. These types of indices can be used to assess changes in water quality over time. The extent of the image in the left is shown in the red box on the right.

20.8 SUMMARY

Hydrologic studies cover a broad range of subject matter, from the movement of currents in bodies of open water to evaporation of moisture from a soil surface. Such topics can be very challenging, especially if the goal is to examine changes in hydrologic variables as they occur over time and space, because the usual methods of surface observation gather data at isolated points or at specific times. Two of the great advantages of remote sensing in this context are the *synoptic view* of the aerial perspective and the *opportunity to examine dynamic patterns at frequent intervals*.

Yet, remote sensing encounters difficulties when applied to hydrologic studies. Many of the standard sensors and analytical techniques have been developed for the study of land areas and are not easily applied to the special problems of understanding water bodies. Analyses often depend on detection of rather subtle differences in color, which are easily lost by the effects of atmospheric degradation of the remotely sensed energy, subsurface currents, and varying shadows/Sun angles. Currents and other dynamic features may change rapidly, requiring frequent observation to record the characteristics of significance to the analyst. Important hydrologic variables, such as groundwater, are not as a rule directly visible, and others, such as evapotranspiration, are not at all visible but must be estimated through other quantities that may themselves be difficult to observe. These problems and others mean that further development of hydrologic remote sensing is likely to be one of the most challenging research areas in remote sensing.

REVIEW QUESTIONS

1. List qualities of water bodies that present difficulties for those who study them using only surface observations collected from a ship. Identify those difficulties that are at least partially alleviated by use of some form of remote sensing.
2. It is probably best to compile bathymetric information using directly observed surface data because of their significance for navigation. Yet there are some special situations in which use of remotely sensed data may be especially advantageous; can you identify at least two such circumstances?
3. Can you think of important applications of the accurate delineation of edges of water bodies (i.e., simple separation of land vs. open water)? Be sure to consider observations over time, as well as use of images from a single date.
4. Contrast SPOT and MODIS (Chapter 7) with respect to their usefulness for oceanographic studies.
5. Write a short description of a design for a multispectral remote sensing system tailored specifically for collecting hydrologic information and accurate location of the edges of water bodies. Disregard all other applications. Suggest some of the factors that might be considered in choosing the optimum time of day for using such a system.
6. Outline the dilemma faced by scientists who wish to use preprocessing in the application of SPOT, Landsat, or Sentinel-2 data for bathymetric information. Describe the reasons a scientist would very much like to use preprocessing in some instances, as well as the counterbalancing reasons he or she would prefer to avoid preprocessing.

7. Outline some of the ways that the methods of image classification might be useful for hydrologic studies. Also outline problems and difficulties that limit the usefulness of these methods for studies of water bodies.

REFERENCES

Alföldi, T. T. 1982. Remote Sensing for Water Quality Monitoring. Chapter 27 in *Remote Sensing for Resource Management* (C. J. Johannsen and J. L. Sanders, eds.). Ankeny, IA: Soil Conservation Society of America, pp. 317–328.

Alföldi, T. T., and J. C. Munday. 1977. Progress toward a Landsat Water Quality Monitoring System. In *Proceedings, Fourth Canadian Symposium on Remote Sensing*. Montreal, Quebec, Canada: Canadian Remote Sensing Society of the Canadian Aeronautics and Space Institute, pp. 325–340.

Alizadeh, M. J., M. R. Kavianpour, M. Danesh, J. Adolf, S. Shamshirband, & K.-W. Chau. 2018. Effect of River Flow on the Quality of Estuarine and Coastal Waters using Machine Learning Models. *Engineering Applications of Computational Fluid Mechanics*, Vol. 12, No. 1, pp. 810–823.

Allen, T. R., and Y. Wang. 2010. Selected Scientific Analyses and Practical Applications of Remote Sensing: Examples from the Coast. Chapter 24 in *Manual of Geospatial Sciences and Technology* (J. D. Bossler, ed.). Boca Raton, FL: CRC Press, pp. 467–485.

Bhargava, D. S., and D. W. Mariam. 1991. Effects of Suspended Particle Sizes and Concentrations on Reflectance Measurements. *Photogrammetric and Remote Sensing*, Vol. 57, pp. 519–529.

Bianchi, T. S. 2013. Estuaries: Where the River Meets the Sea. *Nature Education Knowledge*, Vol. 4, No. 4, p. 12.

Borneman, E. 2014. How Rivers Change the Landscape. *Geography Realm*. Available at *www.geographyrealm.com/rivers-change-landscape*.

Curran, P. J., and E. M. M. Novo. 1988. The Effects of Suspended Particle Size and Concentration on Reflectance Measurements: A Review. *Journal of Coastal Research*, Vol. 4, pp. 351–368.

Fishes, L. T., F. Scarpace, and R. Thomson. 1979. Multidate Landsat Lake Quality Monitoring Program. *Photogrammetric Engineering and Remote Sensing*, Vol. 45, pp. 623–633.

Frazier, P. S., and K. L. Page. 2000. Water Body Detection and Delineation with Landsat TM Data. *Photogrammetric Engineering and Remote Sensing*, Vol. 66, pp. 1461–1465.

Gao, J. 2009. Bathymetric Mapping by Means of Remote Sensing: Methods, Accuracy, and Limitations. *Progress in Physical Geography*, Vol. 33, pp. 103–116.

Hallada, W. A. 1984. Mapping Bathymetry with Landsat 4 Thematic Mapper: Preliminary Findings. In *Proceedings, Ninth Canadian Symposium on Remote Sensing*. Halifax, Nova Scotia, Quebec, Canada: Canadian Aeronautics and Space Institute, pp. 629–643.

Han, L. 1997. Spectral Reflectance with Varying Suspended Sediment Concentrations in Clear and Algae-Laden Waters. *Photogrammetric Engineering and Remote Sensing*, Vol. 63, pp. 701–705.

Han, L., and D. C. Rundquest. 1997. Comparison of NIR/RED Ratio and First Derivative of Reflectance in Estimating Algal-Chlorophyll Concentration: A Case Study in a Turbid Reservoir. *Remote Sensing of Environment*, Vol. 62, pp. 253–261.

Han, L., and D. C. Rundquest. 2003. The Spectral Response of *Ceratophyllum Demersum* at Varying Depths in an Experimental Tank. *International Journal of Remote Sensing*, Vol. 24, pp. 859–864.

Holden, H., and E. Ledrow. 1999. Hyperspectral Identification of Coral Reef Features. *International Journal of Remote Sensing*, Vol. 20, pp. 2545–2563.

Houma, F., R. Belkessa, and N. Bachari. 2006. Contribution of Multispectral Satellite Imagery to

the Bathymetric Analysis of Coastal Sea Bottom. *Revue des Energies Renouvelables,* Vol. 9, pp. 165–172.

Jackson, R. D. 1985. Evaluating Evapotranspiration at Local and Regional Scales. *Proceedings of the IEEE,* Vol. 73, pp. 1086–1096.

Jackson, T. J., R. M. Ragan, and W. N. Fitch. 1977. Test of Landsat-Based Urban Hydrologic Modeling. *Journal of Water Resources Planning and Management Division, American Society of Civil Engineers,* Vol. 103, No. WRI, pp. 141–158.

Jagalingam, P., B. J. Akshaya, and A. V. Hegde. 2015. Bathymetric Mapping Using Landsat 8 Satellite Imagery. *Science Direct. Procedia Engineering,* Vol. 16, pp. 560–566.

Jawak, S., K. Kulkarni, and A. Luis. 2015a. A Review on Extraction of Lakes from Remotely Sensed Optical Satellite Data with a Special Focus on Cryospheric Lakes. *Advances in Remote Sensing*, Vol. 4, No. 3, pp. 196–213.

Jawak, S., S. Vadlamani, and A. Luis. 2015b. A Synoptic Review on Deriving Bathymetry Information Using Remote Sensing Technologies: Models, Methods and Comparisons. *Advances in Remote Sensing*, Vol. 4, No. 2, pp. 147–162.

Jie, W., D. Civco, and W. C. Kennard. 1992, May. Satellite Remote Bathymetry: A New Mechanism for Modeling. *Photogrammetric Engineering and Remote Sensing,* Vol. 58, No. 5, pp. 545–549.

Jupp, D. L. B., K. K. Mayo, D. A. Kuchler, D. Van R. Classen, R. A. Kenchinton, and P. R. Guerin. 1985. Remote Sensing for Planning and Managing the Great Barrier Reef of Australia. *Photogrammetria,* Vol. 40, pp. 21–42.

Khorram, S. 1980. Water Quality Mapping from Landsat Digital Data. *International Journal of Remote Sensing,* Vol. 2, pp. 143–153.

Klemas, V., R. Sicna, W. Treasure, and M. Otley. 1973. Applicability of ERTS-1 Imagery to the Study of Suspended Sediment and Aquatic Forms. In *Symposium on Significant Results Obtained from Earth Resources Technology Satellite-1.* Greenbelt, MD: Goddard Space Flight Center, pp. 1275–1290.

Lee, Z. P., K. L. Carder, and R. A. Arnone. 2002. Deriving Inherent Optical Properties from Water Color: A Multiband Quasi-Analytical Algorithm for Optically Deep Waters. *Applied Optics,* Vol. 41, pp. 5755–5772.

Liedtke, T. A., A. Roberts, and J. Lutenauer. 1995. Practical Remote Sensing of Suspended Sediment Concentration. *Photogrammetric Engineering and Remote Sensing,* Vol. 61, pp. 167–175.

Lillesand, T. M., E. L. Johnson, R. L. Deuell, O. M. Lindstrom, and D. E. Miesner. 1983. Use of Landsat Data to Predict the Trophic State of Minnesota Lakes. *Photogrammetric Engineering and Remote Sensing,* Vol. 49, pp. 219–229.

Lyzenga, D. R. 1979. Shallow-Water Reflectance Modeling with Applications to Remote Sensing of the Ocean Floor. In *Proceedings, Thirteenth International Symposium on Remote Sensing of Environment.* Ann Arbor: Environmental Research Institute of Michigan, pp. 583–602.

Lyzenga, D. R., N. M. Malinas, and F. J. Tarris. 2006. Multispectral Bathymetry Using a Single Physically Based Algorithm. *IEEE Transactions on Geoscience and Remote Sensing,* Vol. 44, pp. 2251–2259.

Lyzenga, D. R., R. A. Shuchman, and R. A. Arnone. 1979. Evaluation of an Algorithm for Mapping Bottom Features under a Variable Depth of Water. In *Proceedings, Thirteenth International Symposium on Remote Sensing of Environment.* Ann Arbor: Environmental Research Institute of Michigan, pp. 1767–1780.

Markham, B. L., and J. L. Barker. 1986. Landsat MSS and TM Post-Calibration Dynamic Ranges, Exoatmospheric Reflectance, and At-Satellite Temperatures. *Landsat Technical Notes,* No. 1, pp. 3–8.

McKim, H. L., C. J. Merry, and R. W. Layman. 1984. Water Quality Monitoring Using an Airborne Spectroradiometer. *Photogrammetric Engineering and Remote Sensing,* Vol. 50, pp. 353–360.

Mishra, D., S. Narumalani, M. Lawson, and D. Rundquist. 2004. Bathymetric Mapping Using IKONOS Multispectral Data. *GIScience and Remote Sensing,* Vol. 41, pp. 301–321.

Moore, G. K. 1978. Satellite Surveillance of Physical Water Quality Characteristics. In *Proceedings of the Twelfth International Symposium on Remote Sensing of Environment*. Ann Arbor: Environmental Research Institute of Michigan, pp. 445–462.

