# Remote Sensing: The Quantitative Approach

Edited by
PHILIP H. SWAIN
and
SHIRLEY M. DAVIS

Edited by
**PHILIP H. SWAIN**
and
**SHIRLEY M. DAVIS**

Contributing authors

| | |
|---|---|
| **SHIRLEY M. DAVIS** | **ROGER M. HOFFER** |
| **DAVID A. LANDGREBE** | **JOHN C. LINDENLAUB** |
| **TERRY L. PHILLIPS** | **LeROY F. SILVA** |
| **PHILIP H. SWAIN** | |

*All from Laboratory for Applications of Remote Sensing*
*Purdue University,*
*West Lafayette, Ind., U.S.A.*

# Remote Sensing:
# The Quantitative Approach

Remote sensing is not a new technology. For many decades man has been ascending above the earth in order to observe it from a distance and thus learn more about its condition. Aerial photography has been used extensively for this purpose and over the years a sophisticated technology has evolved using photographic sensors for remote sensing.

The recent development of satellites carrying earth-observational sensor systems has made available enormous quantities of photographic and other forms of data about the surface of the earth, data which have great potential for helping to solve many human problems: for relieving critical food shortages; for monitoring and controlling environmental pollution; for augmenting shrinking supplies of natural resources; and for planning the orderly growth of cities. In view of these needs, these satellite data are of great human value, provided they can be reduced to useful information both quickly and economically. Modern, high-speed digital computers are well suited to this data-reduction task, and the synthesis of computer technology with the new observation systems has already revolutionized our ability to obtain accurate and current information about the world we live in. The product of this synthesis is the quantitative approach to remote sensing.

To an extent, we have indulged our prejudices in using the phrase "the quantitative approach" to describe the remote sensing approach emphasized here. Other remote sensing methods, some of which have been utilized much longer than the relatively new technology we describe, also have quantitative techniques associated with them. Yet we make no apology for our choice of terms. No other approach has been so fundamentally concerned with the quantitative collection, storage, processing, and application of remote sensing data. The new sensor systems

can be calibrated more precisely and operate over greater dynamic and spectral ranges with greater sensitivity than any instrumentation previously employed for this purpose. The integrity of the data is preserved by converting the data to and recording and storing it in digital form (i.e., computer-compatible form) right at the sensor. The computer-implemented analysis methods are inherently quantitative (objective, numerical) and produce results which are quantitative (numerical, statistical). The results are often most appropriately utilized in applications demanding quantitative information, often distilled in a rapid fashion from relatively large volumes of sensor data.

The quantitative approach to remote sensing has evolved rapidly during the past decade, and its documentation is scattered throughout the technical literature of many disciplines. While this multidisciplinary nature of the technology attests to its depth and the breadth of its application, it also makes it harder for the new student to quickly gain a comprehensive understanding of the approach. Prior to the publication of this book, a unified, tutorial treatment of the quantitative approach had not been available.

*Remote Sensing: The Quantitative Approach* deals with both the theoretical bases and the practical aspects of this approach to remote sensing. The subject is developed through a sequence of chapters that evolves logically from the technology itself. After the multispectral and multitemporal concepts are introduced, a generic description of a quantitative remote sensing system is presented to serve as the framework for the book as a whole. An explanation of the physical theory which underlies data-collection systems leads to an extended description in Chapter 2 of those sensors most appropriate for collecting quantitative data. The statistical theory which supports the various data-analysis steps is presented in Chapter 3, and methods of implementing data-handling and data-analysis tasks are explored in Chapter 4. In Chapters 5 and 6, we see first why the natural spectral conditions of the earth can be described quantitatively and then move to a detailed case study describing an actual large-area agricultural survey which employed the quantitative approach. In the final chapter, multispectral remote sensing, based on both image and multivariate analyses, is viewed in a more complex framework appropriate to the current state of the technology, and future prospects are suggested.

