tispectral data are used to perform the classification and, indeed, the spectral pattern present within the data for each pixel is used as the numerical basis for categorization. That is, different feature types manifest different combinations of DNs based on their inherent spectral reflectance and emittance properties. In this light, a spectral "pattern" is not at all geometric in character. Rather, the term *pattern* refers to the set of radiance measurements obtained in the various wavelength bands for each pixel. *Spectral pattern recognition* refers to the family of classification procedures that utilizes this pixel-by-pixel spectral information as the basis for automated land cover classification.

Spatial pattern recognition involves the categorization of image pixels on the basis of their spatial relationship with pixels surrounding them. Spatial classifiers might consider such aspects as image texture, pixel proximity, feature size, shape, directionality, repetition, and context. These types of classifiers attempt to replicate the kind of spatial synthesis done by the human analyst during the visual interpretation process. Accordingly, they tend to be much more complex and computationally intensive than spectral pattern recognition procedures.

Temporal pattern recognition uses time as an aid in feature identification. In agricultural crop surveys, for example, distinct spectral and spatial changes during a growing season can permit discrimination on multidate imagery that would be impossible given any single date. For example, a field of winter wheat might be indistinguishable from bare soil when freshly seeded in the fall and spectrally similar to an alfalfa field in the spring. An interpretation of imagery from either date alone would be unsuccessful, regardless of the number of spectral bands. If data were analyzed from both dates, however, the winter wheat fields could be readily identified, since no other field cover would be bare in late fall and green in late spring.

As with the image restoration and enhancement techniques we have described, image classifiers may be used in combination in a hybrid mode. Also, there is no single "right" manner in which to approach an image classification problem. The particular approach one might take depends upon the nature of the data being analyzed, the computational resources available, and the intended application of the classified data.

In the remaining discussion we emphasize spectrally oriented classification procedures for land cover mapping. (As stated earlier, this emphasis is based on the relative state of the art of these procedures. They currently form the backbone of most multispectral classification activities.) First, we describe *supervised classification*. In this type of classification the image analyst "supervises" the pixel categorization process by specifying, to the computer algorithm, numerical descriptors of the various land cover types present in a scene. To do this, representative sample sites of known cover type, called *training areas*, are used to compile a numerical "interpretation key" that describes the spectral attributes for each feature type of interest. Each pixel in the data set is then compared numerically to each category in the interpretation key

and labeled with the name of the category it "looks most like." As we see in the next section, there are a number of numerical strategies that can be employed to make this comparison between unknown pixels and training set pixels.

Following our discussion of supervised classification we treat the subject of *unsupervised classification*. Like supervised classifiers, the unsupervised procedures are applied in two separate steps. The fundamental difference between these techniques is that supervised classification involves a training step followed by a classification step. In the unsupervised approach the image data are first classified by aggregating them into the natural spectral groupings, or *clusters*, present in the scene. Then the image analyst determines the land cover identity of these spectral groups by comparing the classified image data to ground reference data. Unsupervised procedures are discussed in Section 7.11.

Following our treatment of supervised and unsupervised classification we discuss *hybrid classification* procedures. Such techniques involve aspects of both supervised and unsupervised classification and are aimed at improving the accuracy or efficiency (or both) of the classification process. Hybrid classification is the subject of Section 7.12.

We reiterate that the various classification procedures we discuss in the next several sections are generally applied to multispectral data sets. We defer discussion of hyperspectral image analysis until Section 7.18.

7.8 SUPERVISED CLASSIFICATION

We use a hypothetical example to facilitate our discussion of supervised classification. In this example, let us assume that we are dealing with the analysis of five-channel airborne multispectral scanner data. (The identical procedures would apply to Landsat, SPOT, or virtually any other source of multispectral data.) Figure 7.37 shows the location of a single line of data collected for our hypothetical example over a landscape composed of several cover types. For each of the pixels shown along this line, the multispectral scanner has measured scene radiance in terms of DNs recorded in each of the five spectral bands of sensing: blue, green, red, near infrared, and thermal infrared. Below the scan line, typical DNs measured over six different land cover types are shown. The vertical bars indicate the relative gray values in each spectral band. These five outputs represent a coarse description of the spectral response patterns of the various terrain features along the scan line. If these spectral patterns are sufficiently distinct for each feature type, they may form the basis for image classification.

Figure 7.38 summarizes the three basic steps involved in a typical supervised classification procedure. In the *training stage* (1), the analyst identifies representative training areas and develops a numerical description of the spectral attributes of each land cover type of interest in the scene. Next, in the *classification stage* (2), each pixel in the image data set is categorized into the land cover

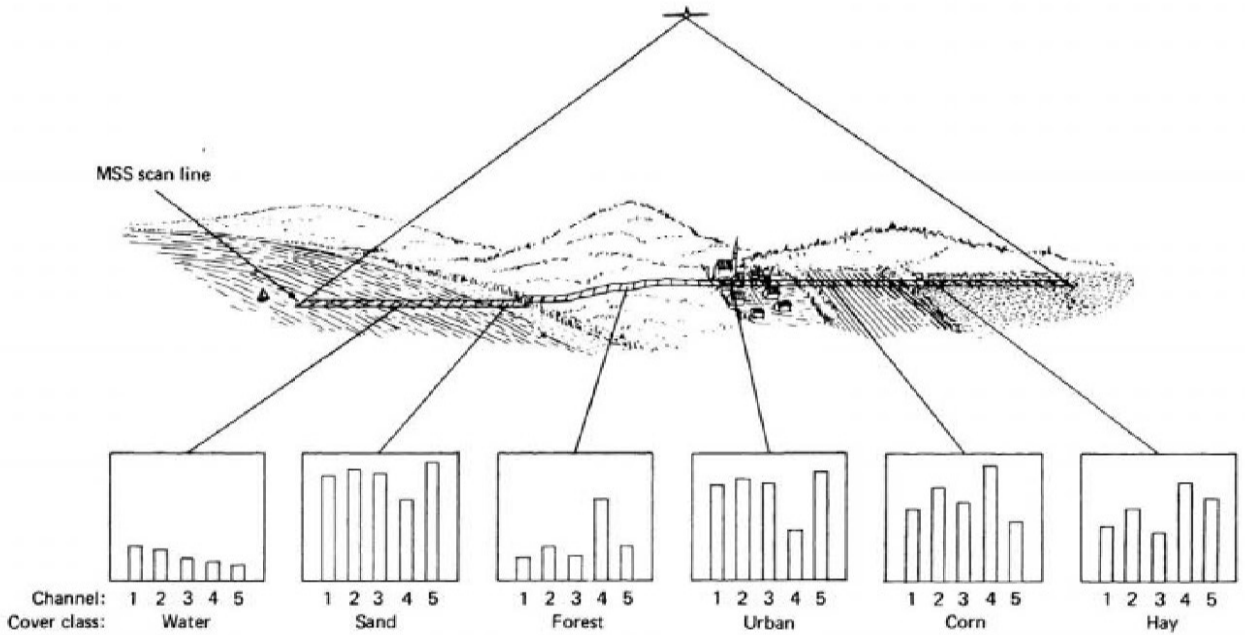

Figure 7.37 Selected multispectral scanner measurements made along one scan line. Sensor covers the following spectral bands: 1, blue; 2, green; 3, red; 4, near infrared; 5, thermal infrared.

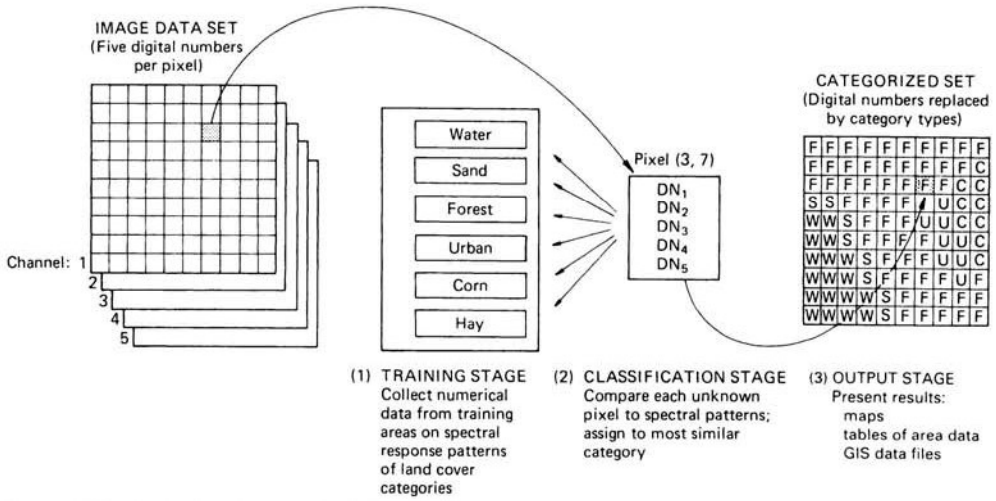

Figure 7.38 Basic steps in supervised classification.

class it most closely resembles. If the pixel is insufficiently similar to any training data set, it is usually labeled "unknown." The category label assigned to each pixel in this process is then recorded in the corresponding cell of an interpreted data set (an "output image"). Thus, the multidimensional image matrix is used to develop a corresponding matrix of interpreted land cover category types. After the entire data set has been categorized, the results are presented in the *output stage* (3). Being digital in character, the results may be used in a number of different ways. Three typical forms of output products are thematic maps, tables of full scene or subscene area statistics for the various land cover classes, and digital data files amenable to inclusion in a GIS. In this latter case, the classification "output" becomes a GIS "input."

We discuss the output stage of image classification in Section 7.14. Our immediate attention is focused on the training and classification stages. We begin with a discussion of the *classification* stage because it is the heart of the supervised classification process—during this stage the spectral patterns in the image data set are evaluated in the computer using predefined decision rules to determine the identity of each pixel. Another reason for treating the classification stage first is because familiarity with this step aids in understanding the requirements that must be met in the training stage.

7.9 THE CLASSIFICATION STAGE

Numerous mathematical approaches to spectral pattern recognition have been developed and extensive discussion of this subject can be found in the various

references found at the end of this chapter. Our discussion only "scratches the surface" of how spectral patterns may be classified into categories.

Our presentation of the various classification approaches is illustrated with a two-channel (bands 3 and 4) subset of our hypothetical five-channel multispectral scanner data set. Rarely are just two channels employed in an analysis, yet this limitation simplifies the graphic portrayal of the various techniques. When implemented numerically, these procedures may be applied to any number of channels of data.

Let us assume that we take a sample of pixel observations from our two-channel digital image data set. The two-dimensional digital values, or *measurement vectors*, attributed to each pixel may be expressed graphically by plotting them on a *scatter diagram* (or *scatter plot*), as shown in Figure 7.39. In this diagram, the band 3 DNs have been plotted on the y axis and the band 4 DNs on the x axis. These two DNs locate each pixel value in the two-dimensional "measurement space" of the graph. Thus, if the band 4 DN for a pixel is 10 and the band 3 DN for the same pixel is 68, the measurement vector for this pixel is represented by a point plotted at coordinate (10, 68) in the measurement space.

Let us also assume that the pixel observations shown in Figure 7.39 are from areas of known cover type (that is, from selected training sites). Each pixel

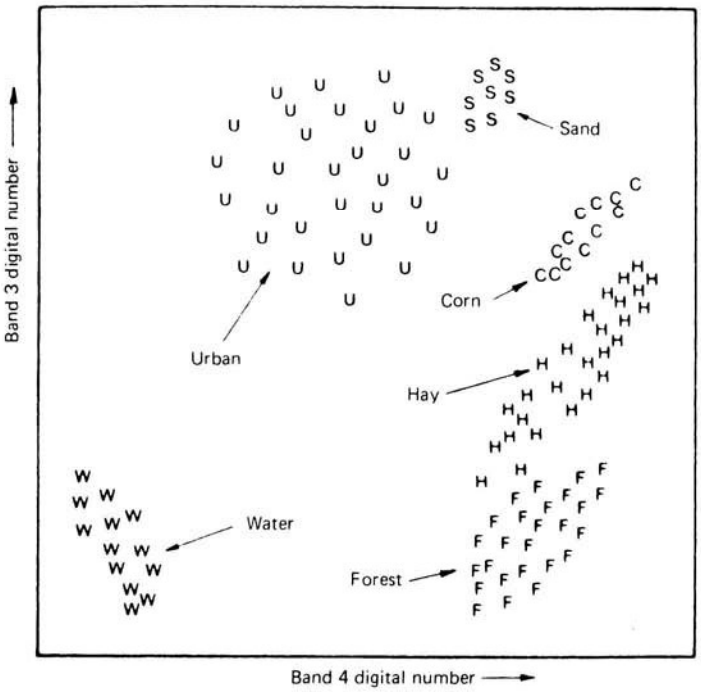

Figure 7.39 Pixel observations from selected training sites plotted on scatter diagram.

value has been plotted on the scatter diagram with a letter indicating the category to which it is known to belong. Note that the pixels within each class do not have a single, repeated spectral value. Rather, they illustrate the natural centralizing tendency—yet variability—of the spectral properties found within each cover class. These "clouds of points" represent multidimensional descriptions of the spectral response patterns of each category of cover type to be interpreted. The following classification strategies use these "training set" descriptions of the category spectral response patterns as interpretation keys by which pixels of unidentified cover type are categorized into their appropriate classes.