Munday, J. C., and T. T Alföldi. 1975. Chromaticity Changes from Isoluminous Techniques Used to Enhance Multispectral Remote Sensing Data. *Remote Sensing of Environment,* Vol. 4, pp. 221–236.

Munday, J. C., T. T. Alföldi, and C. L. Amos. 1979. Bay of Fundy Verification of a System for Multidate Landsat Measurement of Suspended Sediment. In *Satellite Hydrology* (M. Deutsch, D. R. Wiesner, and A. Rango, eds.). Minneapolis, MN: American Water Resource Association, pp. 622–640.

Nace, R. L. 1967. *Are We Running Out of Water?* (U.S. Geological Circular 586). Reston, VA: U.S. Geological Survey, 7 pp.

Nelson, R. 1985. Reducing Landsat MSS Scene Variability. *Photogrammetric Engineering and Remote Sensing,* Vol. 51, pp. 583–593.

Philipson, W. R., and W. R. Hafker. 1981. Manual versus Digital Landsat Analysis for Delineating River Flooding. *Photogrammetric Engineering and Remote Sensing,* Vol. 47, pp. 1351–1356.

Pilkey, O. H., and J. A. G. Cooper. 2014. *The Last Beach.* Durham, NC: Duke University Press. 264 pp.

Pilkey, O. H., L. Pilkey-Jarvis, and K. C. Pilkey. 2016. *Retreat from a Rising Sea: Hard Choices in an Age of Climate Change*. New York: Columbia University Press, 214 pp.

Polcyn, F. C., and D. R. Lyzenga. 1979. Landsat Bathymetric Mapping by Multispectral Processing. In *Proceedings, Thirteenth International Symposium on Remote Sensing of Environment*. Ann Arbor: Environmental Research Institute of Michigan, pp. 1269–1276.

Ritchie, J. C., P. V. Zimba, and J. H. Everitt. 2003. Remote Sensing Techniques to Assess Water Quality. *Photogrammetric Engineering and Remote Sensing,* Vol. 69, pp. 695–704.

Satzman, B. (ed.). 1985. *Satellite Oceanic Remote Sensing: Advances in Geophysics*. New York: Academic Press, 511 pp.

Schmugge, T. J., W. P. Kustas, J. C. Ritchie, T. J. Jackson, and A. Rango. 2002. Remote Sensing in Hydrology. *Advances in Water Resources,* Vol. 25, pp. 1367–1385.

Schwab, D. J., G. A. Leshkevich, and G. C. Muhr. 1992. Satellite Measurements of Surface Water Temperatures in the Great Lakes. *Journal of Great Lakes Research,* Vol. 18, pp. 247–258.

Stanley, S. 2016, April 28. When Rivers and Tides Collide. *EOS*, Vol. 97.

Strandberg, C. 1966. Water Quality Analysis. *Photogrammetric Engineering,* Vol. 32, pp. 234–249.

Stumpf, R. P. 1992. Remote Sensing of Water Clarity and Suspended Sediment in Coastal Waters. In *Proceedings of the First Thematic Conference for Marine and Coastal Environments. Proceedings of the International Society of Optical Engineering,* Vol. 1930, pp. 1436–1473.

Tanis, F. J., and W. A. Hallada. 1984. Evaluation of Landsat Thematic Mapper Data for Shallow Water Bathymetry. In *Proceedings, Eighteenth International Symposium on Remote Sensing of Environment*. Ann Arbor: Environmental Research Institute of Michigan, pp. 629–643.

Tolk, B. L., L. Han, and D. C. Rundquist. 2000. The Impact of Bottom Brightness on Spectral Reflectance of Suspended Sediments. *International Journal of Remote Sensing,* Vol. 21, pp. 2259–2268.

Wang, C.-K., and W. D. Philpot. 2007. Using Airborne Bathymetric Lidar to Detect Bottom Type Variation in Shallow Waters. *Remote Sensing of Environment,* Vol. 106, pp. 123–135.

Westawy, R., M. S. N. Lane, and D. M. Hicks. 2001. Remote Sensing of Clear-Water, Sallow, Gravel-Bed Rivers Using Digital Photogrammetry. *Photogrammetric Engineering and Remote Sensing,* Vol. 67, pp. 1271–1282.

White, S. A., and Y. Wang. 2003. Utilizing DEMs Derived from LIDAR Data to Analyze Mor-

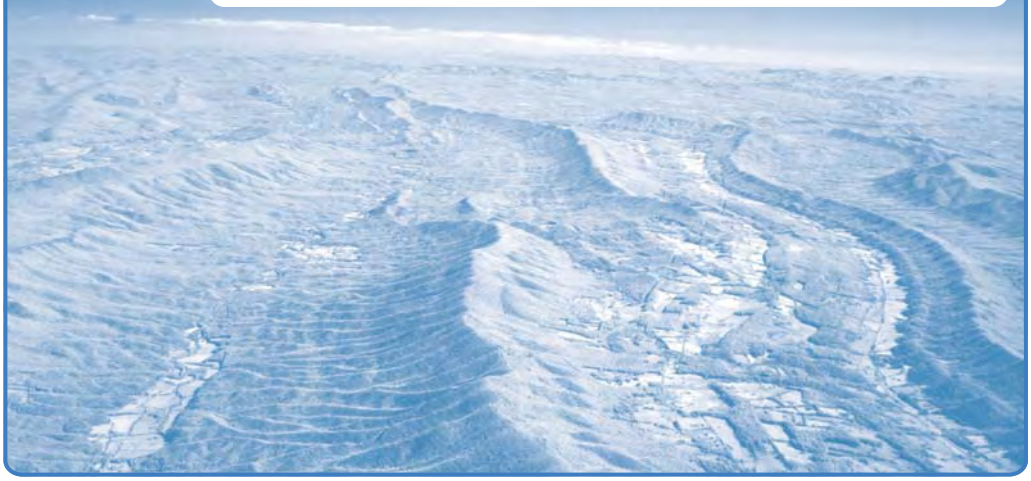
phologic Change in the North Carolina Coastline. *Remote Sensing of Environment,* Vol. 85, pp. 39–47.

Williams, D. D. 2005. *The Biology of Temporary Waters*. Oxford, UK: Oxford University Press, 337 pp.

Zhou, T., J. Wu, and S. Peng. 2012. Assessing the Effects of Landscape Pattern on River Water Quality at Multiple Scales: A Case of the Dongjiang River Watershed, China. *Ecological Indicators,* Vol. 23, pp. 166–175.

Zinnert, J. C., S. M. Via, B. P. Nettleton, P. A. Tuley, L. J. Moore, and J. A. Stallins. 2019. Connectivity in Coastal Systems: Barrier Island Vegetation Influences Upland Migration in a Changing Climate. *Global Change Biology,* Vol. 25, No. 7, pp. 2419–2430.

21 Land Use and Land Cover



MAJOR TOPICS TO UNDERSTAND

- Aerial Imagery for Land-Use Information
- Land-Use Classification
- Visual Interpretation of Land Use and Land Cover
- Land-Use Change by Visual Interpretation
- Historical Land Cover Interpretation for Environmental Analysis
- Other Land-Use Classification Systems
- Land Cover Mapping by Image Classification
- Broad-Scale Land Cover Studies
- Sources of Compiled Land-Use Data

21.1 INTRODUCTION

Land use describes use of the land surface by humans. Normally, use of land is defined in an economic context, so we think of land as it is used for agricultural, residential, commercial, and other purposes. However, strictly speaking, we can seldom see the use of the

land, except on the closest inspection, so we also consider *land cover*—the visible features of the Earth's surface, such as vegetative cover, structures, highways, forests, ponds, and such, often modified by humans (**Figure 21.1**). So, as a practical matter, we consider land use and land cover together, but we also recognize distinctions between the two.

Modern society depends on accurate land-use data for both scientific and administrative purposes. They form essential components of local and regional economic planning to ensure that various activities are positioned on the landscape in a rational manner. For example, accurate knowledge of land-use patterns permits planning to avoid placing residential housing adjacent to heavy industry or in floodplains. Accurate land-use information can ensure that residential neighborhoods are logically placed with respect to commercial centers and access to transportation services. In another context, land use is an important component of climatic and hydrologic modeling to estimate the runoff of rainfall from varied surfaces into stream systems. Accurate land-use data are important for transportation planning, so traffic engineers can estimate the flow of vehicles from one region to another and design highways with appropriate capacities.

Land-use patterns also reflect the character of a society's intimate interaction with its physical environment. Although land-use patterns in our own society seem self-evident and natural, other societies often organize themselves in contrasting patterns. This fact becomes obvious when it is possible to observe differing economic and social systems occupying similar environments.

The practice of land use and analysis has a long history involving a wide variety of places and applications, which Philipson covers in *Manual of Photographic Interpretation* (1997). Philipson's Chapter 9, "Land Use and Land Cover Inventory," is useful for understanding the history of land-use and land cover analysis, much of which was prepared before digital imagery was widely available.

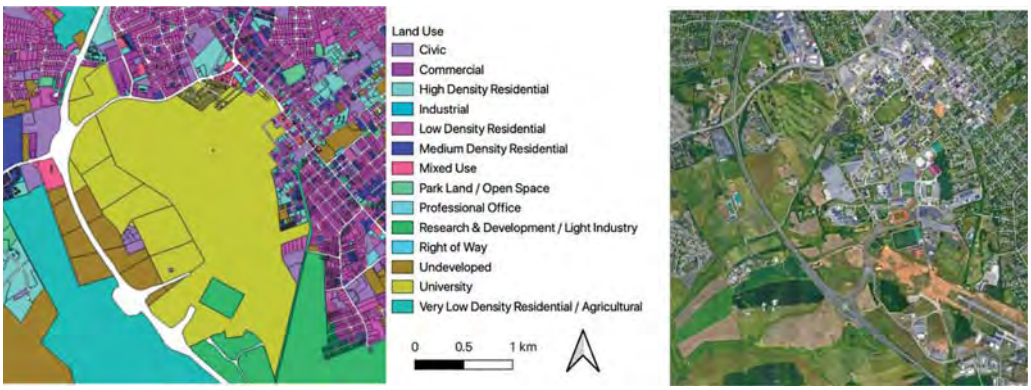


FIGURE 21.1 Land use versus land cover for part of Blacksburg, Virginia. Left: The 2019 land-use categories created by the town of Blacksburg (*https://sites.google.com/vt.edu/townofblacksburg-gisdata*). *Land use* refers to the principal economic enterprises that characterize an area of land, such as Commercial, Residential, Park, and University. Note the large area of "University" in the middle of the map. From Virginia Tech Library. Used by permission. Right: Aerial image of the same area, approximate location, 37°25′15.54″N, 80°25′25.80″W. *Land cover* indicates physical features that occupy the surface of the Earth. In the area designated as "University" on the left, land cover features such as grass, bare soil, trees, buildings, and other impervious surfaces are visible on the right. Google Earth Pro image from Landsat/Copernicus.

Here, we should note briefly that aerial imagery for land-use and land cover analysis had its origins early in civilian uses of aerial surveys but did not develop into systematic applications until the 1970s or so. For many years, land-use inventory was largely based on automobile surveys to acquire rough inventories of urban landscapes and on access to planning documents, with aerial surveys applicable when costs were feasible. An important change occurred when the U.S. Geological Survey (USGS) hired James R. Anderson to develop systematic classification of land use and land cover as observed initially by aerial survey and later from Landsat imagery (Figure 21.2).