It is our intention to provide the serious reader of this book with a systematic way to develop his understanding of quantitative remote sensing. The material presented here is *both* basic in nature and theoretical. The book is not designed to teach the reader how to use a particular system or how to apply the technology in a specific application but rather to give him both an awareness of some of the options open to him as a remote sensing user or designer and the theoretical understanding necessary for making wise choices.

This book can be used in many ways by people with many different objectives and backgrounds. The level of presentation assumes only a general scientific background, and thus the book can be used by advanced undergraduate or graduate students in a variety of disciplines. It is equally suited for use by professionals in governmental agencies or private industry who need to understand

the theoretical bases of quantitative remote sensing, perhaps as a step in evaluating its usefulness in their own work.

As an aid to readers at all levels, each chapter contains statements of educational objectives; these give the reader a quick preview of the material covered and, in the case of group study, provide the teacher or group leader with a convenient aid for focusing the group's attention on those aspects of the technology most relevant to their study. The questions and exercises aid the reader in evaluating his grasp of the concepts presented and serve as departure points for group discussions. These objectives and exercises were prepared by Dr. John C. Lindenlaub, who has long been active in the development and utilization of non-traditional instructional materials in remote sensing and engineering. The references included at the end of each chapter will assist the reader who wishes to continue his study to greater depth in the open literature. The glossary defines technical terms from many disciplines within the context of the quantitative approach to remote sensing of earth resources.

Because of the inherently multidisciplinary nature of quantitative remote sensing, any tutorial presentation of the material can be greatly enhanced by a unified, systematic approach. Ideally such a presentation should also be made in a truly multidisciplinary way, not through the eyes of only the engineer, the computer manager, or the agronomist but as the effective synthesis of the perspectives they all provide. The seven authors and editors of this book have attempted to attain this synthesis. Among them they represent many of the major disciplines associated with remote sensing, and yet, as long-time members of the research and administrative staff of a single laboratory, the Laboratory for Applications of Remote Sensing (LARS) at Purdue University, they have together evolved philosophies and approaches essential to effective interdisciplinary research.

We would like to acknowledge the contributions of several organizations in the development of the remote sensing technology described here: Purdue University for providing the environment conducive to advanced multidisciplinary research; the U.S. Department of Agriculture which promoted the establishment of LARS in 1966; and the National Aeronautics and Space Administration (NASA) for its continued encouragement and support of remote sensing research.

<div align="right">P.H.S.<br>S.M.D.</div>

# CONTENTS

v

## Chapter 3 Fundamentals of Pattern Recognition in Remote Sensing 136
*Philip H. Swain*

## THE QUANTITATIVE APPROACH: CONCEPT AND RATIONALE
*David A. Landgrebe*

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 the object of interest, and although there are other possibilities (e.g., seismic waves, sonic waves, and gravitational force), our attention in this text is focused on systems which measure electromagnetic energy. Sometimes there is playful debate concerning how far an object must be from the measuring device before it can be considered "remote." But this is merely a debate over semantics; we wish to concern ourselves with methodology. Although the applications to be discussed in this book involve measurements made hundreds or thousands of meters from the object of interest, we can also conceive of applications in which the measuring device might be only a fraction of a meter from the object.

**Study objectives**
After studying Sec. 1-1, you should be able to:
1. Briefly describe the impact that the digital computer and pattern recognition have had on remote sensing technology.

## 1-1 BACKGROUND

It is difficult to establish a specific time or event as marking the beginning of remote sensing. Some cite the use of photography from tethered balloons before the American Civil War; others say it began even earlier.[1] However, the development of the techniques involved has been perhaps more intense since early in the twentieth century.