Minimum-Distance-to-Means Classifier

Figure 7.40 illustrates one of the simpler classification strategies that may be used. First, the mean, or average, spectral value in each band for each category is determined. These values comprise the *mean vector* for each category. The category means are indicated by +'s in Figure 7.40. By considering the two-channel pixel values as positional coordinates (as they are portrayed in the scatter diagram), a pixel of unknown identity may be classified by computing

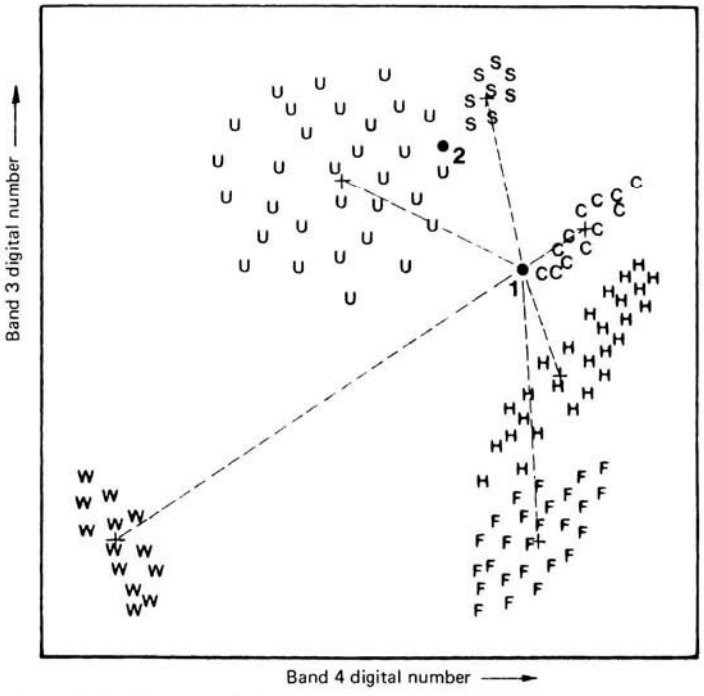

Figure 7.40 Minimum distance to means classification strategy.

the *distance* between the value of the unknown pixel and each of the category means. In Figure 7.40, an unknown pixel value has been plotted at point 1. The distance between this pixel value and each category mean value is illustrated by the dashed lines. After computing the distances, the unknown pixel is assigned to the "closest" class, in this case "corn." If the pixel is farther than an analyst-defined distance from any category mean, it would be classified as "unknown."

The minimum-distance-to-means strategy is mathematically simple and computationally efficient, but it has certain limitations. Most importantly, *it is insensitive to different degrees of variance in the spectral response data*. In Figure 7.40, the pixel value plotted at point 2 would be assigned by the distance-to-means classifier to the "sand" category, in spite of the fact that the greater variability in the "urban" category suggests that "urban" would be a more appropriate class assignment. Because of such problems, this classifier is not widely used in applications where spectral classes are close to one another in the measurement space and have high variance.

Parallelepiped Classifier

We can introduce sensitivity to category variance by considering the *range* of values in each category training set. This range may be defined by the highest and lowest digital number values in each band and appears as a rectangular area in our two-channel scatter diagram, as shown in Figure 7.41. An unknown pixel is classified according to the category range, or *decision region*, in which it lies or as "unknown" if it lies outside all regions. The multidimensional analogs of these rectangular areas are called *parallelepipeds*, and this classification strategy is referred to by that tongue-twisting name. The parallelepiped classifier is also very fast and efficient computationally.

The sensitivity of the parallelepiped classifier to category variance is exemplified by the smaller decision region defined for the highly repeatable "sand" category than for the more variable "urban" class. Because of this, pixel 2 would be appropriately classified as "urban." However, difficulties are encountered when category ranges overlap. Unknown pixel observations that occur in the overlap areas will be classified as "not sure" or be arbitrarily placed in one (or both) of the two overlapping classes. Overlap is caused largely because category distributions exhibiting *correlation* or high *covariance* are poorly described by the rectangular decision regions. Covariance is the tendency of spectral values to vary similarly in two bands, resulting in elongated, slanted clouds of observations on the scatter diagram. In our example, the "corn" and "hay" categories have positive covariance (they slant upward to the right), meaning that high values in band 3 are generally associated with high values in band 4, and low values in band 3 are associated with low values in band 4. The water category in our example exhibits *negative covariance* (its distribution slants down to the right), meaning that high values in band 3 are associated

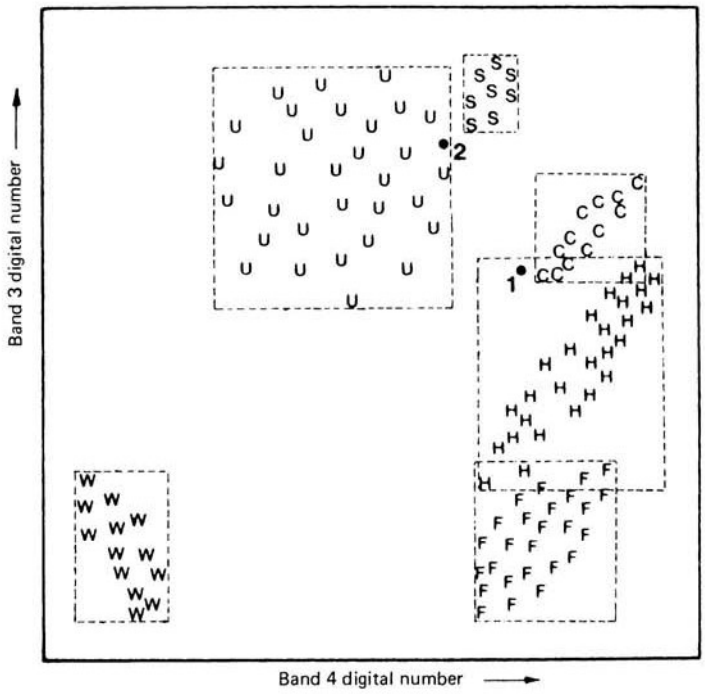

Figure 7.41 Parallelepiped classification strategy.

with low values in band 4. The "urban" class shows a lack of covariance, resulting in a nearly circular distribution on the scatter diagram.

In the presence of covariance, the rectangular decision regions fit the category training data very poorly, resulting in confusion for a parallelepiped classifier. For example, the insensitivity to covariance would cause pixel 1 to be classified as "hay" instead of "corn."

Unfortunately, spectral response patterns are frequently highly correlated, and high covariance is often the rule rather than the exception. The resulting problems can be somewhat alleviated within the parallelepiped classifier by modifying the single rectangles for the various decision regions into a series of rectangles with stepped borders. These borders then describe the boundaries of the elongated distributions more specifically. This approach is illustrated in Figure 7.42.

Gaussian Maximum Likelihood Classifier

The maximum likelihood classifier quantitatively evaluates both the variance and covariance of the category spectral response patterns when classifying an

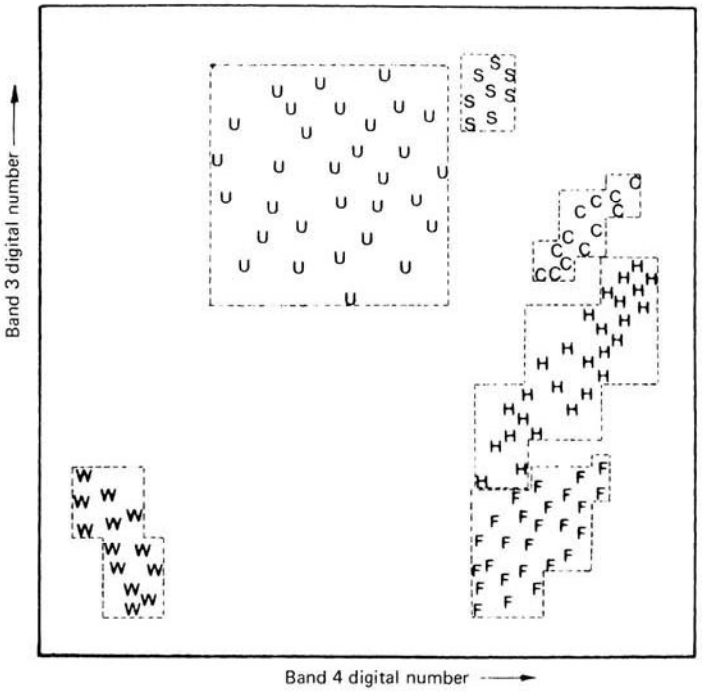

Figure 7.42 Parallelepiped classification strategy employing stepped decision region boundaries.

unknown pixel. To do this, an assumption is made that the distribution of the cloud of points forming the category training data is Gaussian (normally distributed). This *assumption of normality* is generally reasonable for common spectral response distributions. Under this assumption, the distribution of a category response pattern can be completely described by the *mean vector* and the *covariance matrix*. Given these parameters, we may compute the statistical probability of a given pixel value being a member of a particular land cover class. Figure 7.43 shows the probability values plotted in a three-dimensional graph. The vertical axis is associated with the probability of a pixel value being a member of one of the classes. The resulting bell-shaped surfaces are called *probability density functions*, and there is one such function for each spectral category.

The probability density functions are used to classify an unidentified pixel by computing the probability of the pixel value belonging to each category. That is, the computer would calculate the probability of the pixel value occurring in the distribution of class "corn," then the likelihood of its occurring in class "sand," and so on. After evaluating the probability in each category, the pixel would be assigned to the most likely class (highest probability value) or be labeled "unknown" if the probability values are all below a threshold set by the analyst.

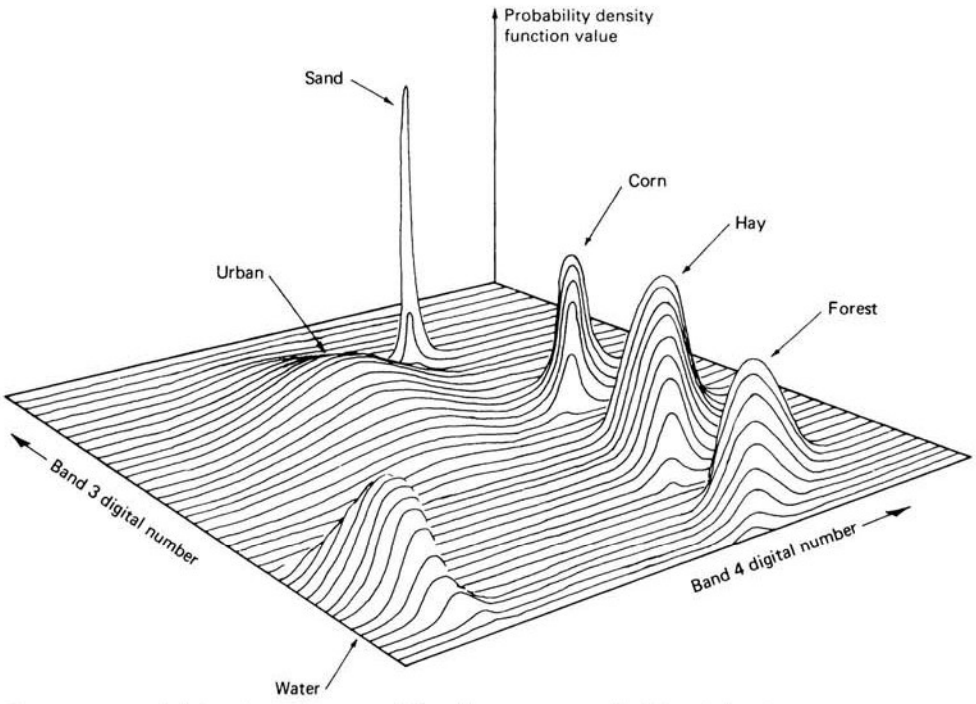

Figure 7.43 Probability density functions defined by a maximum likelihood classifier.

In essence, the maximum likelihood classifier delineates ellipsoidal "equiprobability contours" in the scatter diagram. These decision regions are shown in Figure 7.44. The shape of the equiprobability contours expresses the sensitivity of the likelihood classifier to covariance. For example, because of this sensitivity, it can be seen that pixel 1 would be appropriately assigned to the "corn" category.

An extension of the maximum likelihood approach is the *Bayesian classifier*. This technique applies two weighting factors to the probability estimate. First, the analyst determines the "a priori probability," or the anticipated likelihood of occurrence for each class in the given scene. For example, when classifying a pixel, the probability of the rarely occurring "sand" category might be weighted lightly, and the more likely "urban" class weighted heavily. Second, a weight associated with the "cost" of misclassification is applied to each class. Together, these factors act to minimize the "cost" of misclassifications, resulting in a theoretically optimum classification. In practice, most maximum likelihood classification is performed assuming equal probability of occurrence and cost of misclassification for all classes. If suitable data exist for these factors, the Bayesian implementation of the classifier is preferable.

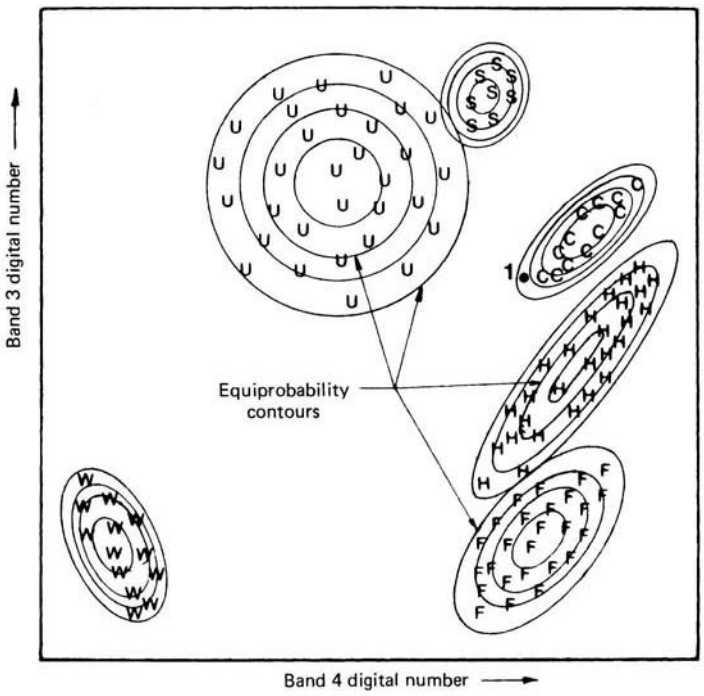

Figure 7.44 Equiprobability contours defined by a maximum likelihood classifier.