21.2 AERIAL IMAGERY FOR LAND-USE INFORMATION

Although some forms of land-use data (such as local zoning or tax data) are acquired through direct observation by officials who visit each site in person, the vast majority of land-use and land cover data is acquired through interpretation of maps, diagrams, aerial photography, and similar imagery. Aerial photography provides an overhead view, a spatial perspective, and comprehensive detail that permits accurate, systematic, and effective study of land-use patterns. Early proponents of the use of aerial photography envisioned that it would provide a rich source of information regarding the structure of cities and landscapes. By the 1930s, the staff members of the Tennessee Valley Authority (TVA) program were among the first to pioneer methods for using aerial photography for land-use surveys. Later, the USGS developed classification systems specifically tailored for use

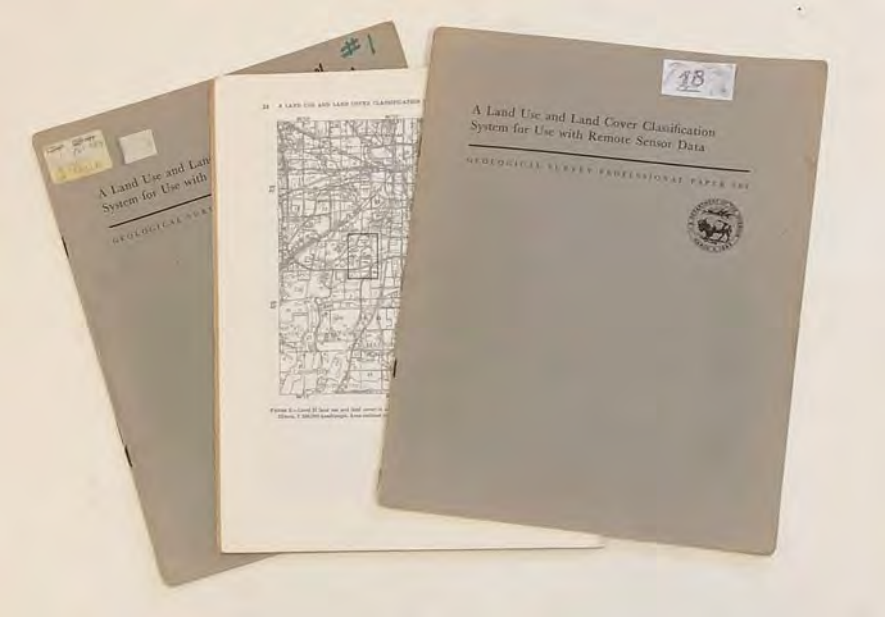


FIGURE 21.2 James R. Anderson's land use and land cover classification system. His contributions led to more effective analyses of aerial photography and interpretations of satellite imagery. The professional paper published by the USGS has been widely used throughout the United States and overseas.

with aerial photography. These techniques form the foundation for today's applications of satellite imagery for land-use surveys.

Remotely sensed images lend themselves to accurate land cover and land-use mapping in part because land cover information can be interpreted more or less directly from aerial images and because aerial observation allows us to view the backgrounds of neighboring features.

Land-use maps are routinely prepared on a wide variety of scales, typically ranging from 1:12,500 and larger to 1:250,000 and smaller. At one end of this spectrum (larger scale maps), remotely sensed imagery may itself contribute relatively little information to the survey; its main role may be to form a highly detailed base for recording data gathered by other means. At such large scales, the land-use map may serve as a kind of reference, having little cartographic generalization. Such products are often used at the lowest levels of local governments, perhaps mainly in urban areas, that have requirements for detailed information and financial resources to acquire it (**Figure 21.3**).

As the scale of the survey becomes broader (i.e., as map scale becomes smaller), the contribution of the image to the informational content of the map becomes greater, although even at the smallest scale there must always be some contribution from collateral information. Differing scales and levels of detail serve different purposes and different users. For the regional planner, losses of resolution and detail at smaller scales may actually provide an advantage in the sense that analysts may prefer integration and simplification of the information that must be examined. For medium- to small-scale land-use surveys, the product is often a thematic map that depicts predominant land cover in

FIGURE 21.3 A planning official using an aerial photograph to discuss land-use policy relating to planning land use near Boston's Logan Airport, 1973. Although modern displays have replaced such images, they are especially valuable for informing the public of plans for broad-scale renewal projects. From National Archives and Records Administration, no. ARC548443.

relatively homogeneous areas, delineated subject to limitations of scale, resolution, and other constraints accepted by users of thematic maps.

21.3 LAND-USE CLASSIFICATION

Preparation of a land-use/land cover (LULC) map requires that mapped areas be subdivided into discrete parcels, each labeled with distinct, mutually exclusive nominal labels. That is, each parcel must be identified using a single, distinct label in a classification system. "Nominal" simply means that the labels are names rather than values. So, for example, when we use a symbol such as "22" to designate a category, we do not mean that it has twice the value of "11," but simply that the separate designations convey different qualitative meanings, such as, for example, "urban" and "agricultural" land uses. Usually, we also use colors, or colors and symbols, to designate land-use classes on maps.

Perhaps the most widely used classification system for land-use and land cover data derived from aerial photography is the USGS's Land Use and Land Cover classification system, developed during the 1970s by Anderson et al. (1976) and also referred to as the Anderson system (Table 21.1).

Such classification systems have many advantages over previous systems. Whereas previous systems did not consider the unique advantages of aerial imagery as a source of land-use and land cover data, the USGS system was specifically prepared for use with aerial photography and related imagery. It has a hierarchical design that lends itself to use with images of varied scales and resolutions. Level I categories (indicated with a single digit, 1–9), for example, are designed for broad-scale, coarse-resolution imagery (e.g., broad-scale satellite imagery or high-altitude aerial photography).

Although the USGS system specifies Level I and II categories, more detailed classes at Level III and below must be defined by the analyst to meet the specific requirements of a particular study and a specific region. When an interpreter defines Level III categories, the Level I and II categories should be used as a framework for the more detailed Level III classes.

21.4 VISUAL INTERPRETATION OF LAND USE AND LAND COVER

Image analysts apply the elements of image interpretation to delineate separate land cover classes, applied in an organized and systematic manner. Visual land-use interpretation proceeds by marking boundaries between categories as they occur on the imagery. Often, interpreters mark transparent overlays that register to the image, or digitize directly as the image is displayed on a computer screen. As each land-use parcel is outlined, its identity is marked with a symbol (often one to three numerals) matching to the classification system (Figure 21.4).

A land-use map is then prepared by delineating regions of consistent land use and assigning them to appropriate classes in land-use classification. An example can show an application of land-use mapping to an aerial photograph, using the principles outlined in Chapter 6 (Figure 21.5). Interpretation steps are as follows: (1) the original aerial photograph is examined (Figure 21.5A); (2) the land-use boundaries, as defined by a photointerpreter, are drawn to separate land-use parcels at Anderson Level II (Figure 21.5B); (3) the separate parcels are then assigned to classes and symbolized with colors

TABLE 21.1 USGS Land Use and Land Cover Classification

1	Urban or Built-Up Land
	11 Residential
	12 Commercial Services
	13 Industrial
	14 Transportation, Communications
	15 Industrial and Commercial
	16 Mixed Urban or Built-Up Land
	17 Other Urban or Built-Up Land
2	Agricultural Land
	21 Cropland and Pasture
	22 Orchards, Groves, Vineyards, Nurseries
	23 Confined Feeding Operations
	24 Other Agricultural Land
3	Rangeland
	31 Herbaceous Rangeland
	32 Shrub and Brush Rangeland
	33 Mixed Rangeland
4	Forest Land
	41 Deciduous Forest Land
	42 Evergreen Forest Land
	43 Mixed Forest Land
5	Water
	51 Streams and Canals
	52 Lakes
	53 Reservoirs
	54 Bays and Estuaries
6	Wetland
	61 Forested Wetlands
	62 Nonforested Wetlands
7	Barren Land
	71 Dry Salt Flats
	72 Beaches
	73 Sandy Areas Other than Beaches
	74 Bare Exposed Rock
	75 Strip Mines, Quarries, and Gravel Pits
	76 Transitional Areas
	77 Mixed Barren Land
8	Tundra
	81 Shrub and Brush Tundra
	82 Herbaceous Tundra
	83 Bare Ground
	84 Wet Tundra
	85 Mixed Tundra
9	Perennial Snow and Ice
	91 Perennial Snowfields
	92 Glaciers

Note: From Anderson et al. (1976, p. 28).

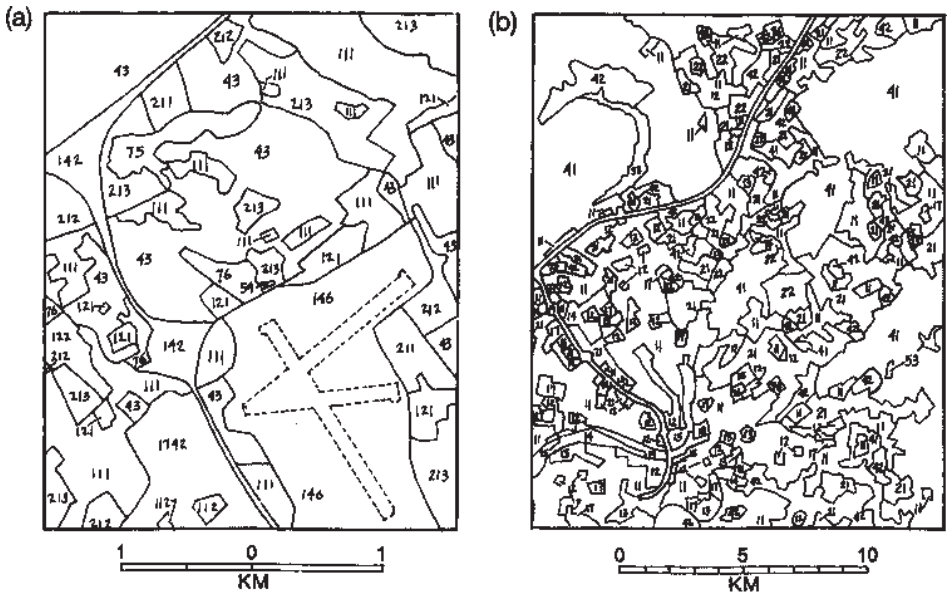


FIGURE 21.4 Land-use and land cover maps. (a) Large scale, fine detail; (b) small scale, coarse detail. Scale determines much of the information conveyed by maps, as indicated in these examples. Coarse detail generalizes, losing specifics. Fine detail loses spatial scope, requiring use of larger map surfaces.