Many of the early techniques used in remote sensing were developed for military reconnaissance during both World Wars I and II, but it was not long before the possibility of applying these methods to the solution of peacetime problems became apparent. Civil engineers were among the first to use remote sensing for nonmilitary purposes when they turned to it early in this century as a means for surveying and mapping; and in the 1930s the U.S. Department of Agriculture began nationwide use of aerial photography for soil survey work. The practice of using a human interpreter to analyze photographs obtained from airborne cameras was well established by the 1940s,[2] and investigations into the use of color and infrared photography increased the effectiveness of the aerial camera as a sensor of the earth's surface.[3]

During the 1950s and early 1960s, the advent of the digital computer and attempts at modeling the processes associated with human intelligence combined to accelerate the development of a data-analysis technology called "pattern recognition."[4,5] During the same period, significant developments were made in the design of sensor systems,[6] particularly in the measurement of energy in the infrared wavelengths. Taken together, these advancements were to give rise to a whole new approach to remote sensing—the quantitative approach—based on the ability to rapidly and precisely measure and process multispectral remote sensing data.

Thus today we find there are two major branches of remote sensing. The branch first mentioned above we refer to as *image oriented* because it capitalizes on the pictorial aspects of the data and utilizes analysis methods which rely heavily on the generation of an image. The second branch is referred to as *numerically oriented* because it results directly from the development of the computer and because it emphasizes the inherently quantitative aspects of the data, treating the data abstractly as a collection of measurements. In this case an image is not thought of as data but rather as a convenient mechanism for viewing the data.

The image-oriented technology is older and more fully developed. The sensors most common to image-oriented systems (photographic cameras) have long been in use, as have the associated analysis techniques (photo interpretation). The technology has proved to be reliable and economical for a wide variety of operational applications. Even so, further development of this technology continues.

By comparison, the technology of numerically oriented systems is still in its infancy, having developed to its present state in a short span of years. A multispectral scanner is often used as the sensor in these systems. This instrument is preferred because it has a greater dynamic range over a larger portion of the

electromagnetic spectrum than is available from photographic systems. Although a computer is usually used for analysis in numerically oriented systems, it may also be used in image-oriented systems; likewise, there are many other tools and techniques which are equally applicable to both technologies, e.g., the methods of photogrammetry. For some applications the numerical approach has already achieved a practical level of utility, and, even though it still holds much undeveloped potential, it is now beginning to assume its proper role alongside the older, image-oriented technology.

Today we are acquiring earth observational data from earth-orbiting satellites; because of the wide view possible from satellite altitudes, the speed with which the satellite-borne sensors travel, and the number of spectral bands used, very large quantities of data are being produced. Since in many instances information obtainable through remote sensing has value only if it can be acquired both rapidly and cost effectively, it is quite natural to turn to computers to achieve this economy and speed in extracting reliable information from the large volume of data available.

The purpose of this book is to describe an approach in which modern sensor systems, pattern recognition, computers, and human beings can be used together to synthesize a numerically oriented remote sensing system. In the next section we turn to a brief overview of some of the basic concepts of this approach.

**Study objectives**

After studying Sec. 1-2, you should be able to:
1. Name and give an example of three major kinds of variation which convey information in remote sensing.
2. Give at least one reason why data values from the same class are not identical but tend to cluster or group themselves around some mean value.

## 1-2 HOW INFORMATION IS CONVEYED IN REMOTE SENSING DATA

Let us begin by considering some of the fundamentals. Figure 1-1 (between pages 36 and 37) is an air view of a portion of the surface of the earth, and one can immediately gain some information about what is in the scene from the image. Stop for a moment and think about what leads you to the conclusions you can reach with just a glance. We can see many linear features, or straight lines. This suggests human activity in the area.

These features divide the area into fields, and some of the fields are rather uniform, again suggesting human activity—in this case, agriculture. Some fields, notably those in the foreground, have a mottled effect. If there are crops there, apparently the crop canopy has not developed enough to mask the natural variations of the soil types in these fields.