The principal drawback of maximum likelihood classification is the large number of computations required to classify each pixel. This is particularly true when either a large number of spectral channels are involved or a large number of spectral classes must be differentiated. In such cases, the maximum likelihood classifier is much slower computationally than the previous techniques.

Several approaches may be taken to increase the efficiency of maximum likelihood classifiers. In the lookup table implementation of such algorithms, the category identity for all possible combinations of digital numbers occurring in an image is determined in advance of actually classifying the image. Hence, the complex statistical computation for each combination is only made once. The categorization of each pixel in the image is then simply a matter of indexing the location of its multichannel gray level in the lookup table.

Another means of optimizing the implementation of the maximum likelihood classifier is to use some method to reduce the dimensionality of the data set used to perform the classification (thereby reducing the complexity of the required computations). As stated earlier, principal or canonical component transformations of the original data may be used for this purpose.

Decision tree, *stratified*, or *layered* classifiers have also been utilized to simplify classification computations and maintain classification accuracy. These classifiers are applied in a series of steps, with certain classes being separated during each step in the simplest manner possible. For example, water might first be separated from all other classes based on a simple threshold set in a near-infrared band. Certain other classes may require only two or three bands for categorization and a parallelepiped classifier may be adequate. The use of more bands or the maximum likelihood classifier would then only be required for those land cover categories where residual ambiguity exists between overlapping classes in the measurement space.

7.10 THE TRAINING STAGE

Whereas the actual classification of multispectral image data is a highly automated process, assembling the training data needed for classification is anything but automatic. In many ways, the training effort required in supervised classification is both an art and a science. It requires close interaction between the image analyst and the image data. It also requires substantial reference data and a thorough knowledge of the geographic area to which the data apply. Most importantly, the quality of the training process determines the success of the classification stage and, therefore, the value of the information generated from the entire classification effort.

The overall objective of the training process is to assemble a set of statistics that describe the spectral response pattern for each land cover type to be classified in an image. Relative to our earlier graphical example, it is during the training stage that the location, size, shape, and orientation of the "clouds of points" for each land cover class are determined.

To yield acceptable classification results, training data must be both representative and complete. This means that the image analyst must develop training statistics for all *spectral* classes constituting each *information* class to be discriminated by the classifier. For example, in a final classification output, one might wish to delineate an information class called "water." If the image under analysis contains only one water body and if it has uniform spectral response characteristics over its entire area, then only one training area would be needed to represent the water class. If, however, the same water body contained distinct areas of very clear water and very turbid water, a minimum of two spectral classes would be required to adequately train on this feature. If multiple water bodies occurred in the image, training statistics would be required for each of the other spectral classes that might be present in the water-covered areas. Accordingly, the single information class "water" might be represented by four or five spectral classes. In turn, the four or five spectral classes would eventually be used to classify all the water bodies occurring in the image.

By now it should be clear that the training process can become quite involved. For example, an information class such as "agriculture" might contain several crop types and each crop type might be represented by several spectral classes. These spectral classes could stem from different planting dates, soil moisture conditions, crop management practices, seed varieties, topographic settings, atmospheric conditions, or combinations of these factors. *The point that must be emphasized is that all spectral classes constituting each information class must be adequately represented in the training set statistics used to classify an image.* Depending upon the nature of the information classes sought, and the complexity of the geographic area under analysis, it is not uncommon to acquire data from 100 or more training areas to adequately represent the spectral variability in an image.

The location of training areas in an image is normally established by viewing *windows*, or portions of the full scene, in an enlarged format on an interactive color display device. The image analyst normally obtains training sample data by outlining training areas using a reference *cursor*. The cursor may be controlled by any of several means (e.g., a mouse, track ball, or joystick). Figure 7.45 shows the boundaries of several training site polygons that have been delineated in this manner. Note that these polygons have been carefully located to avoid pixels located along the edges between land cover types. The row and column coordinates of the vertices for these polygons are used as the basis for extracting (from the image file) the digital numbers for the pixels located within each training area boundary. These pixel values then form the sample used to develop the statistical description of each training area (mean vector and covariance matrix in the case of the maximum likelihood classifier).

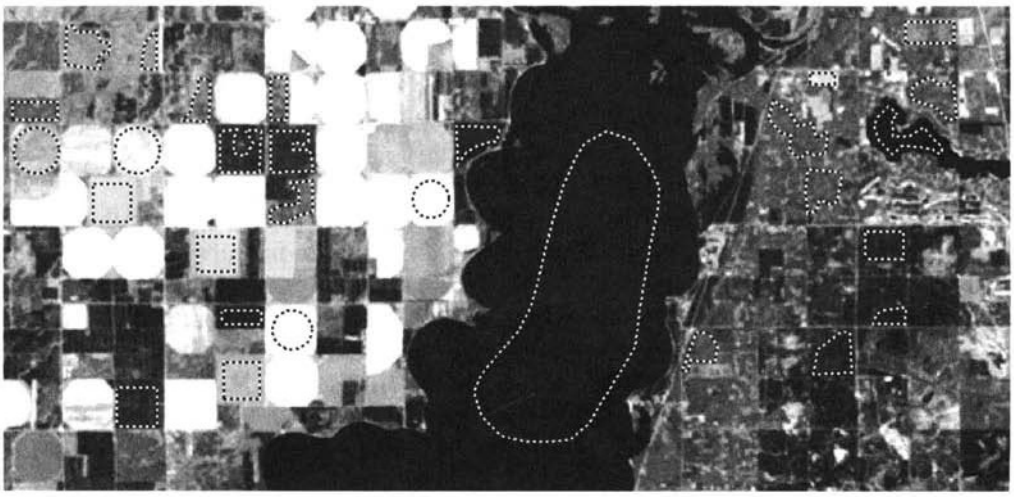

Figure 7.45 Training area polygons delineated on a computer monitor.

An alternative to manually delineating training area polygons is the use of a *seed pixel* approach to training. In this case, the display cursor is placed within a prospective training area and a single "seed" pixel is chosen that is thought to be representative of the surrounding area. Then, according to various statistically based criteria, pixels with similar spectral characteristics that are contiguous to the seed pixel are highlighted on the display and become the training samples for that training area.

Irrespective of how training areas are delineated, when using any statistically based classifier (such as the maximum likelihood method), the theoretical lower limit of the number of pixels that must be contained in a training set is $n + 1$, where n is the number of spectral bands. In our two-band example, *theoretically* only three observations would be required. Obviously, the use of fewer than three observations would make it impossible to appropriately evaluate the variance and covariance of the spectral response values. In practice, a minimum of from $10n$ to $100n$ pixels is used since the estimates of the mean vectors and covariance matrices improve as the number of pixels in the training sets increases. Within reason, the more pixels that can be used in training, the better the statistical representation of each spectral class.

When delineating training set pixels, it is important to analyze several training sites throughout the scene. For example, it would be better to define the training pattern for a given class by analyzing 20 locations containing 40 pixels of a given type than one location containing 800 pixels. Dispersion of the sites throughout the scene increases the chance that the training data will be representative of all the variations in the cover types present in the scene.

The trade-off usually faced in the development of training data sets is that of having sufficient sample size to ensure the accurate determination of the statistical parameters used by the classifier and to represent the total spectral variability in a scene, without going past a point of diminishing returns. In short, one does not want to omit any important spectral classes occurring in a scene, but one also does not want to include redundant spectral classes in the classification process from a computational standpoint. During the process of *training set refinement* the analyst attempts to identify such gaps and redundancies.

As part of the training set refinement process, the overall quality of the data contained in each of the original candidate training areas is assessed and the spectral separability between the data sets is studied. The analyst carefully checks to see if all data sets are essentially normally distributed and spectrally pure. Training areas that inadvertently include more than one spectral class are identified and recompiled. Likewise, extraneous pixels may be deleted from some of the data sets. These might be edge pixels along agricultural field boundaries or within-field pixels containing bare soil rather than the crop trained upon. Training sets that might be merged (or deleted) are

identified, and the need to obtain additional training sets for poorly represented spectral classes is addressed.

One or more of the following types of analyses are typically involved in the training set refinement process:

1. **Graphical representation of the spectral response patterns.** The distributions of training area response patterns can be graphically displayed in many formats. Figure 7.46 shows a hypothetical histogram for one of the "hay" category training sites in our five-channel multispectral scanner data set. (A similar display would be available for all training areas.) Histogram output is particularly important when a maximum likelihood classifier is used, since it provides a visual check on the normality of the spectral response distributions. Note in the case of the hay category that the data appear to be normally distributed in all bands except band 2, where the distribution is shown to be bimodal. This indicates that the training site data set chosen by the analyst to represent "hay" is in fact composed of two subclasses with slightly different spectral characteristics. These subclasses may represent two different varieties of hay or different illumination conditions, and so on. In any case, the classification accuracy will generally be improved if each of the subclasses is treated as a separate category.

   Histograms illustrate the distribution of individual categories very well; yet they do not facilitate comparisons between different category types. To evaluate the spectral separation between categories, it is convenient to use some form of *coincident spectral plot*, as shown in Figure 7.47. This plot illustrates, in each spectral band, the mean spectral response of each category (with a letter) and the variance of the distribution (± 2 standard deviations shown by black bars). Such plots indicate the overlap between category response patterns. For example, Figure 7.47 indicates that the hay and corn response patterns overlap in all spectral bands. The plot also shows which combination of bands might be best for discrimination because of relative reversals of spectral response (such as bands 3 and 5 for hay/corn separation).

   The fact that the spectral plots for hay and corn overlap in all spectral bands indicates that the categories could not be accurately classified on any *single* multispectral scanner band. However, this does not preclude successful classification when two or more bands are analyzed (such as bands 3 and 4 illustrated in the last section). Because of this, two-dimensional scatter diagrams (as shown in Figures 7.39 to 7.42) provide better representations of the spectral response pattern distributions.

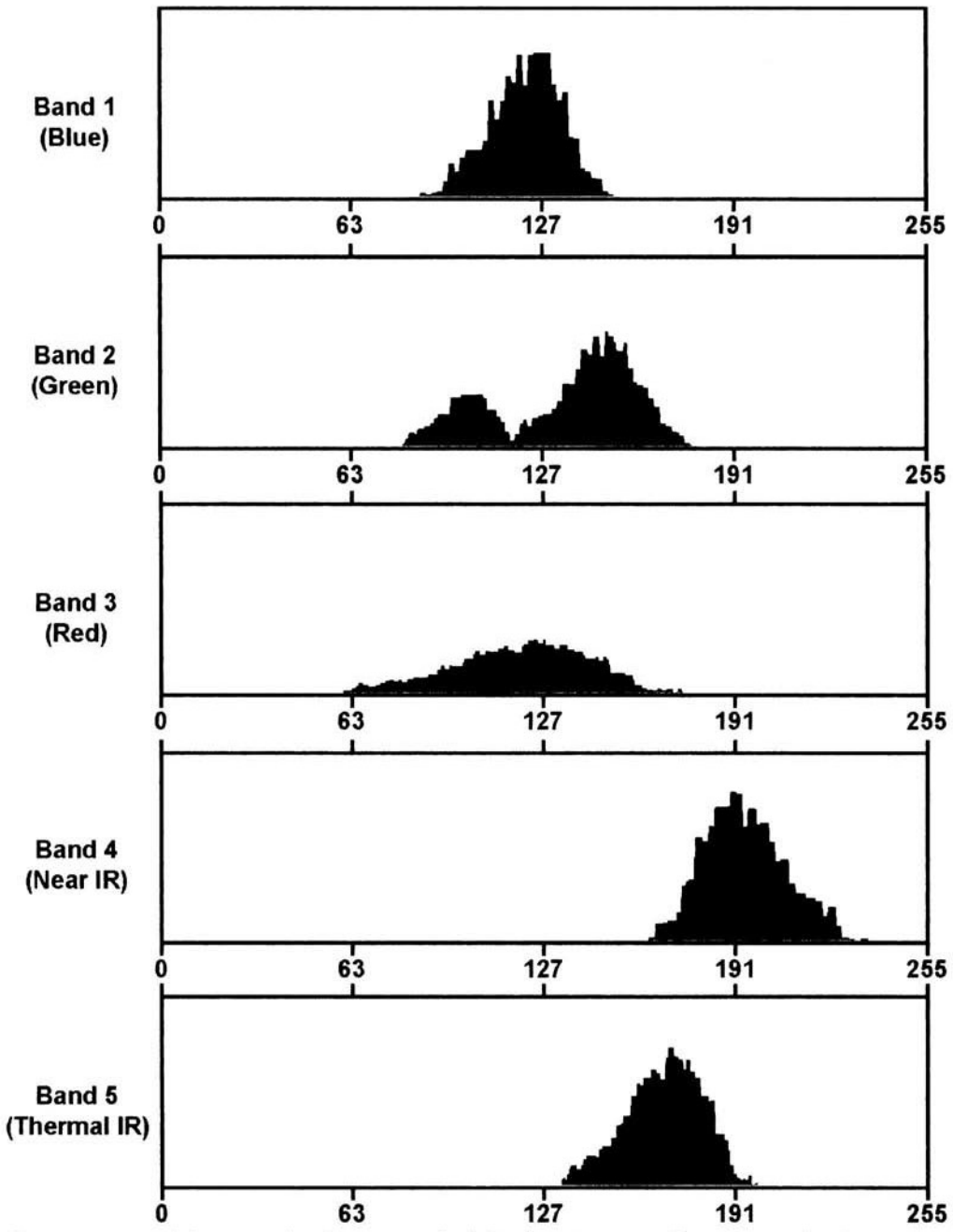

Figure 7.46 Sample histograms for data points included in the training areas for cover type "hay."