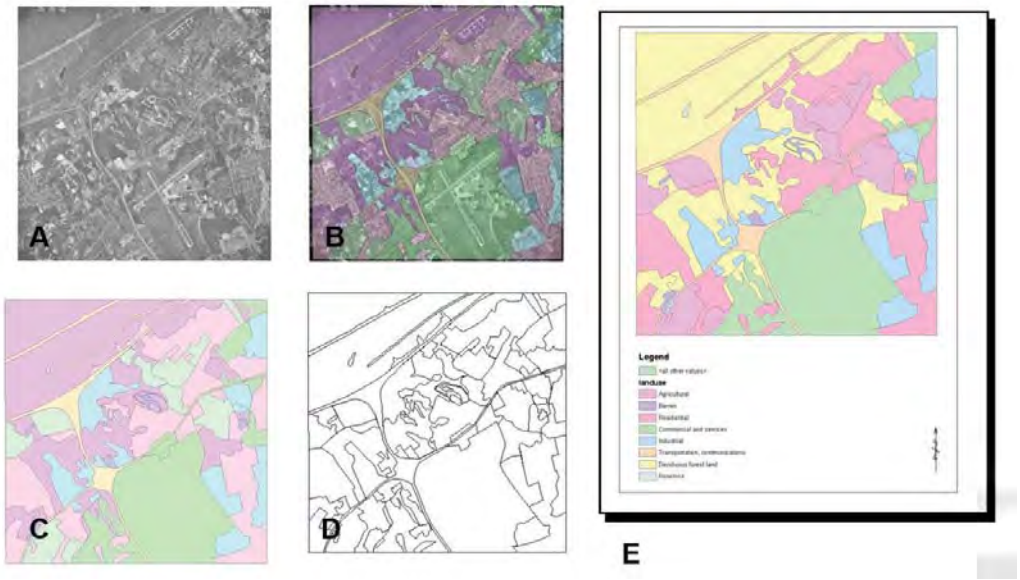


FIGURE 21.5 Representation of key steps in interpretation of land use from aerial photography. (A) Aerial photograph only; (B) delineation of boundaries; (C) symbolization of principal classes; (D) boundaries without the backdrop of the aerial photograph; (E) map with symbols and key. From Jessica Dorr Perkins. Used by permission.

or symbols (Figure 21.5C and 21.5D); and finally (4) a map is displayed with a legend (Figure 21.5E). LULC information must be presented to the user in a planimetrically correct form, so the imagery that is used for the interpretation must be corrected beforehand (e.g., as in the case of a digital orthophoto quadrangle [Chapter 4]), or the boundaries must later be rectified.

Identification of land cover parcels is based on the elements of image interpretation, as presented in Chapter 6. Although land use may sometimes be characterized by specific objects, usually the primary task is one of consistent delineation of uniform parcels that match the classification system. As an example, *cropped agricultural land* (as it occurs in midlatitude landscapes) is usually easily recognizable by the systematic division of fields and smooth, even textures. Tone varies with the crop and the growth stage (Figure 21.6A). *Pasture* (Anderson symbol 21) usually has more irregular boundaries and a mottled texture with medium tones and is often characterized by isolated trees or small groves both at edges and in the interiors of parcels (Figure 21.6B). *Deciduous forest* (Anderson symbol 41) is characterized by rough textures and medium dark tones that usually occur in relatively large parcels with irregular edges (i.e., there is a small patch in Figure 21.6B). Roads and clearings are common.

Land in transition (Anderson symbol 76) is usually recognizable by the bright tones characteristic of bare soil exposed during construction and irregular outlines. Sometimes the outlines of roadways, foundations for buildings, and partially completed structures are visible (Figure 21.7A, B). *Transportation* (Anderson symbol 14) is often recognizable by linear patterns that cut across predominant land-use patterns and by distinctive loops of interchanges. *Residential* (Anderson symbol 11) land uses, especially in suburban developments, are often distinctive because of the even placement of structures, the curving street patterns, and the even backgrounds of lawns and ornamental trees (Figure 21.7C). *Commercial services* (Anderson symbol 12) is characterized by bright tones (because of the predominance of pavement, parking lots, and rooftops), uneven textures (because of shadowing of buildings), and close proximity to transportation and other high-density land-use classes.

As a map is prepared, remember that some land-use maps have legal standing and that some may be intended to match data collected previously or by a neighboring juris-

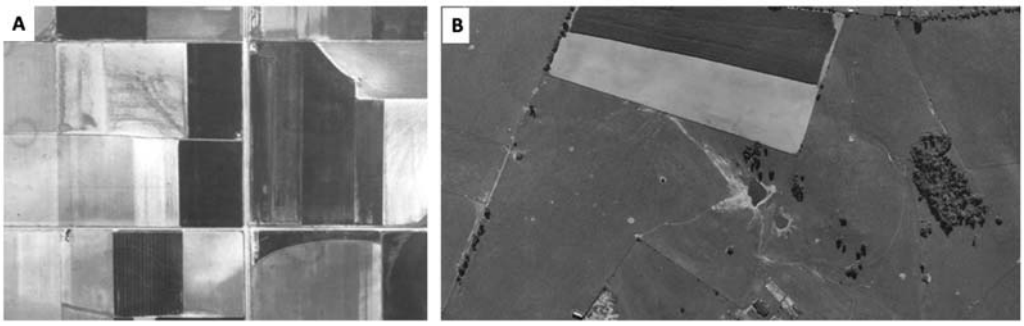


FIGURE 21.6 Agricultural land showing distinctive features that aid in visual interpretation. (A) Systematic fields with regular shapes and smooth, even textures; (B) two fields (top center) and open pasture. A small patch of trees can also be seen. From Virginia Department of Transportation. Used by permission.

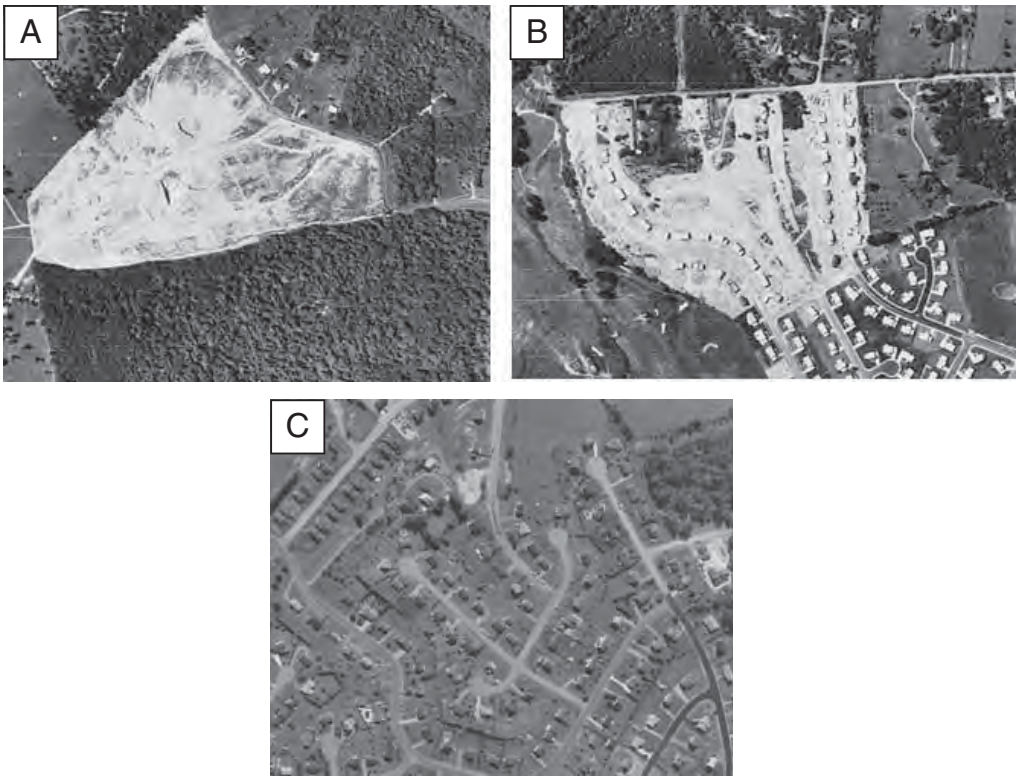


FIGURE 21.7 Visual interpretation of aerial imagery for land-use–land cover information. (A and B) Land in transition (Anderson symbol 76) is often characterized by bright tones, uneven texture, and irregular parcels; (C) residential land (Anderson symbol 11). Suburban residential land, such as that shown here, is often characterized by evenly spaced rooftops and systematic patterns of paved streets. From Virginia Department of Transportation. Used by permission.

diction. Therefore, the analyst should not make off-the-cuff adjustments solely to address local interpretation problems.

Interpreters must take care to define delineations that are consistent, clear, and legible. Each parcel is completely enclosed by a boundary and is labeled with a symbol keyed to the classification system. The interpreter delineates only those features that occupy areas at the scale of the final map, so linear features (such as streams and rivers, railways, and powerlines) or point data are usually not mapped.

Some parcels will encompass categories other than those named by the parcel label. Such inclusions are permissible but must be clearly described in an accompanying report and must be consistent throughout the map. When several interpreters work on the same project, consistency is especially important. Interpreters must coordinate their work with those assigned to neighboring areas, so that detail is uniform throughout the mapped area.

The entire area devoted to a specific use is delineated on the overlay. Thus, delineation of an airfield normally includes not only the runway but also hangars, passenger terminals, parking areas, access roads, and, in general, features enclosed by limits of the

perimeter fence (i.e., the outline of the parcel encompasses areas occupied by all of these features, even though they may not be shown individually on the map). In a similar manner, delineations of interstate highways include not only the two parallel roadways, but also a median strip and a fenced right of way.

The example of an aerial image of a quarry can also illustrate this practice (**Figure 21.8**). The quarry itself consists of several features, each of which has a distinctive appearance on the aerial photograph. Even though it would be possible to individually delineate the transportation and loading structures, the milling and administration buildings, the quarries, and the spoil, these features constitute a single land use, so all the features are outlined and labeled as a single parcel. Only at the very largest scales, possibly required

FIGURE 21.8 Delineation of an entire area devoted to a given use. This quarry consists of a complex of individual activities that together constitute a single enterprise, so the individual component need not be delineated separately. From Virginia Department of Transportation. Used by permission.

for planning and management of the facility itself, would the individual components be delineated as separate features.

The issue of *multiple use* occurs because we assign a single label to each parcel, even though we know that there may in fact be several uses. For example, a forested area may simultaneously serve as a source of timber, a recreational area for hunters and hikers, and a source of runoff that supplies water for an urban region.

The interpreter must select an appropriate *minimum size* for the smallest parcels to be represented on the final map. Areas smaller than this size will be omitted. Even though they are visible to the interpreter, they are too small to depict legibly.

Usually, the label of each category identifies the *predominant category* present within each parcel. Although at small scales there may be unavoidable inclusions of other categories, the interpreter should aspire to define categories that include consistent mixtures of such inclusions.

21.5 LAND-USE CHANGE BY VISUAL INTERPRETATION

Land-use patterns change over time in response to economic, social, and environmental forces. The practical significance of such changes is obvious. For planners and administrators, they reveal the areas that require the greatest attention if communities are to develop in a harmonious and orderly manner. From a conceptual perspective, study of land-use changes permits identification of long-term trends in time and space and the formation of policy in anticipation of the problems that accompany changes in land use (Anderson, 1977; Estes and Senger, 1972; Jensen and Toll, 1982).

Remote sensing and photointerpretation provide the primary vehicles for compiling *land-use change maps*—maps that represent changes in land use from one date to another. Such maps form important tools in planning land-use policy, in hydrologic management of watersheds, in transportation planning (to understand changes in the population patterns that generate traffic flow), and in environmental studies.

The concept behind land-use and land cover change is very simple: two maps representing the same region, prepared to depict land-use patterns at different dates, are compared, point by point, to summarize differences between the two dates (Figure 21.9). In practice, compiling land-use and land cover change requires mastery of several practical procedures that often reveal previously hidden difficulties.