Some of the linear features are roads, and near the roads there are small, angular objects which are buildings on the farmsteads. People not only work but

live in this area, and, furthermore, it is possible to estimate the number of people from the number of buildings.

We could go on extracting more and more information from the scene. Looking back now at the information we derived, we realize that what led us to these conclusions were by and large the geometrical forms apparent in the image, i.e., the spatial organization of the energy emanating from the scene and recorded by the photographic film. Thus we may conclude at this point that the spatial organization of the energy contains much information about what is in the scene.

In Fig. 1-2 (between pages 36 and 37) we see the same scene at the same time of year but with something added—color. Now we can see not only the spatial distribution of energy but a significant indication of its spectral distribution as well. Studying this figure in the same way we studied Fig. 1-1, we can derive even more information about what was in the scene and what its condition was. For example, in the color image the tan color of the harvested and unharvested fields of small grain is easier to identify.

Figure 1-3 (between pages 36 and 37) shows the scene once more, but this time the image was formed using color infrared film. This type of film will be discussed in more detail in Chap. 2. For now, though, we will simply state that instead of being sensitive to the energy distribution in the blue, green, and red portions of the spectrum, this film is sensitive to the energy from the near-infrared portions of the spectrum as well, and thus provides new information about the scene. For example, since green vegetation has a relatively high reflectance in the near-infrared wavelengths, it is easier to detect relatively minor changes in a vegetative canopy in this portion of the spectrum than it is in the visible portion of the spectrum.

The color and color infrared images of Figs. 1-4 and 1-5 (between pages 36 and 37) give a clear example of this. Note the trapezoidal field on the right-hand side of each image. This is a corn field in which alternate rows have been planted with different varieties. One variety has become infected by a disease and its condition is much more apparent in Fig. 1-5 because of the infrared sensitivity of the film. The important point to notice is that by having data available from the near-infrared wavelengths, more information is available to us than when we are limited to viewing the scene in only the visible wavelengths.†

So far we have confined our discussion to the regions of the spectrum in which photographic sensors can respond, but other parts of the spectrum contain still more information about the earth scene. Figure 1-6 describes the entire electromagnetic spectrum. At the top are shown the various types of radiation, displayed as a function of the wavelength. It is the portion of the optical region, shown in detail in the lower portion of the figure, which is of primary interest to us, although other portions of the electromagnetic spectrum such as the microwave region are also useful in remote sensing. Notice that radiation measurements available from photographic sensors include only those in the visible region and a small portion of

---

† Because remote sensing makes use of spectral information within as well as beyond the visible region of the spectrum, the terminology of spectroscopy (wavelength, power per unit wavelength, etc.) has been chosen in the text rather than that of colorimetry (hue, chroma, etc.).
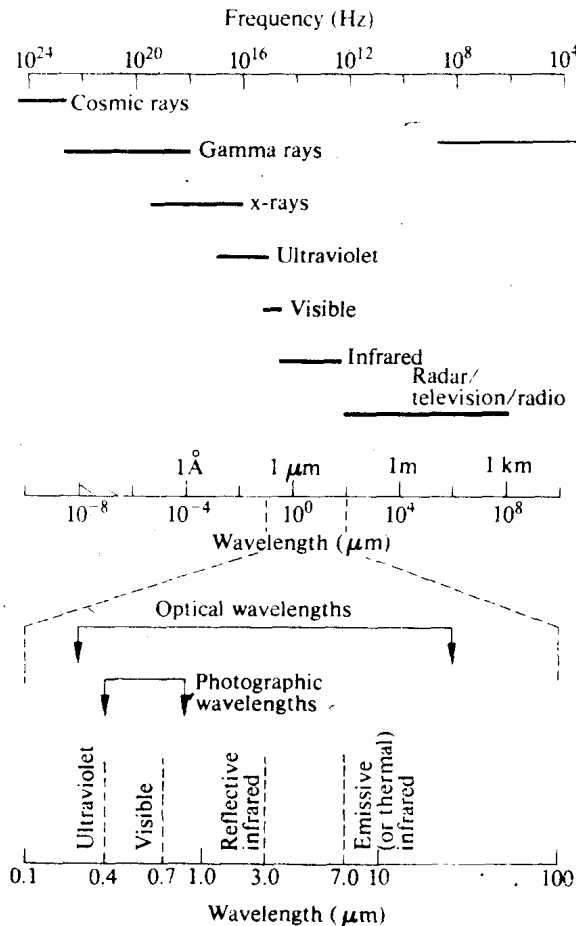
**Figure 1-6** The electromagnetic spectrum.