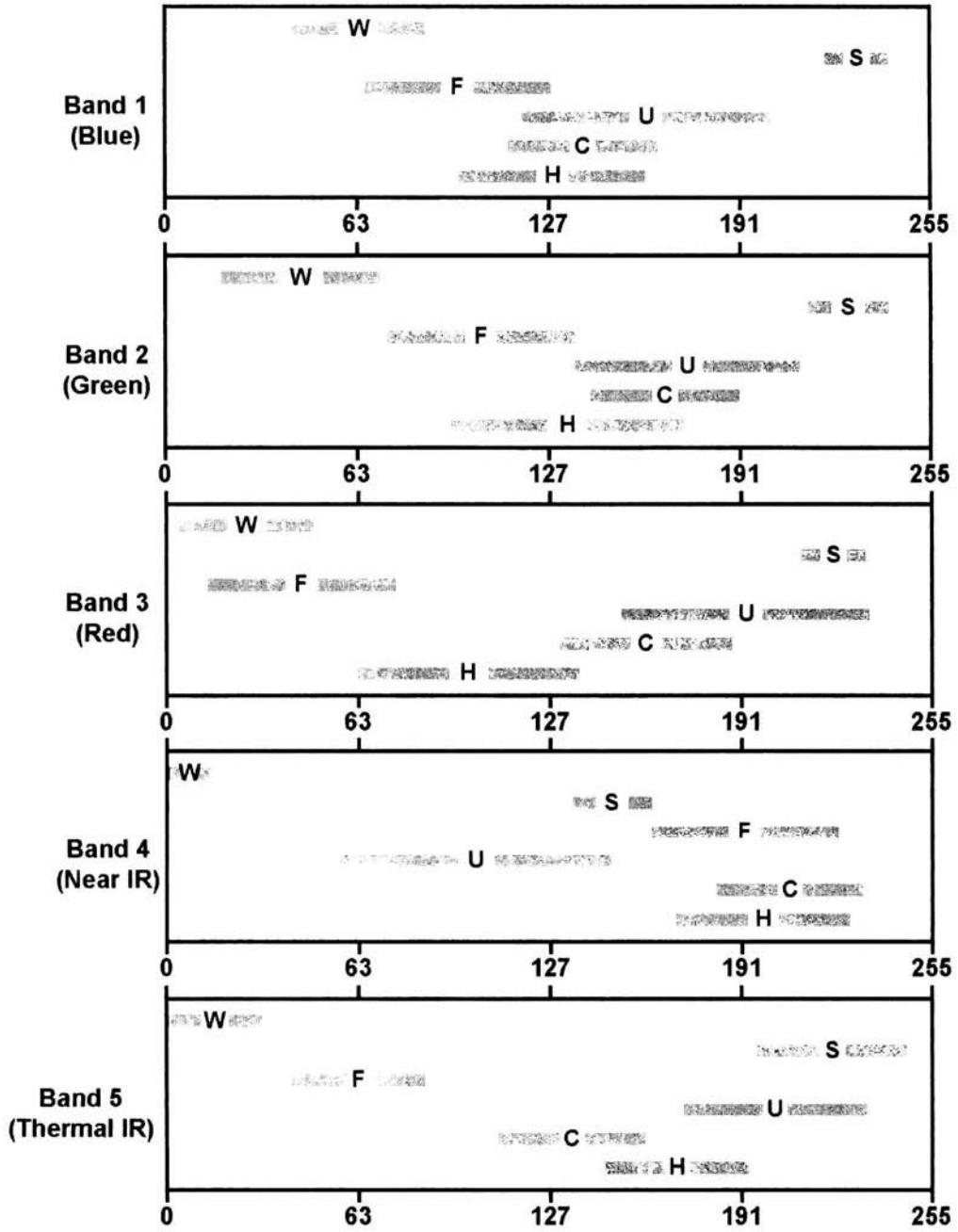

Figure 7.47 Coincident spectral plots for training data obtained in five bands for six cover types.

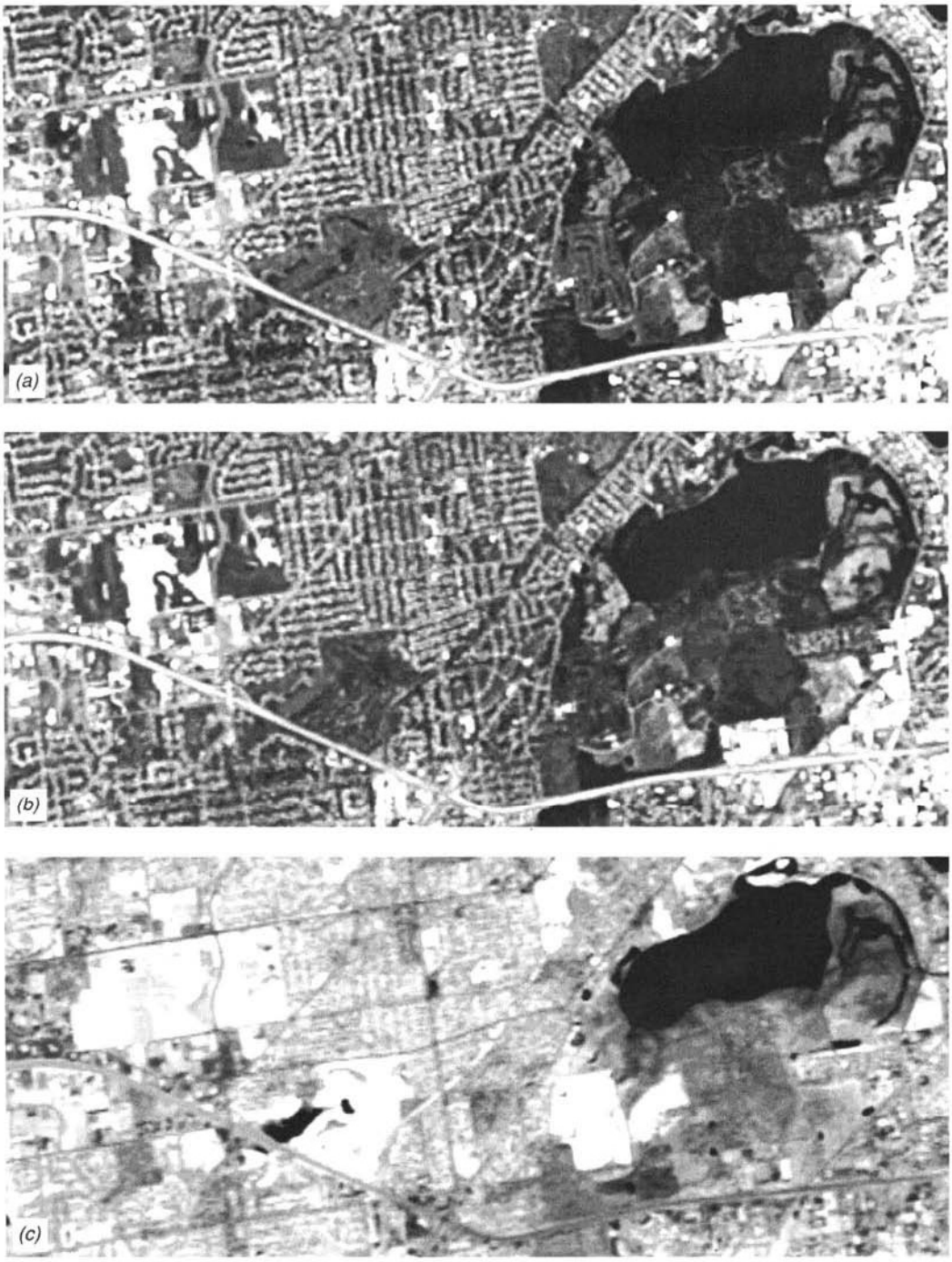

Figure 7.48 SPOT HRV multispectral images of Madison, WI: (*a*) band 1 (green); (*b*) band 2 (red); (*c*) band 3 (near IR).

The utility of scatter diagrams (or scatter plots) is further illustrated in Figures 7.48 to 7.50. Shown in Figure 7.48 are SPOT multispectral HRV images depicting a portion of Madison, Wisconsin. The band 1 (green), band 2 (red), and band 3 (near-IR) images are shown in (*a*), (*b*), and (*c*), respectively. Figure 7.49 shows the histograms for bands 1 and 2 as well as the associated scatter diagram for these two bands. Note that the data in these two bands are highly correlated and a very compact and near-linear "cloud of points" is shown in the scatter diagram.

Figure 7.50 shows the histograms and the scatter diagram for bands 2 and 3. In contrast to Figure 7.49, the scatter diagram in Figure 7.50

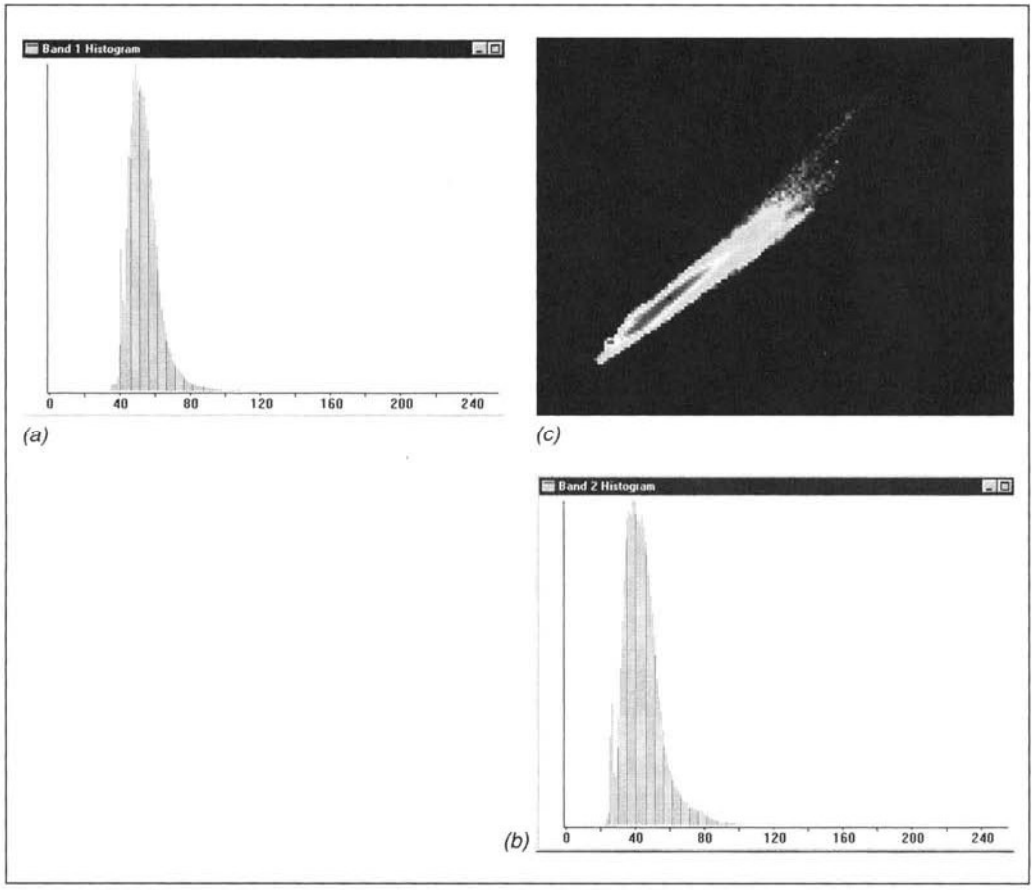

Figure 7.49 Histograms and two-dimensional scatter diagram for the images shown in Figures 7.48*a* and *b*: (*a*) band 1 (green) histogram; (*b*) band 2 (red) histogram; (*c*) scatter diagram plotting band 1 (vertical axis) versus band 2 (horizontal axis). Note the high correlation between these two visible bands.