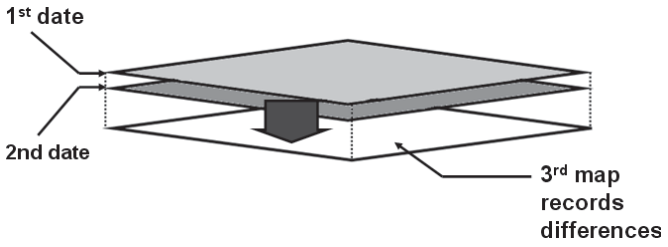


FIGURE 21.9 Schematic representation of compilation of land-use change using sequential aerial imagery. Information from two similar maps is aggregated, or extracted, to create a third map combining information from the two earlier maps.

First, the two maps must use the same classification system, or at a minimum, two compatible systems. Compatible classifications mean that the classes can be clearly matched to one another without omission or ambiguity. For example, if one map represents forested land as a single category, "forest," and a second uses three classes—"coniferous," "deciduous," and "mixed coniferous/deciduous," then the two maps are compatible in the sense that the three categories of the second map can be combined into a single forested class that is logically compatible with the "forest" class on the first map. As long as the classes are matched at that level of generalization, one can be confident that the comparisons would reveal changes in the extent of forested land.

However, if the second map employs classes such as "densely forested," "partially forested," and "sparsely forested," we could not reasonably match these classes to the "forest class" on the first map; the two sets of classes are not logically compatible. The meanings conveyed by these classes, as represented on the two maps, are not equivalent, so they cannot be compared to determine whether the differences represent true change or simply are differences in the way the maps were prepared.

Second, the two maps must be compatible with respect to scale, geometry, and level of detail. Can we match one point on one map with the corresponding point on the second and be sure that both points refer to the same place on the Earth's surface? If not, we cannot conduct a change analysis using the two maps because we cannot be sure that differences in the two maps reflect genuine differences in land use and are not simply differences in the projection, scale, or geometric properties of the two maps.

These two basic conditions, compatibility of the information portrayed and geometric compatibility, must be satisfied before change data can be compiled using any two maps or images. Often, a land-use map is prepared to meet the very specific objectives of the sponsoring organization. Therefore, it may be difficult to reuse that map in a land-use change study many years later. Objectives may have changed, or the level of detail, the classification system, and so forth, reflect the needs of a previous era. It is for this reason that land-use change studies for maps are commonly prepared at the time of the study, employing consistent materials, methods, and procedures, even if earlier studies already exist and are available.

21.6 HISTORICAL LAND COVER INTERPRETATION FOR ENVIRONMENTAL ANALYSIS

A special application of land-use change analysis has been developed to address the environmental hazards arising from land-use change in regions where hazardous industrial materials have been abandoned. Aerial imagery is especially valuable because it provides a spatial perspective and has the ability both to provide a detailed historical record and to observe outside the visible spectrum. These capabilities allow analysts to examine problems over time, to detect the existence of specific environmental problems, and to see them in their geographical context. Image analysis can provide a historical record of conditions and actions at specific sites to assess the nature and severity of conditions, often when other sources of information are not available.

Over the decades, in industrial regions of the United States, hazardous industrial wastes have been deposited in varied disposal sites, including ponds, lagoons, and landfills, that were not designed to safely store dangerous substances. Although many of these areas were once positioned in rural or sparsely populated regions, urban growth has

brought many of them close to residential neighborhoods. As a result, many such sites are now near populated regions and in some instances have been converted to other uses, including residential housing, schoolyards, and recreational areas. Deterioration of containers has released these materials to contaminate nearby soils and to enter groundwater systems, migrating beyond the original site. Often, the use of the site for waste disposal was not properly recorded or the nature of the wastes was never documented, so remediation is limited by ignorance of the specifics of the situation.

Because of the difficulty of establishing the historical pattern of use at such sites and the absence or inaccuracy of official records, aerial photography has formed one of the most valuable resources for investigating them. For many areas in the United States, the archives of aerial photography extend as far back as the 1930s. Thus, it may be possible to assemble photographs to form a sequence of snapshots that record the use of a site over time. Such images, when combined with field data, official records, and health surveys, can permit development of an understanding of the sequence of events and assessment of the risks to the environment and to nearby populations. In some instances, use of photography permits estimation of the amount of waste present and even specification of the class of materials by identifying the kinds of containers used. Aerial photography permits development of an understanding of the relationship of each site with respect to drainage, population, economic activities, and other waste disposal sites.

A sequence of photographs (**Figure 21.10**) illustrates the Environmental Protection Agency's (EPA's) ability to track the history of waste disposal sites. (See *https://semspub.epa.gov/work/04/11104345.pdf* for the complete EPA report.) The sequence of the most

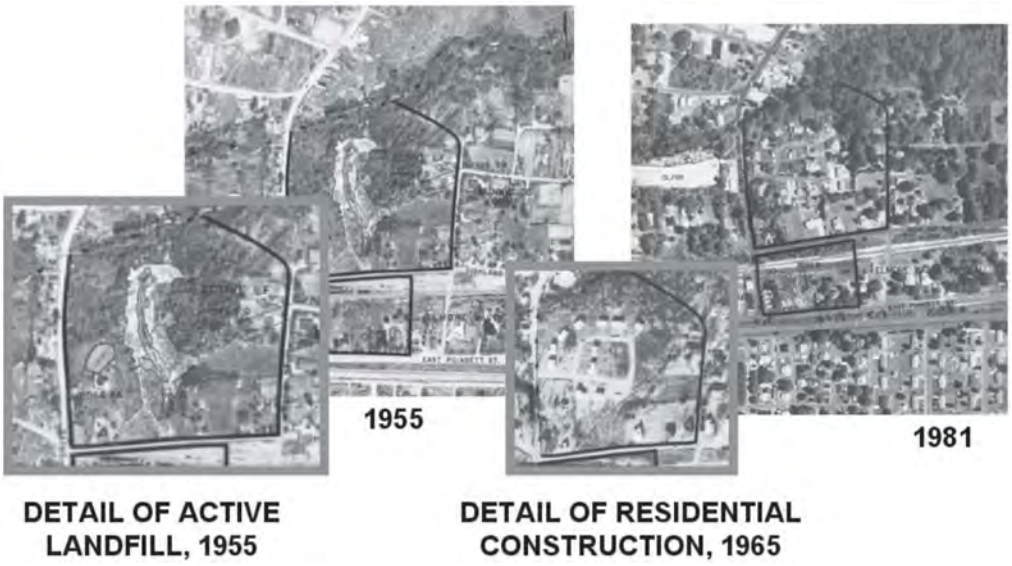


FIGURE 21.10 Aerial photographs depicting land-use changes at the Elmore/Sunnyside waste disposal area in Spartanburg County, South Carolina, 1955–1981. These photographs, selected from a longer sequence, show the active landfill (1955, left), newly constructed residential housing at the same site (1965, center), and the mature residential neighborhood (1981, right). Historical aerial photography permits analysts to unravel the historical sequence that otherwise is no longer visible at ground level for the on-site observation. From EPA, Environmental Photographic Interpretation Center.

recent photographs for this site shows how current land use can mask evidence that reveals the history of the site's use for waste disposal. These photographs show aerial photographs of the Elmore Waste Disposal site and the Sunnyside Dump in Spartanburg County, South Carolina. These sites are outlined on each photograph. The Elmore Waste Disposal site is about one-half acre in area, situated near an area that is now a residential neighborhood. The Elmore site was later designated as a superfund site, so it has been monitored by the EPA.

A large number of drums containing liquid wastes were placed at this site between 1975 and 1977. In 1977, the property owner signed a consent order with the state of South Carolina for cleanup of the site. However, state health and environmental officials judged the actions taken to be inadequate for the situation and ordered the owner to cease use of the site. Between 1981 and 1984, state and federal EPA officials investigated the site and found arsenic, chromium, and other heavy metals, as well as a number of volatile organic compounds, in the site soils.

At various times during this interval, between 150 and 300 drums were present on-site, as well as a 6,000-gallon tank, which was partially buried and contained contaminated waste. When the owner died in 1983, ownership passed to his heirs, one of whom continued to operate the site and accept waste drums. After many failed efforts to compel this owner to clean the site, in June 1986 state officials completed a state-funded removal of 5,500 tons of contaminated soil and 16,800 pounds of liquid waste, which were taken to an appropriate hazardous waste facility.

After this removal, groundwater monitoring wells and EPA's remedial investigation in 1991–1992 identified a contaminated groundwater plume extending some 700 feet north from the site. The estimated area underlain by this plume, north of the Elmore property, is 6–10 acres. Although no private water wells are located near this plume, the groundwater discharges to a creek. Additionally, surface soil in a one-quarter-acre area at one end of the Elmore property was found to be contaminated by lead and arsenic at levels exceeding health-based residential standards for those metals.

This area and its history illustrate the value of aerial photography in providing documentation of waste disposal sites and an overview of disposal sites, residential areas, and the drainage systems in their spatial context.

21.7 OTHER LAND-USE CLASSIFICATION SYSTEMS

A land-use classification system is not simply a list of categories. It must organize its classes according to an underlying logic that defines the relationships between classes. The organizing logic must impose a consistent organization on the classes. Classes must be mutually exclusive, and subclasses must nest within each other to form the hierarchical structure that we encountered in the Anderson system, for example. Furthermore, the distinctions between the classes must be consistent from class to class. For example, if some classes are defined on the basis of economic function, then the entire classification structure must reflect that logic.

General-Purpose Land-Use Classification

The Andersen (USGS) land-use classification system (Table 21.1) is specifically designed for use with aerial photography and related imagery. It is a good example of a *general-*

purpose land-use classification system, intended to provide a comprehensive classification of land use. General-purpose classifications are among the most widely used and are probably the classifications most likely to be encountered by photointerpreters. General-purpose land-use classification attempts to provide a classification that serves many purposes, although it is not specifically tailored for any specific application.

There were several predecessors to the Andersen system, with respect to scope and purpose.

- L. Dudley Stamp, a British economist, organized a national survey of Great Britain's land use from 1931 to 1938. He gathered large-scale land-use information for Britain using volunteers who recorded land use near their residences by annotating large-scale topographic maps. The maps were then mailed to a central office, where information was edited and posted to maps at 1:63,360. Although completion of this ambitious project was delayed by World War II, the information was still valuable when it was used in planning Britain's economic recovery after the war. Later, in 1950, Stamp's work supported broad-scale international mapping. His campaign extended its scope over a broad reach. His *Land Utilization Survey* led Stamp to develop a strategy for a World Land Use Survey (Stamp, 1951).

- In the United States, during the same interval, photointerpreters at the TVA defined procedures to interpret land use from aerial photography. TVA photointerpreters devised a system (no longer in use) that classified each parcel according to both its economic use and its physical properties.

- During the mid-1960s, scientists at Cornell University designed New York's Land Use and Natural Resources (LUNR) Survey, a database to record land use and related information for the entire state of New York. Much of the information was interpreted from 1:24,000 panchromatic aerial photography. Photointerpreters coded information according to predominant land use within 1-km cells, using 100 classes defined specifically by the LUNR project. This project was one of the first computerized land data inventories, forming an early precedent for the geographic information systems now in use today.

- Kreig (1970) described the approaches for LUNR and suggested best practices. He discussed timing of photography and scale of analysis, focusing on interpretation of land use at a scale of 1:24,000 and using mainly black-and-white imagery and stereoscopes when feasible. He proposed summer photography for agriculture and forestry (for northern landscapes). Larger scales (e.g., 1:6,000) create costs in time and effort, and so they were not favored. Kreig's manuscript could be used as an aid for analysts in the New York region.