the infrared region. The optical portion of the electromagnetic spectrum will receive more detailed attention early in Chap. 2.

Summarizing to this point, then, we have illustrated the fact that one may derive information about the materials covering the surface of the earth from the *spatial* and *spectral* distributions of energy emanating from those materials. Shortly, we will point out that *temporal* (time) variations in the scene are also useful in the information-extractive process, since much information about what is in an area can be derived by seeing how the area changes with time.

Before doing this, however, let us consider in a bit more detail the manner in which data may be analyzed and information realized about a scene.[7] Figure 1-7 shows a very small portion of an agricultural area imaged in three portions of the spectrum: the visible, the near infrared, and the far infrared. Suppose for the moment that all the data gathered will represent only the four classes of materials contained in this small scene, i.e., corn, alfalfa, bare soil. and stubble. Thus, in analyzing the data gathered from a region of unknown ground cover, we have only four possibilities to choose from. On what can we base our choice?
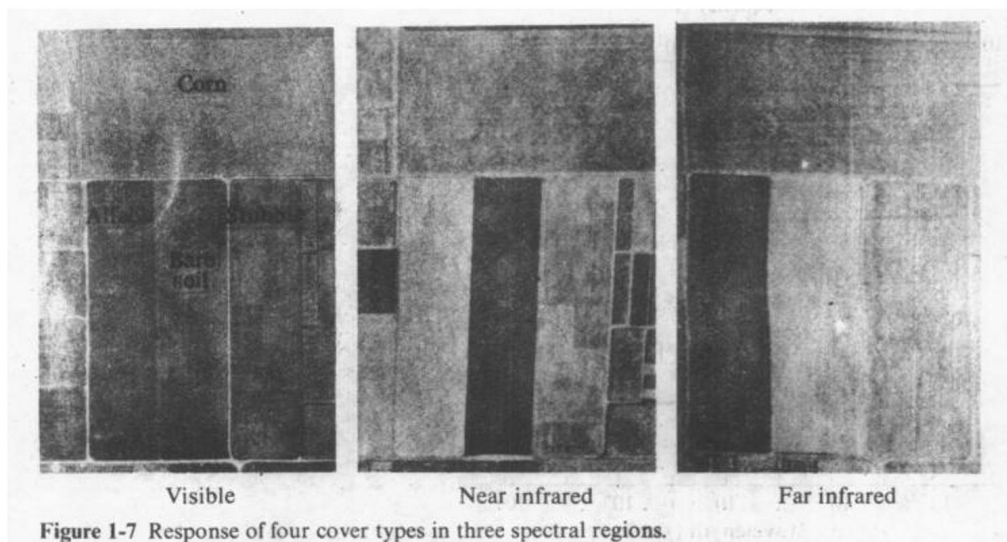
Visible        Near infrared        Far infrared

**Figure 1-7** Response of four cover types in three spectral regions.

First, we can carefully inspect the image from the visible region. Notice that the area marked "corn" has a distinctive texture which is not seen in the other three materials. Thus each time we see this texture in an unknown area, we can make a decision in favor of corn over the other classes. Texture, however, is a very difficult image characteristic to quantify. To do so requires a precise definition of what is meant by texture which, if made adequately robust, may be too complex to be utilized efficiently.