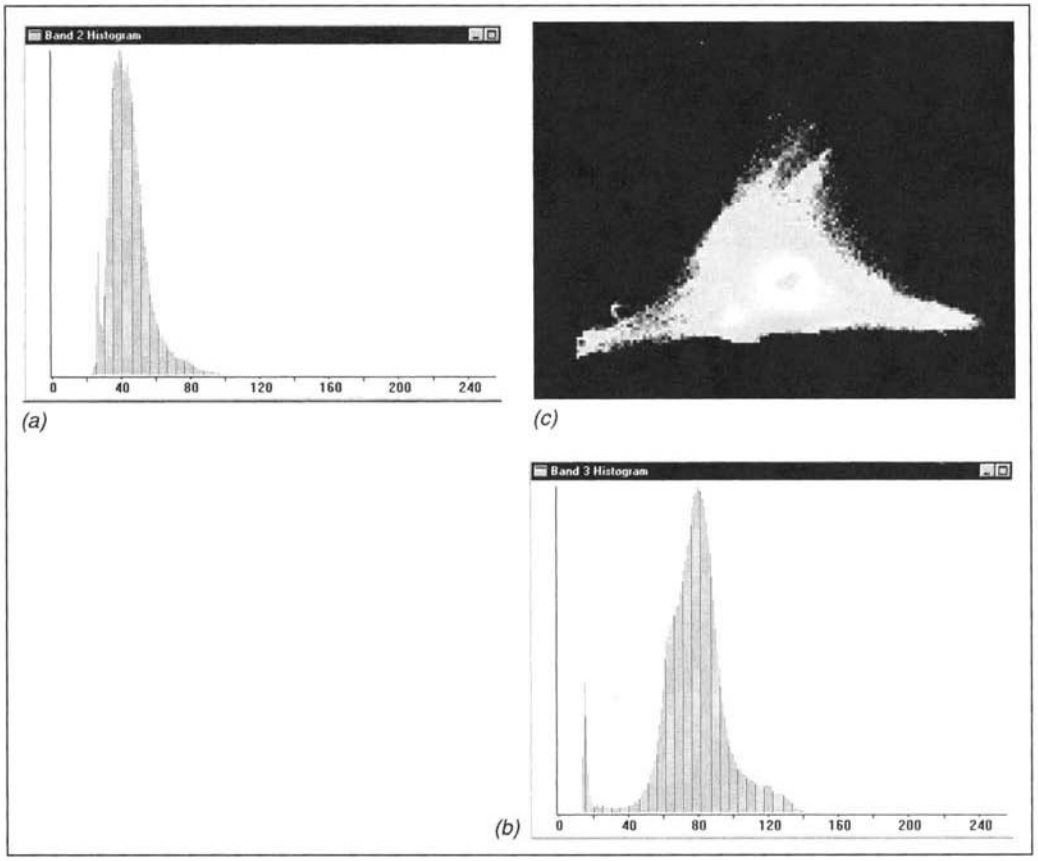

Figure 7.50 Histograms and two-dimensional scatter diagram for the images shown in Figures 7.48*b* and *c*: (*a*) band 2 (red) histogram; (*b*) band 3 (near-IR) histogram; (*c*) scatter diagram plotting band 2 (vertical axis) versus band 3 (horizontal axis). Note the relative lack of correlation between these visible and near-IR bands.

shows that bands 2 and 3 are much less correlated than bands 1 and 2. Whereas various land cover types might overlap one another in bands 1 and 2, they would be much more separable in bands 2 and 3. In fact, these two bands alone may be adequate to perform a generalized land cover classification of this scene.

2. **Quantitative expressions of category separation.** A measure of the statistical separation between category response patterns can be computed for all pairs of classes and can be presented in the form of a matrix. One statistical parameter commonly used for this purpose is *transformed divergence*, a covariance-weighted distance between category means. In general, the larger the transformed divergence, the greater the "statistical distance" between training patterns and the

TABLE 7.1 Portion of a Divergence Matrix Used to Evaluate Pairwise Training Class Spectral Separability

Spectral Class[a]	W1	W2	W3	C1	C2	C3	C4	H1	H2 ...
W1	0								
W2	1185	0							
W3	1410	680	0						
C1	1997	2000	1910	0					
C2	1953	1890	1874	860	0				
C3	1980	1953	1930	1340	1353	0			
C4	1992	1997	2000	1700	1810	1749	0		
H1	2000	1839	1911	1410	1123	860	1712	0	
H2	1995	1967	1935	1563	1602	1197	1621	721	0
⋮	⋮								

[a]W, water; C, corn; H, hay.

higher the probability of correct classification of classes. A portion of a sample matrix of divergence values is shown in Table 7.1. In this example, the maximum possible divergence value is 2000, and values less than 1500 indicate spectrally similar classes. Accordingly, the data in Table 7.1 suggest spectral overlap between several pairs of spectral classes. Note that W1, W2, and W3 are all relatively spectrally similar. However, note that this similarity is all among spectral classes from the same information class ("water"). Furthermore, all the "water" classes appear to be spectrally distinct from the spectral classes of the other information classes. More problematic is a situation typified by the divergence between the H1 and C3 spectral classes (860). Here, a "hay" spectral class severely overlaps a "corn" class.

[Another statistical distance measure of the separability of two spectral classes is the *Jeffries–Matusita (JM) distance*. It is similar to transformed divergence in its interpretation but has a maximum value of 1414.]

3. **Self-classification of training set data.** Another evaluation of spectral separability is provided by classifying the training set pixels. In such an effort, a preliminary classification of only the training set pixels (rather than the full scene) is made, to determine what percentage of the training pixels are actually classified as expected. These percentages are normally presented in the form of an *error matrix* (to be described in Section 7.16).

   It is important to avoid considering an error matrix based on training set values as a measure of *overall* classification accuracy throughout an image. For one reason, certain land cover classes might

be inadvertently missed in the training process. Also, the error matrix simply tells us how well the classifier can classify the *training areas* and nothing more. Because the training areas are usually good, homogeneous examples of each cover type, they can be expected to be classified more accurately than less pure examples that may be found elsewhere in the scene. Overall accuracy can be evaluated only by considering *test areas* that are different from and considerably more extensive than the training areas. This evaluation is generally performed after the classification and output stages (as discussed in Section 7.16).

4. **Interactive preliminary classification.** Most modern image processing systems incorporate some provision for interactively displaying how applicable training data are to the full scene to be classified. Often, this involves performing a preliminary classification with a computationally efficient algorithm (e.g., parallelepiped) to provide a visual approximation of the areas that would be classified with the statistics from a given training area. Such areas are typically highlighted in color on the display of the original raw image.

   This is illustrated in Plate 29, which shows a partially completed classification of a subset of the data included in Figure 7.50 (bands 2 and 3). Shown in (*a*) are selected training areas delineated on a color infrared composite of bands 1, 2, and 3 depicted as blue, green, and red, respectively. Part (*b*) shows the histograms and scatter plot for bands 2 and 3. Shown in (*c*) are the parallelepipeds associated with the initial training areas an image analyst has chosen to represent four information classes: water, trees, grass, and impervious surfaces. Part (*d*) shows how the statistics from these initial training areas would classify various portions of the original scene.

5. **Representative subscene classification.** Often, an image analyst will perform a classification of a representative subset of the full scene to eventually be classified. The results of this preliminary classification can then be used interactively on an overlay to the original raw image. Selected classes are then viewed individually or in logical groups to determine how they relate to the original image.

In general, the training set refinement process cannot be rushed with the "maximum efficiency" attitude appropriate in the classification stage. It is normally an iterative procedure in which the analyst revises the statistical descriptions of the category types until they are sufficiently spectrally separable. That is, the original set of "candidate" training area statistics is revised through merger, deletion, and addition to form the "final" set of statistics used in classification.

Training set refinement for the inexperienced data analyst is often a difficult task. Typically, an analyst has little difficulty in developing the statistics for the distinct "nonoverlapping" spectral classes present in a scene. If there

are problems, they typically stem from spectral classes on the borders between information classes—"transition" or "overlapping" classes. In such cases, the impact of alternative deletion and pooling of training classes can be tested by trial and error. In this process the sample size, spectral variances, normality, and identity of the training sets should be rechecked. Problem classes that occur only rarely in the image may be eliminated from the training data so that they are not confused with classes that occur extensively. That is, the analyst may accept misclassification of a class that occurs rarely in the scene in order to preserve the classification accuracy of a spectrally similar class that appears over extensive areas. Furthermore, a classification might initially be developed assuming a particular set of detailed information classes will be maintained. After studying the actual classification results, the image analyst might be faced with aggregating certain of the detailed classes into more general ones (for example, "birch" and "aspen" may have to be merged into a "deciduous" class or "corn" and "hay" into "agriculture").

One final note to be made here is that training set refinement is usually the key to improving the accuracy of a classification. However, if certain cover types occurring in an image have inherently similar spectral response patterns, no amount of retraining and refinement will make them spectrally separable! Alternative methods, such as using data resident in a GIS, performing a visual interpretation, or making a field check, must be used to discriminate these cover types. Multitemporal or spatial pattern recognition procedures may also be applicable in such cases. Increasingly, land cover classification involves some merger of remotely sensed data with ancillary information resident in a GIS.

7.11 UNSUPERVISED CLASSIFICATION

As previously discussed, unsupervised classifiers do *not* utilize training data as the basis for classification. Rather, this family of classifiers involves algorithms that examine the unknown pixels in an image and aggregate them into a number of classes based on the natural groupings or clusters present in the image values. The basic premise is that values within a given cover type should be close together in the measurement space, whereas data in different classes should be comparatively well separated.

The classes that result from unsupervised classification are *spectral classes*. Because they are based solely on the natural groupings in the image values, the identity of the spectral classes will not be initially known. The analyst must compare the classified data with some form of reference data (such as larger scale imagery or maps) to determine the identity and informational value of the spectral classes. Thus, in the *supervised* approach we define useful information categories and then examine their spectral separability; in the *unsupervised* approach we determine spectrally separable classes and then define their informational utility.

We illustrate the unsupervised approach by again considering a two-channel data set. Natural spectral groupings in the data can be visually identified by plotting a scatter diagram. For example, in Figure 7.51 we have plotted pixel values acquired over a forested area. Three groupings are apparent in the scatter diagram. After comparing the classified image data with ground reference data, we might find that one cluster corresponds to deciduous trees, one to conifers, and one to stressed trees of both types (indicated by D, C, and S in Figure 7.51). In a supervised approach, we may not have considered training for the "stressed" class. This highlights one of the primary advantages of unsupervised classification: The *classifier* identifies the distinct spectral classes present in the image data. Many of these classes might not be initially apparent to the analyst applying a supervised classifier. Likewise, the spectral classes in a scene may be so numerous that it would be difficult to train on all of them. In the unsupervised approach they are found automatically.

There are numerous *clustering* algorithms that can be used to determine the natural spectral groupings present in a data set. One common form of clustering, called the "K-means" approach, accepts from the analyst the number of clusters to be located in the data. The algorithm then arbitrarily "seeds," or locates, that number of cluster centers in the multidimensional measure-

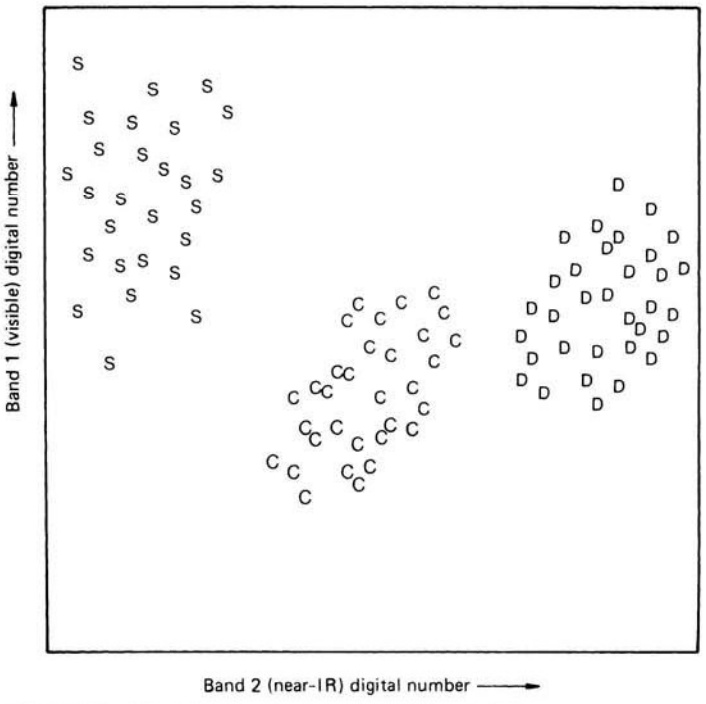

Figure 7.51 Spectral classes in two-channel image data.

ment space. Each pixel in the image is then assigned to the cluster whose arbitrary mean vector is closest. After all pixels have been classified in this manner, revised mean vectors for each of the clusters are computed. The revised means are then used as the basis to reclassify the image data. The procedure continues until there is no significant change in the location of class mean vectors between successive iterations of the algorithm. Once this point is reached, the analyst determines the land cover identity of each spectral class.

A widely used variant on the K-means method for unsupervised clustering is an algorithm called *Iterative Self-Organizing Data Analysis*, or *ISODATA* (Tou and Gonzalez, 1974). This algorithm permits the number of clusters to change from one iteration to the next, by merging, splitting, and deleting clusters. The general process follows that described above for K-means. However, in each iteration, following the assignment of pixels to the clusters, the statistics describing each cluster are evaluated. If the distance between the mean points of two clusters is less than some predefined minimum distance, the two clusters are merged together. On the other hand, if a single cluster has a standard deviation (in any one dimension) that is greater than a predefined maximum value, the cluster is split in two. Clusters with fewer than the specified minimum number of pixels are deleted. Finally, as with K-means, all pixels are then reclassified into the revised set of clusters, and the process repeats, until either there is no significant change in the cluster statistics or some maximum number of iterations is reached.