Special-Purpose Land-Use Classification

Another approach to land-use classification—*special-purpose classifications*—is also significant. Special-purpose land-use classifications are designed to address a specific classification issue, with no attempt to provide comprehensive scope.

A good example of a special-purpose land-use classification system, and one that is of practical significance to photointerpreters, is the wetlands classification developed by Cowardin (1979), devoted specifically to classification of wetlands (Table 21.2). Its

logic therefore reflects the concerns of scientists who wish to portray the distribution of wetlands in detail and to indicate specific characteristics of wetlands that are of significance for the hydrologist and ecologist. For example, in the Anderson system, wetlands at Levels I and II are subdivided simply as the distinction between forested and nonforested wetlands, whereas in Cowardin's system, wetlands are distinguished first by two hydrologic criteria ("marine," "riverine," etc., and "tidal," "subtidal," etc.), and then by geomorphic and ecologic criteria ("reef," "streambed," etc.). Special-purpose systems such as Cowardin's can be seen as specialized alternatives to the general-purpose strategy, or in some instances, they might serve as Level III or IV within a hierarchical system such as the Anderson system.

It is important to select or design a system that is tailored to the needs of the client who will use the data. It is especially important to consider compatibility with previous classification systems if the results are to be compared with data from earlier dates. It is also important to consider compatibility with neighboring jurisdictions or with higher or lower units (such as state or city systems, in the case of a county survey), so that the classes match with these other data sets. A careful selection of classification that meets immediate needs may not be ideal if it does not permit comparison with data from these other units.

TABLE 21.2 Wetland Classification

Systems
1. Marine
2. Esturaine
3. Riverine
4. Lacustrine
5. Palustrine
Subsystems
1. Subtidal
2. Intertidal
3. Tidal
4. Lower perennial
5. Upper perennial
6. Intermittent
7. Limnetic
8. Littoral
Classes
1. Rock bottom
2. Unconsolidated bottom
3. Aquatic bed
4. Reef
5. Rocky shore
6. Unconsolidated shore
7. Streambed
8. Emergent wetland
9. Scrub-Shrub wetland
10. Forested wetland
11. Moss-Lichen wetland

Note: From Cowardin et al. (1979)

21.8 LAND COVER MAPPING BY IMAGE CLASSIFICATION

Land cover can be mapped by applying image classification techniques discussed in Chapter 12 to digital remote sensing images. In principle, the process is straightforward; in practice, many of the most significant factors are concealed among apparently routine considerations:

1. *Selection of images*. Success of classification for land cover analysis depends on the astute selection of images with respect to season and date. Therefore, the discussion in Chapter 7 of the design and interpretation of searches of image archives, though ostensibly mundane in nature, assumes vital significance for the success of a project. What season will provide the optimum contrasts between the classes to be mapped? Two or more dates might be required to separate all the classes of significance.

2. *Preprocessing*. Accurate registration of images and correction for atmospheric and system errors (Chapter 11) are required preliminary steps for successful classification. Subsetting of the region to be examined requires careful thought.

3. *Selection of classification algorithm*. The discussion in Chapter 12 reviewed many of the classification algorithms available for land cover analysis. The classification procedure should also be selected on the basis of local experience. AMOEBA, for example, tends to be accurate in landscapes dominated by large homogeneous patches, such as the agricultural landscapes of the midwestern United States, and less satisfactory in landscapes composed of many smaller heterogeneous parcels, such as those found in mountainous regions. Local experience and expertise are likely to be more reliable guides for selection of classification procedures than are universal declarations about their performance. Even when comparative information on classification effectiveness is available, it is difficult to anticipate the balance among effects of the choice of classifier, selection of image date, characteristics of the landscape, and other factors.

4. *Selection of training data*. Accurate selection of training data is universally significant for image classification, as we emphasized in Chapter 12. Training data for each class must be carefully examined to be sure that it is represented by an appropriate selection of spectral subclasses to account for variations in spectral appearance due to shadowing, composition, and the like. Many individual laboratories and image analysis software packages have applied unsupervised classification in various forms to define homogeneous regions from which to select training fields for supervised classification (e.g., Chuvieco and Congalton, 1988). Another approach to the same question is an algorithm that permits the analyst to select a pixel or group of pixels that forms the focal point for a region that grows outward until a sharp discontinuity is encountered. This process identifies a region of homogeneous pixels from which the analyst may select training fields for that class.

5. *Assignment of spectral classes to informational classes*. Because of the many subclasses that must be defined to accurately map an area by digital classification, a key process is the aggregation of spectral classes and their assignment to informational classes. For example, accurate classification of the informational class *deciduous forest* may require several spectral subclasses, such as *north-facing forest, south-facing forest, shadowed forest,* and the like. When the classification is complete, these subclasses should be assigned a common symbol to represent a single informational class.

6. *Display and symbolization.* The wide range of colors that can be presented on color displays and the flexibility in their assignment provide unprecedented opportunity for effective display of land cover information. Although unconventional choices of colors can sometimes be effective, it is probably sensible to seek some consistency in symbolization of land cover information to permit users to quickly grasp the meaning of a specific map or image without detailed examination of the legend. Therefore, the color symbols recommended by Anderson et al. (1976) may be useful guides. Another strategy for assignment of colors to classes is to mimic the colors used for USGS 7.5-minute quadrangles.

Within such general strategies, it is usually effective to assign related colors to related classes to symbolize Level II and Level III categories. For example, subclasses of agricultural land can be represented in shades of brown (using Anderson et al.'s strategy [1976] as a starting point), water and wetlands in shades of blue, subclasses of forest in shades of green, and so forth.

21.9 BROAD-SCALE LAND COVER STUDIES

The availability of multispectral Advanced Very High Resolution Radiometer (AVHRR),[1] Moderate Resolution Imaging Spectroradiometer (MODIS), and Visible Infrared Imaging Radiometer Suite (VIIRS) data (and data from similar meteorological satellites) on a regular basis has provided the capability to directly compile broad-scale land cover maps and data. In this context, *broad scale* refers to images that represent entire continents, or even entire hemispheres, based on data collected over a short period of time, perhaps about 10 days to 2 weeks. Previously, data for such large regions could be acquired only by generalizing more detailed information—a task that was difficult and inaccurate because of the incompleteness of coverage and the inconsistencies of the many detailed maps required to prepare small-scale maps. The finer-resolution data from the Landsat and Sentinel systems provide information that is of local and regional interest but is not now suitable for compilation of data at continental scales because of the effects of cloud cover, differences in Sun angle, and other factors that prevent convenient comparisons and mosaicking of many scenes into a single data set representing a large region.

AVHRR, MODIS, and VIIRS data, described in Chapter 7, provide coverage of entire continents over relatively short time periods. Accumulation of data over a period of a week to 10 days usually permits each pixel to be observed at least once under cloud-free conditions. Although the scan angle varies greatly, data are acquired at such frequent intervals that it is often possible to select coverage of the region of interest from the central section of each scene to reduce the effects of the extreme perspective at the edges of each scene.

Tucker, Townshend, and Goff (1985) examined AVHRR data for Africa using images acquired over a 19-month period. They examined changes in the vegetation index, with

[1]AVHRR: 1-km multispectral data from the U.S. National Oceanic and Atmospheric Adminstration (NOAA) satellite series (1979 to present). The NOAA 19 satellite that currently supports the AVHRR sensor has been degrading in orbit to the point where the nadir view is on the day–night terminator in its orbit.

results clearly illustrating climatic and ecological differences between major biomes, as observed using the vegetation index and seasonal variations in the vegetation index. They condensed the many variables into a concise yet potent data set that describes both seasonal and place-to-place variations in the vegetation index. Their land cover map, based on the first three principal components, is an extraordinary representation of key environmental conditions over an entire continent.

21.10 SOURCES OF COMPILED LAND-USE DATA

Earlier sections have often introduced land use and land cover analysis that address rather localized regions of small or modest size. Once aerial imagery, multispectral satellite data, and repeat coverage became available, it became feasible to conduct land use and land cover analyses at much broader scales.

Thus, during recent decades, remote sensing has contributed to the ability of organizations to prepare comprehensive land cover surveys of large regions. Satellite imagery, in particular, has provided near-simultaneous acquisition, broad areal coverage, uniform detail, and other qualities not available from alternative sources. In the United States, several governmental initiatives have exploited this capability to prepare broad-scale land cover surveys.

Land Use and Land Cover

In the 1970s, the USGS developed a national land cover mapping program, described in Anderson et al. (1976), based largely on manual interpretation of high-altitude aerial photography at scales of 1:60,000 and smaller. The resulting USGS LULC data files describe vegetation, water, natural surfaces, and cultural features for large regions of the United States. The USGS National Mapping Program provides these data sets and associated maps using map bases at 1:250,000. For more information, see:

https://usgs.gov/programs/national-geospatial-program/national-map

Multi-Resolution Land Characteristics Consortium

To economize acquisition of image data, in 1972 several agencies of the U.S. government formed a consortium to acquire satellite-based remotely sensed data to support environmental monitoring. The *Multi-Resolution Land Characteristics Consortium* (MRLC, *www.mrlc.gov*) is a team of federal agencies that generate land management cover for all 50 U.S. states and Puerto Rico. Initially, the MRLC supported the *National Land Cover Database,* a 30-m Landsat-based classification product encompassing all 50 states and Puerto Rico. Over time, the MRLC expanded to incorporate additional programs that support land cover mapping for five major themes of interest within the United States. These include (1) the National Land Cover Database (NLCD) project, for general land use and land cover, (2) the Coastal Change Analysis Program (C-CAP), with a focus on coastal wetlands, (3) the Gap Analysis project, for biodiversity, (4) the LANDFIRE project, for vegetation conditions related to fire and fuel characteristics, and (5) the Cropland Data Layer (CDL), for major crop types.

National Land Cover Database

The USGS, in partnership with several federal agencies, has developed NLCD products that provide spatially explicit and reliable information concerning land cover and land cover change at 30-m resolution. Earlier versions of this product were released first by decade and then roughly every 5 years (i.e., 1992, 1997, 2001, 2006, and 2011). The most recent generation of products, named *NLCD 2019*, offers eight integrated epochs of land cover from 2001 through 2019. NLCD 2019 released land cover products for 2001, 2003, 2006, 2008, 2011, 2013, 2016, and 2019 and impervious surface data for some but not all of those years, superseding earlier products. With the release of NLCD 2016, several products became available for public use, including land cover, rangeland, tree canopy cover, urban impervious, and land cover change, all of which help monitor the dynamics of our nation's resources and can be found at *www.mrlc.gov/data*.

Although there are some differences in the algorithmic approaches, there are general consistencies across the various releases of NLCD. The classification scheme is modified from Anderson et al. (1976) Level II classes (**Table 21.1**). Land cover is mapped using general land cover classes (**Figure 21.11**). For example, forest is classified as either deciduous, evergreen, or mixed. Land cover classification is based on the Landsat data archive and a host of ancillary sources. NLCD was the first national land cover data set produced since the early 1970s, effectively replacing the land cover data known to many as LUDA or GIRAS.

The suite of data products from NLCD has been widely used for many environmental applications, including land-use planning, hydrological analyses, habitat assessments, and many others. **Figure 21.12** shows an example of NLCD 2016 percentage of tree canopy cover (left) and percentage of impervious surface (right) for the Virginia coastline, including the heavily developed areas of Norfolk, Virginia Beach, and Newport News,

FIGURE 21.11 Left: NCLD 2011 Land Cover (2011 edition, amended 2014). Mosaic of the 48 contiguous U.S. states, using the USGS suggested color scheme. Right: The NLCD 2011 classification legend. From MRLC.