There are, however, other approaches to the discrimination problem. Notice, for example, that the response from alfalfa in the three spectral regions shown in the figure changes from very dark to very bright to very dark and that this sequence of responses as a function of wavelength does not occur for any of the other three materials. Thus we have a distinctive way to discriminate alfalfa from the other three materials, and the process relies on a feature more easily quantified than texture, namely, the radiant energy in each of the spectral regions.

Pursuing this line of reasoning further, consider Fig. 1-8. This figure provides a hypothetical but conceptually accurate graph of the relative reflectance as a function of wavelength for three simple ground-cover classes: vegetation, soil, and water. Let us select two wavelengths of interest, marked $\lambda_1$ and $\lambda_2$ in Fig. 1-8($a$), and display the spectral responses in two-dimensional measurement space, as shown in Fig. 1-8($b$). In this case, we are showing the response at wavelength $\lambda_1$ versus the response at wavelength $\lambda_2$. As we saw earlier (Fig. 1-7), distinct response sequences can aid in class identification; they produce points in different portions of the two-dimensional space. The discrimination or identification process can therefore be reduced to determining the class associated with the region in the space where the data points fall.

Notice immediately that we can use response data at more than two wavelengths for the analysis. For example, we have marked a third in Fig. 1-8($a$). Use of this wavelength in addition to the others would require a three-dimensional version
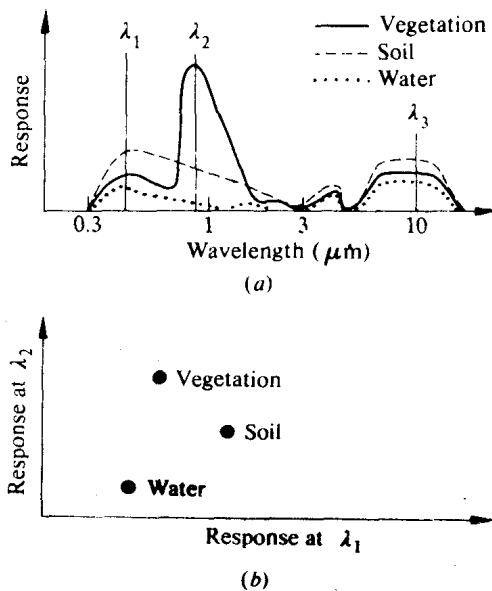
Figure 1-8  (a) Hypothetical spectra and (b) their representation in two-dimensional measurement space. (*After Landgrebe.*[7])

of the space shown in Fig. 1-8(b). Indeed, four or more wavelengths could be used and would present no particular difficulty if, for example, one were using a computer in analyzing the data.

There is another characteristic of the reflectance data which must be incorporated into this theoretical example in order to present an adequate basic overview of the quantitative approach. While the spectral response patterns from all healthy, green vegetation tend to be generally the same, there are small variations both within and between vegetative types. The variations, of course, are what make it possible to identify different plant species. In terms of Fig. 1-8(b) this means, however, that the data point for vegetation in two-dimensional space should not really be a point but a cluster of points, which could be described by a statistical distribution. Thus a statistical model for the response of a given material is needed in order to deal properly with, and indeed to take advantage of, the natural variability within classes.

The analysis task therefore becomes one of deciding how to partition the multivariate space into regions associated with each class so that a data point occurring in any part of the space will be uniquely assigned to a class. The actual situation is illustrated in Fig. 1-9, in which points from three classes of material are plotted in two-dimensional space. Note that the distribution of the points within a class, as illustrated here, is more typical of the actual situation. The question we must raise now is: By what rule should we associate an unknown point, marked U in the figure, with one of the three classes?

There are many algorithms (calculation procedures) which have been proposed in the engineering and statistical literature for answering this question. One of the simplest is illustrated in Fig. 1-10. In this case the conditional mean (class average value) for each class of material has been determined and is marked by the bold-