Another common approach to unsupervised classification is the use of algorithms that incorporate a sensitivity to image "texture" or "roughness" as a basis for establishing cluster centers. Texture is typically defined by the multidimensional variance observed in a moving window passed through the image (e.g., a 3×3 window). The analyst sets a variance threshold below which a window is considered "smooth" (homogeneous) and above which it is considered "rough" (heterogeneous). The mean of the first smooth window encountered in the image becomes the first cluster center. The mean of the second smooth window encountered becomes the second cluster center, and so forth. As soon as an analyst-specified maximum number of cluster centers is reached (e.g., 50), the classifier considers the distances between all previously defined cluster centers in the measurement space and merges the two closest clusters, combining their statistics. The classifier continues through the image combining the closest two clusters encountered until the entire image is analyzed. The resulting cluster centers are then analyzed to determine their separability on the basis of an analyst-specified statistical distance. Those clusters separated by less than this distance are combined and their statistics are merged. The final clusters resulting from the analysis are used to classify the image (e.g., with a minimum distance or maximum likelihood classifier).

Data from supervised training areas are sometimes used to augment the results of the above clustering procedure when certain land cover classes are poorly represented in the purely unsupervised analysis. (We discuss other such

hybrid approaches in Section 7.12.) Roads and other linear features, for example, may not be represented in the original clustering statistics if these features do not happen to meet the smoothness criteria within the moving window. Likewise, in some unsupervised classifiers the order in which different feature types are encountered can result in poor representation of some classes. For example, the analyst-specified maximum number of classes may be reached in an image long before the moving window passes throughout the scene.

Before ending our discussion of unsupervised classification, we reiterate that the result of such efforts is simply the identification of spectrally distinct classes in image data. The analyst must still use reference data to associate the spectral classes with the cover types of interest. This process, like the training set refinement step in supervised classification, can be quite involved.

Table 7.2 illustrates several possible outcomes of associating spectral classes with information classes for data from a scene covering a forested area. The ideal result would be outcome 1, in which each spectral class is found to be associated uniquely with a feature type of interest to the analyst.

TABLE 7.2 Spectral Classes Resulting from Clustering a Forested Scene

Spectral Class	Identity of Spectral Class	Corresponding Desired Information Category
Possible Outcome 1		
1	Water	Water
2	Coniferous trees	Coniferous trees
3	Deciduous trees	Deciduous trees
4	Brushland	Brushland
Possible Outcome 2		
1	Turbid water	Water
2	Clear water	
3	Sunlit conifers	Coniferous trees
4	Shaded hillside conifers	
5	Upland deciduous	Deciduous trees
6	Lowland deciduous	
7	Brushland	Brushland
Possible Outcome 3		
1	Turbid water	Water
2	Clear water	
3	Coniferous trees	Coniferous trees
4	Mixed coniferous/deciduous	
5	Deciduous trees	Deciduous trees
6	Deciduous/brushland	Brushland

This outcome will occur only when the features in the scene have highly distinctive spectral characteristics.

A more likely result is presented in outcome 2. Here, several spectral classes are attributable to each information category desired by the analyst. These "subclasses" may be of little informational utility (sunlit versus shaded conifers) or they may provide useful distinctions (turbid versus clear water and upland versus lowland deciduous). In either case, the spectral classes may be aggregated after classification into the smaller set of categories desired by the analyst.

Outcome 3 represents a more troublesome result in which the analyst finds that several spectral classes relate to more than one information category. For example, spectral class 4 was found to correspond to coniferous trees in some locations and deciduous trees in others. Likewise, class 6 included both deciduous trees and brushland vegetation. This means that these information categories are spectrally similar and cannot be differentiated in the given data set.

As with supervised classification, access to efficient hardware and software is an important factor in determining the *ease* with which an unsupervised classification can be performed. The *quality* of the classification still depends upon the analyst's understanding of the concepts behind the classifiers available and knowledge about the land cover types under analysis.

7.12 HYBRID CLASSIFICATION

Various forms of hybrid classification have been developed to either streamline or improve the accuracy of purely supervised or unsupervised procedures. For example, *unsupervised training areas* might be delineated in an image in order to aid the analyst in identifying the numerous spectral classes that need to be defined in order to adequately represent the land cover information classes to be differentiated in a supervised classification. Unsupervised training areas are image subareas chosen intentionally to be quite different from supervised training areas.

Whereas supervised training areas are located in regions of homogeneous cover type, the unsupervised training areas are chosen to contain numerous cover types at various locations throughout the scene. This ensures that all spectral classes in the scene are represented somewhere in the various subareas. These areas are then clustered independently and the spectral classes from the various areas are analyzed to determine their identity. They are subjected to a pooled statistical analysis to determine their spectral separability and normality. As appropriate, similar clusters representing similar land cover types are combined. Training statistics are developed for the combined classes and used to classify the entire scene (e.g., by a minimum distance or maximum likelihood algorithm).

Hybrid classifiers are particularly valuable in analyses where there is complex variability in the spectral response patterns for individual cover types present. These conditions are quite common in such applications as vegetation mapping. Under these conditions, spectral variability within cover types normally comes about both from variation within cover types per se (species) and from different site conditions (e.g., soils, slope, aspect, crown closure). *Guided clustering* is a hybrid approach that has been shown to be quite effective in such circumstances.

In guided clustering, the analyst delineates numerous "supervised-like" training sets for each cover type to be classified in a scene. Unlike the training sets used in traditional supervised methods, these areas need not be perfectly homogeneous. The data from all the training sites for a given information class are then used in an unsupervised clustering routine to generate several (as many as 20 or more) spectral signatures. These signatures are examined by the analyst; some may be discarded or merged and the remainder are considered to represent spectral subclasses of the desired information class. Signatures are also compared among the different information classes. Once a sufficient number of such spectral subclasses have been acquired for all information classes, a maximum likelihood classification is performed with the full set of refined spectral subclasses. The spectral subclasses are then aggregated back into the original information classes.

Guided clustering may be summarized in the following steps:

1. Delineate training areas for information class X.
2. Cluster all class X training area pixels at one time into spectral subclasses $X_1, \ldots, X_n$ using an automated clustering algorithm.
3. Examine class X signatures and merge or delete signatures as appropriate. A progression of clustering scenarios (e.g., from 3 to 20 cluster classes) should be investigated, with the final number of clusters and merger and deletion decisions based on such factors as (1) display of a given class on the raw image, (2) multidimensional histogram analysis for each cluster, and (3) multivariate distance measures (e.g., transformed divergence or JM distance).
4. Repeat steps 1 to 3 for all additional information classes.
5. Examine all class signatures and merge or delete signatures as appropriate.
6. Perform maximum likelihood classification on the entire image with the full set of spectral subclasses.
7. Aggregate spectral subclasses back to the original information classes.

Bauer et al. (1994) demonstrated the utility of guided clustering in the classification of forest types in northern Minnesota. Based on this success,

Lillesand et al. (1998) made the technique a central part of the Upper Midwest Gap Analysis Program image processing protocol development for classifying land cover throughout the states of Michigan, Minnesota, and Wisconsin. Given the extent and diversity of cover types in this area, the method was found to not only increase classification accuracy relative to either a conventional supervised or unsupervised approach but also increase the efficiency of the entire classification process. Among the advantages of this approach is its ability to help the analyst identify the various spectral subclasses representing an information class "automatically" through clustering. At the same time, the process of labeling the spectral clusters is straightforward because these are developed for one information class at a time. Also, spurious clusters due to such factors as including multiple-cover-type conditions in a single training area can be readily identified (e.g., openings containing understory vegetation in an otherwise closed forest canopy, bare soil in a portion of a crop-covered agricultural field). The method also helps identify situations where mixed pixels might inadvertently be included near the edges of training areas.

7.13 CLASSIFICATION OF MIXED PIXELS

As we have previously discussed (Sections 1.9 and 5.2), mixed pixels result when a sensor's IFOV includes more than one land cover type or feature on the ground. The extent to which mixed pixels are contained in an image is both a function of the spatial resolution of the remote sensing system used to acquire an image and the spatial scale of the surface features in question. For example, if a sensor with a narrow field of view is positioned within a few meters of a healthy soybean crop canopy, the sensor's field of view may be entirely covered by soybean leaves. A lower resolution sensor operating at higher altitude might focus on the same field yet have its field of view occupied by a mixture of soybean leaves, bare soil, and grass. These mixed pixels present a difficult problem for image classification, since their spectral characteristics are not representative of any single land cover type. *Spectral mixture analysis* and *fuzzy classification* are two procedures designed to deal with the classification of mixed pixels. They represent means by which "subpixel classification" is accomplished.

Spectral Mixture Analysis

Spectral mixture analysis involves a range of techniques wherein mixed spectral signatures are compared to a set of "pure" reference spectra (measured in the laboratory, in the field, or from the image itself). The basic assumption is that the spectral variation in an image is caused by mixtures of a limited

number of surface materials. The result is an estimate of the approximate proportions of the ground area of each pixel that are occupied by each of the reference classes.

Spectral mixture analysis differs in several ways from other image processing methods for land cover classification. Conceptually, it is a deterministic method rather than a statistical method, since it is based on a physical model of the mixture of discrete spectral response patterns. It provides useful information at the subpixel level, since multiple land cover types can be detected within a single pixel. Many land cover types tend to occur as heterogeneous mixtures even when viewed at very fine spatial scales; thus, this method provides a more realistic representation of the true nature of the surface than would be provided by the assignment of a single dominant class to every pixel.

Many applications of spectral mixture analysis make use of linear mixture models, in which the observed spectral response from an area on the ground is assumed to be a *linear* mixture of the individual spectral signatures of the various land cover types present within the area. These pure reference spectral signatures are referred to as *endmembers*, because they represent the cases where 100 percent of the sensor's field of view is occupied by a single cover type. In this model, the weight for any given endmember signature is the proportion of the area occupied by the class corresponding to the endmember. The input to a linear mixture model consists of a single observed spectral signature for each pixel in an image. The model's output then consists of "abundance" or "fraction" images for each endmember, showing the fraction of each pixel occupied by each endmember.

Linear mixture analysis involves the simultaneous satisfaction of two basic conditions for each pixel in an image. First, the sum of the fractional proportions of all potential endmembers included in a pixel must equal 1. Expressed mathematically,

$$\sum_{i=1}^{N} F_i = F_1 + F_2 + \cdots + F_N = 1 \tag{7.10}$$

where $F_1, F_2, \ldots, F_N$ represent the fraction of each of N possible endmembers contained in a pixel.

The second condition that must be met is that for a given spectral band λ the observed digital number DN_λ for each pixel represents the sum of the DNs that would be obtained from a pixel that is completely covered by a given endmember weighted by the fraction actually occupied by that endmember plus some unknown error. This can be expressed by

$$DN_\lambda = F_1 DN_{\lambda,1} + F_2 DN_{\lambda,2} + \cdots + F_N DN_{\lambda,N} + E_\lambda \tag{7.11}$$

where DN_λ is the composite digital number actually observed in band λ; $F_1, \ldots, F_N$ equal the fractions of the pixel actually occupied by each of the N endmembers; $DN_{\lambda,1}, \ldots, DN_{\lambda,N}$ equal the digital numbers that would be ob-

served if a pixel were completely covered by the corresponding endmember; and E_λ is the error term.

With multispectral data, there would be one version of Eq. 7.11 for each spectral band. So, for B spectral bands, there would be B equations, plus Eq. 7.10. This means that there are $B + 1$ equations available to solve for the various endmember fractions ($F_1, \ldots, F_N$). If the number of endmember fractions (unknowns) is equal to the number of spectral bands plus 1, the set of equations can be solved simultaneously to produce an exact solution without any error term. If the number of bands $B + 1$ is greater than the number of endmembers N, the magnitude of the error term along with the fractional cover for each endmember can be estimated (using the principles of least squares regression). On the other hand, if the number of endmember classes present in a scene exceeds $B + 1$, the set of equations will not yield a unique solution.

For example, a spectral mixture analysis of a four-band SPOT HRVIR multispectral image could be used to find estimates of the fractional proportions of five different endmember classes (with no estimate of the amount of error), or of four, three, or two endmember classes (in which case an estimate of the error would also be produced). Without additional information, this image alone could not be used in linear spectral mixture analysis to derive fractional cover estimates for more than five endmember classes.

Figure 7.52 shows an example of the output from a linear spectral mixture analysis project in which Landsat TM imagery was used to determine the fractional cover of trees, shrubs, and herbaceous plants in the Steese National Conservation Area of central Alaska. Figure 7.52*a* shows a single band (TM band 4, near IR), while 7.52*b* through 7.52*d* show the resulting output for each of the endmember classes. Note that these output images are scaled such that higher fractional cover values appear brighter while lower fractional cover values appear darker.

One drawback of linear mixture models is that they do not account for certain factors such as multiple reflections, which can result in complex nonlinearities in the spectral mixing process. That is, the observed signal from a pixel may include a mixture of spectral signatures from various endmembers, but it may also include additional radiance reflected multiple times between scene components such as leaves and the soil surface. In this situation, a more sophisticated *nonlinear* spectral mixture model may be required (Borel and Gerstl, 1994). Artificial neural networks (Section 7.17) may be particularly well suited for this task, because they do not require that the input data have a Gaussian distribution and they do not assume that spectra mix linearly (Moody et al., 1996).