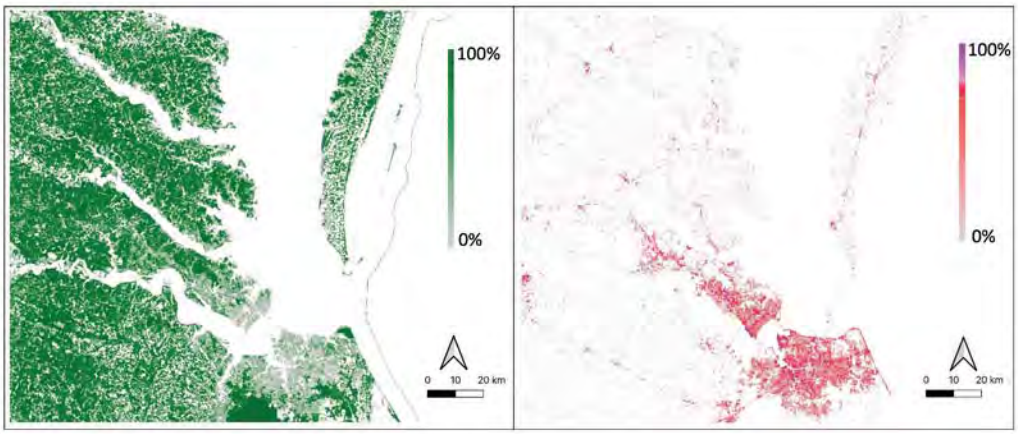


FIGURE 21.12 Sample images of NLCD 2016 Tree Canopy Cover (left) and Impervious Surface (right) for the Virginia coastline, United States. Note the heavily developed urban area of Norfolk and Newport News in the bottom center of the image, which corresponds to high percentage of impervious and low tree canopy cover. Paved road networks and other linear features can also be seen in the impervious layer.

as well as surrounding forested and agricultural areas. The Delmarva Peninsula and barrier islands are also visible. These two NLCD products (and the rangeland product) have continuous fields (i.e., they are percentages), as compared to the categorical NLCD classification shown in Figure 21.11. This has implications for how the data are analyzed and how change is interpreted over time. As an example, the categorical NLCD product can be used to assess change from one category to another (i.e., a transition from *forest* to *developed*). In contrast, the two examples in Figure 21.12 might show a percentage drop in tree canopy cover and a corresponding increase in percentage of impervious surface for the same type of change over time. Refer to Chapter 15 for more discussion on this topic.

Over the years, the algorithmic approaches to NLCD classification have been honed to increase accuracy. These strategies include using multiple images throughout the year to incorporate seasonal impacts; ancillary data (particularly digital elevation models to incorporate topographic information, that is, slope or aspect); indices such as the normalized tasseled cap transformation of the data; and stratification by mapping zones of common physiography and overall landscape characteristics (Figure 21.13). Further details are available at *www.mrlc.gov*.

Coastal Change Analysis Program

The C-CAP provides 30-m maps over time for areas of the coastal United States. The available dates for this product vary by location. Most coastal areas have coverage starting in 1996, with data updated roughly every 5 years. However, some areas have earlier dates, including as far back as 1975 for the Great Lakes. As with NLCD, land cover and change products are available. C-CAP focuses on intertidal areas, wetlands, and uplands. As such, it includes additional wetland water classes identifying palustrine versus estua-

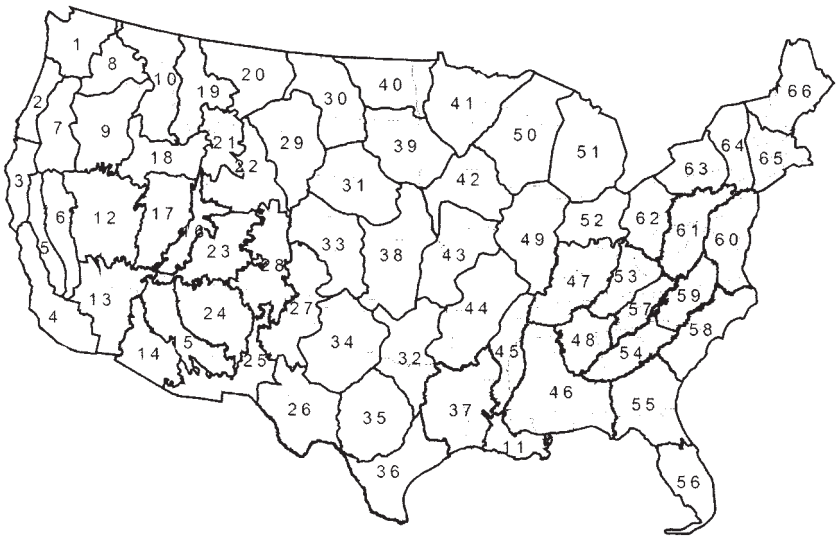


FIGURE 21.13 66 regions defined as parcels with consistent terrain and land-use conditions for the NLCD. From Homer et al. (2004).

rine areas (**Figure 21.14**), as compared to the generic "woody wetlands" and "emergent herbaceous wetlands" provided in NLCD.

The primary agency responsible for C-CAP is the NOAA Office for Coastal Management. Data and additional information can be found here:

https://coast.noaa.gov/digitalcoast/data/ccapregional.html

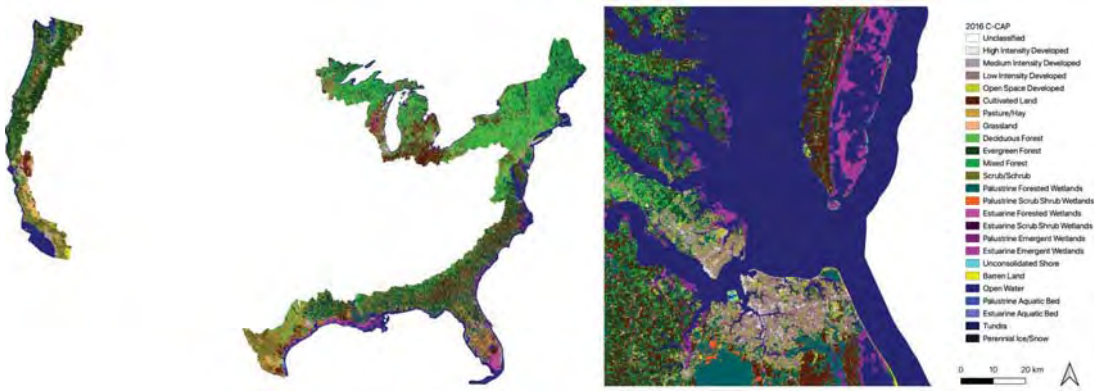


FIGURE 21.14 The 2016 Coastal Change Analysis Program product for the contiguous United States. Left: The spatial extent of the coastal area coverage. Right: A closer view of the Virginia Beach area with the individual classes shown. Note the increased detail in wetlands and aquatic classes as compared to NLCD. From NOAA.

Gap Analysis

In the United States, a national program to assess biodiversity and threats to biodiversity has been based on assessment of land cover derived from multispectral Landsat data. The objective of *gap analysis* is to identify "gaps" in biodiversity by surveying land cover patterns to assess cover types. For example, regions with high species diversity are often associated with high ranges of elevations, contrasts in soil characteristics, and the presence of varied vegetation classes. Gap analysis permits identification of high priorities for conservation and for development and land-use policies (Scott and Jennings, 1998). The vegetation and land cover classes used for this project are very detailed and are jointly part of the GAP/LANDFIRE National Terrestrial Ecosystems data used for both the GAP and LANDFIRE programs.

Each state conducts its own survey, using its own classification, mapping, and accuracy assessment strategy, developed in collaboration with neighboring states. The national project is coordinated by the National Gap Analysis Program within the USGS Biological Resources Division (USGS-BRD). For more information, see:

www.usgs.gov/core-science-systems/science-analytics-and-synthesis/gap/science/land-cover

LANDFIRE

The LANDFIRE program provides data to support decision making and land management for fire. It emerged because of ongoing concern over the trends in the increasing frequency, severity, and size of fires over the last few decades. The primary agencies responsible for the program are the U.S. Department of Agriculture Forest Service and the U.S. Department of the Interior. As part of the program, they produce data layers (over 20, including such things as disturbances, fire fuel, and topographic variables), as well as ecological models and tools for the public. More information can be found at *https://landfire.gov*.

Cropland Data Layer

The CDL was developed to provide estimates of the acreage of major commodity crops in the United States to the Agricultural Statistics Board. It is part of the suite of products by the MRLC Consortium. As described in Section 17.12, this program produces a yearly national product, beginning in 1997, which has a similar classification scheme to NLCD, but with more than 100 crop classes. For noncrop classes, there is strong consistency with NLCD because the NLCD is an input into the CDL classification algorithm (Lark et al., 2017). The CDL addresses agricultural-specific questions. The primary agency responsible is the U.S. Department of Agriculture National Agriculture Statistics Service (NASS). This agency also conducts validation and uncertainty analysis for the crop classes. The data can be accessed either directly from NASS as downloadable layers for each year or through a web-based service that enables viewing and some analysis of multiple dates at a time and crop frequency layers for specific common crops (corn, soybean, cotton, and wheat) (see **Figure 17.23** for an example). More information, as well as the CDL data, can be found at:

- *www.nass.usda.gov/Research_and_Science/Cropland/Release*
- *https://nassgeodata.gmu.edu/CropScape*

21.11 SUMMARY

Study of land use and land cover reveals the overall pattern of human occupation of the Earth's surface and the geographic organization of their activities. At broad scales, the land cover map provides a delineation of the broad patterns of climate and vegetation that form the environmental context for human activities. At local and regional scales, knowledge of land use and land cover forms a basic dimension of resources available to any political unit; both the citizens and the leaders of any community must understand the land resources available to them and the constraints that limit uses of land and environmental resources.

Although the formal study of land use and land cover dates from the early 1800s, systematic mapping at a large scale was not attempted until the 1920s, and aerial photography and remote sensing were not routinely applied until the 1960s. Thus, effective land use mapping is a relatively recent capability. We have yet to fully assemble and evaluate all the data that are available and to develop the techniques for acquiring and interpreting imagery.

Without the aerial images acquired by remote sensing, there can be no really practical method of observing the pattern of land cover or of monitoring changes. Systems such as Landsat, SPOT, Sentinel, and AVHRR have provided a capability for observing land cover at broad scales and at intervals that previously were not practical. Images from such systems have not only provided vital information but also data at new scales that have changed the intellectual perspective on the environment by recording broad-scale patterns and relationships that otherwise could not be accurately perceived or analyzed.

REVIEW QUESTIONS

1. Using aerial photographs and other information provided by your instructor, design Level III categories compatible with the USGS classification for a nearby region.
2. Review Chapters 7 and 11 to refresh your memory of resolutions of satellite sensors and the effects of mixed pixels. A typical city block is said to be about 300 ft × 800 ft in size. Make rough assessments of the effectiveness of the Landsat and Sentinel systems for depicting land use and land cover in urban regions. List factors other than the sizes of objects that would be important in making such assessments.
3. Outline some of the difficulties that would be encountered in compiling land use and land cover maps if aerial photography and remotely sensed data were not available.
4. Compare the relative advantages and disadvantages of alternative aerial imagery, including photography, thermal imagery, and radar imagery (at comparable scales and resolutions), for compiling land-use and land cover maps. List advantages and problems that might be encountered in using all three kinds of alternative aerial images in combination.