Fuzzy Classification

Fuzzy classification attempts to handle the mixed-pixel problem by employing the fuzzy set concept, in which a given entity (a pixel) may have partial

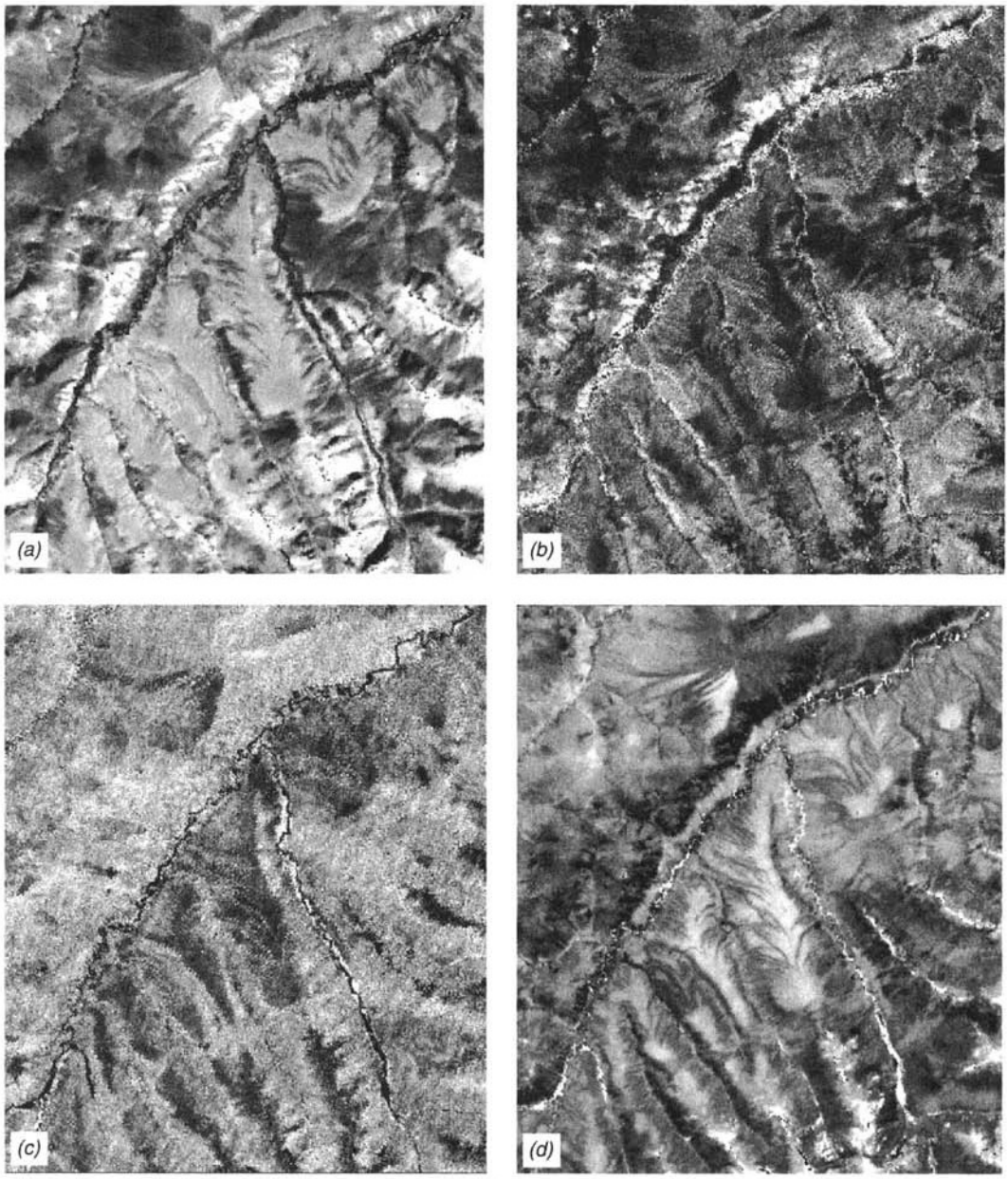

Figure 7.52 Linear spectral mixture analysis of a Landsat TM image including the Steese National Conservation Area of central Alaska: (*a*) band 4 (near IR) of original image; fractional cover images for trees (*b*), shrubs (*c*), and herbaceous plants (*d*). Brighter pixels represent higher fractional cover. (Courtesy Bureau of Land Management–Alaska and Ducks Unlimited, Inc.)

membership in more than one category (Jensen, 1996; Schowengerdt, 1997). One approach to fuzzy classification is *fuzzy clustering*. This procedure is conceptually similar to the K-means unsupervised classification approach described earlier. The difference is that instead of having "hard" boundaries between classes in the spectral measurement space, fuzzy regions are established. So instead of each unknown measurement vector being assigned solely to a single class, irrespective of how close that measurement may be to a partition in the measurement space, *membership grade* values are assigned that describe how close a pixel measurement is to the means of all classes.

Another approach to fuzzy classification is *fuzzy supervised* classification. This approach is similar to application of maximum likelihood classification; the difference being that fuzzy mean vectors and covariance matrices are developed from statistically weighted training data. Instead of delineating training areas that are purely homogeneous, a combination of pure and mixed training sites may be used. Known mixtures of various feature types define the fuzzy training class weights. A classified pixel is then assigned a membership grade with respect to its membership in each information class. For example, a vegetation classification might include a pixel with grades of 0.68 for class "forest," 0.29 for "street," and 0.03 for "grass." (Note that the grades for all potential classes must total 1.)

7.14 THE OUTPUT STAGE

The utility of any image classification is ultimately dependent on the production of output products that effectively convey the interpreted information to its end user. Here the boundaries between remote sensing, computer graphics, digital cartography, and GIS management become blurred. A virtually unlimited selection of output products may be generated. Three general forms that are commonly used include hardcopy graphic products, tables of area statistics, and digital data files.

Graphic Products

Since classified data are in the form of a two-dimensional data array, hardcopy graphic output can be easily computer generated by displaying different colors, tones, or characters for each cell in the array according to its assigned land cover category. A broad range of peripheral equipment can be used for this purpose, including a variety of black and white and color printers, film recorders, and large-format scanners/writers. Printouts can be prepared either in black and white or in color and can vary significantly in color fidelity and geometric precision.

Plate 30 shows a land cover classification of the states of New York and New Jersey that was derived from Landsat TM data. This color output product was prepared on a film recorder as part of a project at the USGS EROS Data Center that will create a generalized and consistent (i.e., "seamless") land cover data layer for the entire conterminous United States. Twenty-eight TM scenes were analyzed to create this classification, and 18 land cover classes are shown at this level of detail.

Tabular Data

Another common form of classification output is a table that lists summary statistics on the areal extent of the cover types present in a scene or in user-defined subscene areas. It is a simple task to derive area statistics from the grid-based interpreted data file. First, the boundary of a region of interest, such as a watershed or a county, is digitized in terms of its image matrix coordinates. Within the boundary, the number of cells in each land cover class is tabulated and multiplied by the ground area covered by a single cell. This process is considerably simpler than manually measuring areas on a map and represents a major advantage of processing land cover data in a digital format.

Digital Information Files

The final general class of output is interpreted data files containing the classification results recorded on some type of computer storage medium. As we illustrate in Section 7.17, the interpreted data in this form may be conveniently input to a GIS for merger with other geographic data files.

7.15 POSTCLASSIFICATION SMOOTHING

Classified data often manifest a salt-and-pepper appearance due to the inherent spectral variability encountered by a classifier when applied on a pixel-by-pixel basis (Figure 7.53*a*). For example, in an agricultural area, several pixels scattered throughout a corn field may be classified as soybeans, or vice versa. In such situations it is often desirable to "smooth" the classified output to show only the dominant (presumably correct) classification. Initially, one might consider the application of the previously described low pass spatial filters for this purpose. The problem with this approach is that the output from an image classification is an array of pixel locations containing numbers serving the function of *labels*, not *quantities*. That is, a pixel containing land cover 1 may be coded with a 1; a pixel containing land cover 2 may be

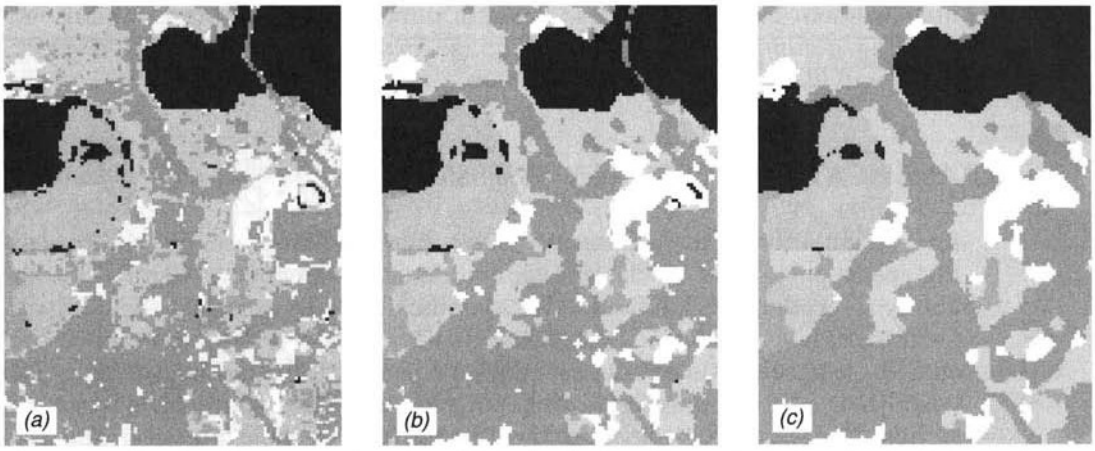

Figure 7.53 Postclassification smoothing: (*a*) original classification; (*b*) smoothed using a 3 × 3 pixel-majority filter; (*c*) smoothed using a 5 × 5-pixel majority filter.

coded with a 2; and so on. A moving low pass filter will not properly smooth such data because, for example, the averaging of class 3 and class 5 to arrive at class 4 makes no sense. In short, postclassification smoothing algorithms must operate on the basis of logical operations, rather than simple arithmetic computations.

One means of classification smoothing involves the application of a *majority filter*. In such operations a moving window is passed through the classified data set and the majority class within the window is determined. If the center pixel in the window is not the majority class, its identity is changed to the majority class. If there is no majority class in the window, the identity of the center pixel is not changed. As the window progresses through the data set, the original class codes are continually used, not the labels as modified from the previous window positions. (Figure 7.53*b* was prepared in this manner, applying a 3 × 3-pixel majority filter to the data shown in Figure7.53*a*. Figure 7.53*c* was prepared by applying a 5 × 5-pixel filter.)

Majority filters can also incorporate some form of class and/or spatial weighting function. Data may also be smoothed more than once. Certain algorithms can preserve the boundaries between land cover regions and also involve a user-specified minimum area of any given land cover type that will be maintained in the smoothed output.

One way of obtaining smoother classifications is to integrate the types of logical operations described above directly into the classification process. This involves the use of spatial pattern recognition techniques that are sensitive to such factors as image texture and pixel context. Compared to purely spectrally based procedures, these types of classifiers have received only limited attention in remote sensing in the past. However, with the continued

improvement in the spatial resolution of remote sensing systems and the increasing computational power of image processing systems, such procedures will likely become more common.

7.16 CLASSIFICATION ACCURACY ASSESSMENT

Another area that is continuing to receive increased attention by remote sensing specialists is that of classification accuracy assessment. Historically, the ability to produce digital land cover classifications far exceeded the ability to meaningfully quantify their accuracy. In fact, this problem sometimes precluded the application of automated land cover classification techniques even when their cost compared favorably with more traditional means of data collection. The lesson to be learned here is embodied in the expression "A classification is not complete until its accuracy is assessed."

Congalton and Green (1999) have prepared a thorough overview of the principles and practices currently in use for assessing classification accuracy. Many of the concepts we present here in brief are more fully described in this reference.

Classification Error Matrix

One of the most common means of expressing classification accuracy is the preparation of a classification *error matrix* (sometimes called a *confusion matrix* or a *contingency table*). Error matrices compare, on a category-by-category basis, the relationship between known reference data (ground truth) and the corresponding results of an automated classification. Such matrices are square, with the number of rows and columns equal to the number of categories whose classification accuracy is being assessed.

Table 7.3 is an error matrix that an image analyst has prepared to determine how well a classification has categorized a representative subset of pixels used in the training process of a supervised classification. This matrix stems from classifying the sampled training set pixels and listing the known cover types used for training (columns) versus the pixels actually classified into each land cover category by the classifier (rows).

Several characteristics about classification performance are expressed by an error matrix. For example, one can study the various classification errors of omission (exclusion) and commission (inclusion). Note in Table 7.3 that the training set pixels that are classified into the proper land cover categories are located along the major diagonal of the error matrix (running from upper left to lower right). All nondiagonal elements of the matrix represent errors of omission or commission. Omission errors correspond to nondiagonal column elements (e.g., 16 pixels that should have been classified as "sand" were omit-

TABLE 7.3 Error Matrix Resulting from Classifying Training Set Pixels

	Training Set Data (Known Cover Types)[a]						
	W	S	F	U	C	H	Row Total
Classification data							
W	480	0	5	0	0	0	485
S	0	52	0	20	0	0	72
F	0	0	313	40	0	0	353
U	0	16	0	126	0	0	142
C	0	0	0	38	342	79	459
H	0	0	38	24	60	359	481
Column total	480	68	356	248	402	438	1992

Producer's Accuracy	**User's Accuracy**
W = 480/480 = 100%	W = 480/485 = 99%
S = 052/068 = 76%	S = 052/072 = 72%
F = 313/356 = 88%	F = 313/353 = 87%
U = 126/248 = 51%	U = 126/142 = 89%
C = 342/402 = 85%	C = 342/459 = 74%
H = 359/438 = 82%	H = 359/481 = 75%

Overall accuracy = (480 + 52 + 313 + 126 + 342 + 359)/1992 = 84%

[a]W, water; S, sand; F, forest; U, urban; C, corn; H, hay.

ted from that category). Commission errors are represented by nondiagonal row elements (e.g., 38 "urban" pixels plus 79 "hay" pixels were improperly included in the "corn" category).