5. Review Section 21.5, and then prepare a diagram or flow chart that illustrates manual preparation of a land-use change map.
6. The following issues all require, directly or indirectly, use of accurate land-use and land cover information. For each, identify, in a few sentences or in short paragraphs, the role of accurate land-use maps and data.
   a. Solid waste disposal
   b. Selection of a location for a new electrical power plant
   c. The boundaries of a state park or wildlife preserve
   d. Zoning decisions in a suburban region near a large city
   e. Abandoned toxic waste dumps
7. Some scientists have advocated development of a classification system with categories based on the appearance of features on specific kinds of remotely sensed images. In contrast, the approach used by Anderson et al. (1976; see **Table 21.1**) is based on the idea that remotely sensed data should be categorized using classes that remain the same for all forms of remotely sensed images and match those used by planners. Compare the advantages and disadvantages of both strategies, considering the ease of application to the imagery and the ease of use by those who must actually apply the information.
8. About 18% of urban land is devoted to streets. About 20% of the land areas of large cities is said to be undeveloped. Assess the ability of remotely sensed images to contribute to assessing the amount and patterns of these two kinds of land use. Consider Landsat and Sentinel data, as well as aerial photography at 1:10,000.

REFERENCES

Anderson, J. R. 1977. Land Use and Land Cover Changes: A Framework for Monitoring. *Journal of Research, U.S. Geological Survey*, Vol. 5, No. 3, pp. 143–153.

Anderson, J. R., E. E. Hardy, J. T. Roach, and R. E. Witmer. 1976. *A Land Use and Land-Cover Classification for Use with Remote Sensor Data* (USGS Professional Paper 964). Washington, DC: U.S. Government Printing Office, 28 pp.

Baker, R. D., J. E. deSteiger, D. E. Grant, and M. J. Newton. 1979. Land-Use/Land-Cover Mapping from Aerial Photographs. *Photogrammetric Engineering and Remote Sensing*, Vol. 45, pp. 661–668.

Budreski, K. A., R. H. Wynne, J. O. Browder, and J. B. Campbell. 2007. Comparison of Segment and Pixel-Based Nonparametric Land Cover Classification in the Brazilian Amazon Using Multitemporal Landsat TM/ETM+ Imagery. *Photogrammetric Engineering and Remote Sensing*, Vol. 73, pp. 813–827.

Campbell, J. B. 1983. *Mapping the Land: Aerial Imagery for Land Use Information* (Resource Publications in Geography). Washington, DC: Association of American Geographers, 96 pp.

Campbell, J. B. 2001. Land Use and Land Cover Inventory. In *Manual of Photographic Interpretation* (2nd ed.; Warren R. Philipson, ed.). Bethesda, MD: American Society of Photogrammetry and Remote Sensing, pp. 335–364.

Chuvieco, E., and R. G. Congalton. 1988. Using Cluster Analysis to Improve the Selection of Training Statistics in Classifying Remotely Sensed Data. *Photogrammetric Engineering and Remote Sensing*, Vol. 54, pp. 1275–1281.

Cihlar, J. 2000. Land Cover Mapping of Large Areas from Satellites: Status and Research Priorities. *International Journal of Remote Sensing,* Vol. 21, pp. 1093–1114.

Congalton, R. G., M. Balough, C. Bell, K. Green, J. Milliken, and R. Ottman. 1998. Mapping and Monitoring Agricultural Crops and Other Land Cover in the Lower Colorado River Basin. *Photogrammetric Engineering and Remote Sensing,* Vol. 64, pp. 1107–1113.

Cowardin, L. M., V. Carter, F. C. Golet, and E. T. LaRoe. 1979. *Classification of Wetlands and Deepwater Habitat of the United States.* Washington, DC: U.S. Department of the Interior, Fish and Wildlife Service.

Erb, T. L., W. R. Philipson, W. Teng, and T. Liang. 1981, October. Analysis of Landfills with Historic Airphotos. *Photogrammetric Engineering and Remote Sensing,* Vol. 47, No. 9, pp. 1363–1369.

Estes, J. E., and L. Senger. 1972. Remote Sensing in the Detection of Regional Change. In *Proceedings, 8th International Symposium on Remote Sensing of Environment.* Ann Arbor: Environmental Research Institute of Michigan, pp. 317–324.

Hoeschele, W. 2000. Geographic Information Systems and Social Ground Truth in Attappadi, Kerala State, India. *Annals, Association of American Geographers,* Vol. 90, pp. 293–321.

Homer, C., C. Q. Huang, L. M. Yang, B. Wylie, and M. Coan. 2004. Development of a 2001 National Land-Cover Database for the United States. *Photogrammetric Engineering and Remote Sensing,* Vol. 70, pp. 829–840.

Jensen, J. R. 1979. Spectral and Textural Features to Classify Elusive Land Cover at the Urban Fringe. *Professional Geographer,* Vol. 31, pp. 400–409.

Jensen, J. R. 1981. Urban Change Detection Mapping Using Landsat Digital Data. *American Cartographer,* Vol. 8, pp. 127–147.

Jensen, J. R. 1996. *Introductory Digital Image Processing: A Remote Sensing Perspective.* Upper Saddle River, NJ: Prentice-Hall, 316 pp.

Jensen, J. R., and D. L. Toll. 1982. Detecting Residential Land-Use Development at the Urban Fringe. *Photogrammetric Engineering and Remote Sensing,* Vol. 48, pp. 629–643.

Kreig, R. A. 1970. Aerial Photographic Interpretation for Land Use Classification in the New York State Land Use and Natural Resources Inventory. *Photogrammetria,* Vol. 26, pp. 101–111.

Lark, T. J., R. M. Mueller, D. M. Johnson, and H. K. Gibbs. 2017. Measuring Land-Use and Land-Cover Change in the U.S. Department of Agriculture's Cropland Data Layer: Cautions and Recommendations. *International Journal of Applied Earth Observation and Geoinformation,* Vol. 62, pp. 224–235.

Loelkes, G. L. 1977. *Specifications for Land Cover and Associated Maps* (U.S. Geological Survey Open File Report 77–555). Reston, VA: U.S. Geological Survey.

Loveland, T. R., T. L. Shol, S. V. Stehman, A. L. Gallant, K. L. Sayler, and D. E. Napton. 2002. A Strategy for Estimating the Rates of Recent United States Land Cover Changes. *Photogrammetric Engineering and Remote Sensing,* Vol. 68, pp. 1091–1100.

Lunetta, R. S., and S. D. Elvidge. 1998. *Remote Sensing Change Detection: Environmental Monitoring Methods and Applications.* Ann Arbor, MI: Ann Arbor Press, 318 pp.

Lyon, J. G. 1987. Use of Maps, Aerial Photographs, and Other Remote Sensor Data for Practical Evaluations of Hazardous Waste Sites. *Photogrammetric Engineering and Remote Sensing,* Vol. 53, pp. 515–519.

Merteng, B., and E. F. Lambin. 2000. Land-Cover-Change Trajectories in Southern Cameroon. *Annals, Association of American Geographers,* Vol. 90, pp. 467–494.

Nunnally, N. R., and R. E. Witmer. 1970. Remote Sensing for Land Use Studies. *Photogrammetric Engineering and Remote Sensing,* Vol. 36, pp. 449–453.

Philipson, W. R. 1997. *Manual of Photographic Interpretation* (2nd ed). Bethesda, MD: American Society of Photogrammetry and Remote Sensing, 689 pp.

Robinove, C. J. 1981. The Logic of Multispectral Classification and Mapping of Land. *Remote Sensing of Environment,* Vol. 11, pp. 231–244.

Scott, J. M., and M. D. Jennings. 1998. Large Area Mapping of Biodiversity. *Annals of the Missouri Botanical Garden,* Vol. 85, pp. 34–47.

Smith, J. H., J. D. Wickham, S. V. Stehman, and L. M. Yang. 2002. Impacts of Patch Size and Land-Cover Heterogeneity on Thematic Image Classification Accuracy. *Photogrammetric Engineering and Remote Sensing,* Vol. 68, pp. 65–70.

Stamp, L. D. 1951. The World Land Use Survey. *Nature*, Vol. 167, No. 4260, pp. 1010–1011.

Townshend, J. R. G. 1994. Global Data Sets for Land Applications from the Advanced Very High Resolution Radiometer: An Introduction. *International Journal of Remote Sensing*, Vol. 15, pp. 3319–3332.

Tucker, C. J., J. R. G. Townshend, and T. E. Goff. 1985. African Land Cover Classification Using Satellite Data. *Science,* Vol. 227, pp. 369–375.

Vogelmann, J. E., D. Helder, R. Morfit, M. J. Choate, J. W. Merchant, and H. Bulley. 2001. Effects of Landsat 5 Thematic Mapper and Landsat 7 Enhanced Thematic Mapper Plus Radiometric and Calibrations and Corrections on Landscape Characterization. *Remote Sensing of Environment,* Vol. 78, pp. 55–70.

Vogelmann, J. E., S. M. Howard, L. Yang, C. R. Larson, B. K. Wylie, and N. Van Driel. 2001. Completion of the 1990s National Land Cover Data Set for the Coterminous United States from the Landsat Thematic Data and Ancillary Data Sources. *Photogrammetric Engineering and Remote Sensing,* Vol. 70, pp. 650–662.

Vogelmann, J. E., T. Sohl, and S. M. Howard. 1998. Regional Characterization of Land Cover Using Multiple Sources of Data. *Photogrammetric Engineering and Remote Sensing,* Vol. 64, pp. 45–57.

Index

Note. *f* or *t* following a page number indicates a figure or table.

B

C

D

E

F

G

H

I

J

K

L

M

N

O

P

Q

R

S

T

U

V

W

X

Z

About the Authors

James B. Campbell, PhD, is Professor of Geography at Virginia Tech, where he teaches remote sensing, quantitative methods, and geomorphology. He has worked closely with students and faculty in forestry, geology, agronomy, and environmental sciences. The author of numerous technical articles and several books, Dr. Campbell has received the Outstanding Service Award and the Fellow Award from the American Society for Photogrammetry and Remote Sensing, as well as the Outstanding Service Medal from the Remote Sensing Specialty Group of the Association of American Geographers. In 2020, Dr. Campbell received the AmericaView Lifetime Achievement Award. He has served as a principal investigator for the VirginiaView consortium and as a member and chair of the AmericaView Board of Directors.

Randolph H. Wynne, PhD, is Professor in the Department of Forest Resources and Environmental Conservation at Virginia Tech. He also serves as Director of the Interdisciplinary Graduate Education Program in Remote Sensing. He teaches courses focused on the environmental and natural resources applications of remote sensing at the senior and graduate levels. Dr. Wynne's research interests are in the applications of remote sensing to forestry, natural resource management, ecosystem ecology, and earth system science. He is a recipient of the Award in Forest Science from the Society of American Foresters. Dr. Wynne is Coeditor of the journal *Science of Remote Sensing* and Associate Editor of *Remote Sensing of Environment*.

Valerie A. Thomas, PhD, is Professor in the Department of Forest Resources and Environmental Conservation at Virginia Tech. She also serves as Co-Director of the Center for Environmental Analytics and Remote Sensing within the College of Natural Resources and Environment. Dr. Thomas teaches remote sensing courses in forest lidar applications and hyperspectral applications for natural resources. She also teaches about the linkages between forests, society, and climate. Dr. Thomas's research related to remote sensing of forest cover, function, and change has been funded through federal and state agencies and by industry.