Several other descriptive measures can be obtained from the error matrix. For example, the *overall accuracy* is computed by dividing the total number of correctly classified pixels (i.e., the sum of the elements along the major diagonal) by the total number of reference pixels. Likewise, the accuracies of individual categories can be calculated by dividing the number of correctly classified pixels in each category by either the total number of pixels in the corresponding row or column. What are often termed *producer's accuracies* result from dividing the number of correctly classified pixels in each category (on the major diagonal) by the number of training set pixels used for that category (the column total). This figure indicates how well training set pixels of the given cover type are classified.

User's accuracies are computed by dividing the number of correctly classified pixels in each category by the total number of pixels that were classified in that category (the row total). This figure is a measure of commission error and indicates the probability that a pixel classified into a given category actually represents that category on the ground.

Note that the error matrix in Table 7.3 indicates an overall accuracy of 84 percent. However, producer's accuracies range from just 51 percent ("urban") to 100 percent ("water") and user's accuracies vary from 72 percent ("sand") to 99 percent ("water"). Furthermore, this error matrix is based on training data. *It should be remembered that such procedures only indicate how well the statistics extracted from these areas can be used to categorize the same areas!* If the results are good, it means nothing more than that the training areas are homogeneous, the training classes are spectrally separable, and the classification strategy being employed works well in the training areas. This aids in the training set refinement process, but it indicates little about how the classifier performs elsewhere in a scene. One should expect training area accuracies to be overly optimistic, especially if they are derived from limited data sets. (Nevertheless, training area accuracies are sometimes used in the literature as an indication of overall accuracy. They should not be!)

Sampling Considerations

Test areas are areas of representative, uniform land cover that are different from and considerably more extensive than training areas. They are often located during the training stage of supervised classification by intentionally designating more candidate training areas than are actually needed to develop the classification statistics. A subset of these may then be withheld for the postclassification accuracy assessment. The accuracies obtained in these areas represent at least a first approximation to classification performance throughout the scene. However, being homogeneous, test areas might not provide a valid indication of classification accuracy at the individual pixel level of land cover variability.

One way that would appear to ensure adequate accuracy assessment at the pixel level of specificity would be to compare the land cover classification at every pixel in an image with a reference source. While such "wall-to-wall" comparisons may have value in research situations, assembling reference land cover information for an entire project area is expensive and defeats the whole purpose of performing a remote-sensing-based classification in the first place.

Random sampling of pixels circumvents the above problems, but it is plagued with its own set of limitations. First, collection of reference data for a large sample of randomly distributed points is often very difficult and costly. For example, travel distance and access to random sites might be prohibitive. Second, the validity of random sampling depends on the ability to precisely register the reference data to the image data. This is often difficult to do. One way to overcome this problem is to sample only pixels whose identity is not influenced by potential registration errors (for example, points at least several pixels away from field boundaries).

Another consideration is making certain that the randomly selected test pixels or areas are geographically representative of the data set under analysis. Simple random sampling tends to undersample small but potentially important areas. Stratified random sampling, where each land cover category may be considered a stratum, is frequently used in such cases. Clearly, the sampling approach appropriate for an agricultural inventory would differ from that of a wetlands mapping activity. Each sample design must account for the area being studied and the cover type being classified.

One common means of accomplishing random sampling is to overlay classified output data with a grid. Test cells within the grid are then selected randomly and groups of pixels within the test cells are evaluated. The cover types present are determined through ground verification (or other reference data) and compared to the classification data.

Several papers have been written about the proper sampling scheme to be used for accuracy assessment under various conditions, and opinions vary among researchers. Several suggest the concept of combining both random and systematic sampling. Such a technique may use systematically sampled areas to collect some accuracy assessment data early in a project (perhaps as part of the training area selection process) and random sampling within strata after the classification is complete.

Consideration must also be given to the *sample unit* employed in accuracy assessment. Depending upon the application, the appropriate sample unit might be individual pixels, clusters of pixels, or polygons. Polygon sampling is the most common approach in current use.

Sample size must also weigh heavily in the development and interpretation of classification accuracy figures. Again, several researchers have published recommendations for choosing the appropriate sample size. However, these techniques primarily produce the sample size of test areas or pixels needed to compute the overall accuracy of a classification or of a single category. In general, they are not appropriate for filling in a classification error matrix wherein errors of omission and commission are of interest.

As a broad guideline, it has been suggested that a minimum of 50 samples of each vegetation or land use category be included in the error matrix. Further, "if the area is especially large (i.e., more than a million acres) or the classification has a large number of vegetation or land use categories (i.e., more than 12 categories), the minimum number of samples should be increased to 75 or 100 samples per category" (Congalton and Green, 1999, p. 18). Similarly, the number of samples for each category might be adjusted based on the relative importance of that category for a particular application (i.e., more samples taken in more important categories). Also, sampling might be allocated with respect to the variability within each category (i.e., more samples taken in more variable categories such as wetlands and fewer in less variable categories such as open water).

Evaluating Classification Error Matrices

Once accuracy data are collected (either in the form of pixels, clusters of pixels, or polygons) and summarized in an error matrix, they are normally subject to detailed interpretation and further statistical analysis. For example, a number of features are readily apparent from inspection of the error matrix included in Table 7.4 (resulting from randomly sampled test pixels). First, we can begin to appreciate the need for considering overall, producer's, and user's accuracies simultaneously. In this example, the overall accuracy of the classification is 65 percent. However, if the primary purpose of the classification is to map the locations of the "forest" category, we might note that the producer's accuracy of this class is quite good (84 percent). This would potentially lead one to the conclusion that although the overall accuracy of the classification was poor (65 percent), it is adequate for the purpose of mapping the forest class. The problem with this conclusion is the fact that the user's accuracy for this class is only 60 percent. That is, even though 84 percent of the forested areas have been correctly identified as "forest," only 60 percent of the areas identified as "forest" within the classification are truly of that category. A more careful inspection of

TABLE 7.4 Error Matrix Resulting from Classifying Randomly Sampled Test Pixels

	Reference Data[a]						
	W	S	F	U	C	H	Row Total
Classification data							
W	226	0	0	12	0	1	239
S	0	216	0	92	1	0	309
F	3	0	360	228	3	5	599
U	2	108	2	397	8	4	521
C	1	4	48	132	190	78	453
H	1	0	19	84	36	219	359
Column total	233	328	429	945	238	307	2480

PRODUCER'S ACCURACY	USER'S ACCURACY
W = 226/233 = 97%	W = 226/239 = 94%
S = 216/328 = 66%	S = 216/309 = 70%
F = 360/429 = 84%	F = 360/599 = 60%
U = 397/945 = 42%	U = 397/521 = 76%
C = 190/238 = 80%	C = 190/453 = 42%
H = 219/307 = 71%	H = 219/359 = 61%

Overall accuracy = (226 + 216 + 360 + 397 + 190 + 219)/2480 = 65%

[a]W, water; S, sand; F, forest; U, urban; C, corn; H, hay.

the error matrix shows that there is significant confusion between the "forest" and "urban" classes. Accordingly, although the producer of the classification can reasonably claim that 84 percent of the time an area that was forested was identified as such, a user of this classification would find that only 60 percent of the time will an area visited on the ground that the classification says is "forest" actually be "forest." In fact, the only highly reliable category associated with this classification from both a producer's and a user's perspective is "water."

A further point to be made about interpreting classification accuracies is the fact that even a completely random assignment of pixels to classes will produce percentage correct values in the error matrix. In fact, such a random assignment could result in a surprisingly good apparent classification result. The $\hat{k}$ ("*KHAT*") statistic is a measure of the difference between the actual agreement between reference data and an automated classifier and the chance agreement between the reference data and a random classifier. Conceptually, $\hat{k}$ can be defined as

$$\hat{k} = \frac{\text{observed accuracy} - \text{chance agreement}}{1 - \text{chance agreement}} \tag{7.12}$$

This statistic serves as an indicator of the extent to which the percentage correct values of an error matrix are due to "true" agreement versus "chance" agreement. As true agreement (observed) approaches 1 and chance agreement approaches 0, $\hat{k}$ approaches 1. This is the ideal case. In reality, $\hat{k}$ usually ranges between 0 and 1. For example, a $\hat{k}$ value of 0.67 can be thought of as an indication that an observed classification is 67 percent better than one resulting from chance. A $\hat{k}$ of 0 suggests that a given classification is no better than a random assignment of pixels. In cases where chance agreement is large enough, $\hat{k}$ can take on negative values—an indication of very poor classification performance. (Because the possible range of negative values depends on the specific matrix, the magnitude of negative values should not be interpreted as an indication of relative classification performance).

The KHAT statistic is computed as

$$\hat{k} = \frac{N \sum_{i=1}^{r} x_{ii} - \sum_{i=1}^{r} (x_{i+} \cdot x_{+i})}{N^2 - \sum_{i=1}^{r} (x_{i+} \cdot x_{+i})} \tag{7.13}$$

where

r = number of rows in the error matrix
x_{ii} = number of observations in row i and column i (on the major diagonal)
x_{i+} = total of observations in row i (shown as marginal total to right of the matrix)
x_{+i} = total of observations in column i (shown as marginal total at bottom of the matrix)
N = total number of observations included in matrix

To illustrate the computation of KHAT for the error matrix included in Table 7.4,

$$\sum_{i=1}^{r} x_{ii} = 226 + 216 + 360 + 397 + 190 + 219 = 1608$$

$$\sum_{i=1}^{r} (x_{i+} \cdot x_{+i}) = (239 \cdot 233) + (309 \cdot 328) + (599 \cdot 429) + (521 \cdot 945) + (453 \cdot 238) + (359 \cdot 307) = 1{,}124{,}382$$

$$\hat{K} = \frac{2480(1608) - 1{,}124{,}382}{(2480)^2 - 1{,}124{,}382} = 0.57$$

Note that the KHAT value (0.57) obtained in the above example is somewhat lower than the overall accuracy (0.65) computed earlier. Differences in these two measures are to be expected in that each incorporates different forms of information from the error matrix. The overall accuracy only includes the data along the major diagonal and excludes the errors of omission and commission. On the other hand, KHAT incorporates the nondiagonal elements of the error matrix as a product of the row and column marginal. Accordingly, it is not possible to give definitive advice as to when each measure should be used in any given application. Normally, it is desirable to compute and analyze both of these values.

One of the principal advantages of computing KHAT is the ability to use this value as a basis for determining the statistical significance of any given matrix or the differences among matrices. For example, one might wish to compare the error matrices resulting from different dates of images, classification techniques, or individuals performing the classification. Such tests are based on computing an estimate of the variance of $\hat{k}$ and then using a Z test to determine if an individual matrix is significantly different from a random result and if $\hat{k}$ values from two separate matrices are significantly different from one another. Readers interested in performing such analyses and learning more about accuracy assessment in general are urged to consult the various references on this subject in the Selected Bibliography.

There are three other facets of classification accuracy assessment that we wish to emphasize before leaving the subject. The first relates to the fact that the quality of any accuracy estimate is only as good as the information used to establish the "true" land cover types present in the test sites. To the extent possible, some estimate of the errors present in the reference data should be incorporated into the accuracy assessment process. It is not uncommon to have the accuracy of the reference data influenced by such factors as spatial misregistration, photo interpretation errors, data entry errors, and changes in land cover between the date of the classified image and the date of the reference data. The second point to be made is that the accuracy assessment procedure must be designed to reflect the intended use of the classification. For

example, a single pixel misclassified as "wetland" in the midst of a "corn" field might be of little significance in the development of a regional land use plan. However, this same error might be intolerable if the classification forms the basis for land taxation or for enforcement of wetland preservation legislation. Finally, it should be noted that remotely sensed data are normally just a small subset of many possible forms of data resident in a GIS. How errors accumulate through the multiple layers of information in a GIS is the subject of ongoing research.

7.17 DATA MERGING AND GIS INTEGRATION

Many applications of digital image processing are enhanced through the merger of multiple data sets covering the same geographical area. These data sets can be of a virtually unlimited variety of forms. Similarly, the merger of the data may or may not take place in the context of a GIS. For example, one frequently applied form of data merger is the combining of *multiresolution data* acquired by the same sensor. We earlier illustrated (Plate 19) the application of this procedure to IKONOS 1-m panchromatic and 4-m multispectral data.

Also, in Plate 1, we illustrated the merger of automated land cover classification data with soil erodibility and slope information in a GIS environment in order to assist in the process of soil erosion potential mapping. We also illustrated the use of raster remotely sensed imagery as a backdrop for vector overlay data (Figure 1.24). Thus, the reader should already have a basic appreciation for the diversity of forms of data and types of mergers that characterize current spatial analysis procedures. In the remainder of this section, we briefly discuss the following additional data merging operations: multitemporal data merging, change detection procedures, multisensor image merging, merging of image and ancillary data in the image display process, and incorporating GIS data in land cover classification.

We have made the above subdivisions of the topic of data merging and GIS integration purely for convenience in discussing the various procedures involved. As we will see, many of the operations we discuss are extensively used in combination with one another. Similarly, the boundaries between digital image processing and GIS operations have become blurred, and fully integrated spatial analysis systems have become the norm.

Multitemporal Data Merging

Multitemporal data merging can take on many different forms. One such operation is simply combining images of the same area taken on more than one date to create a product useful for visual interpretation. For example, agricultural crop interpretation is often facilitated through merger of images taken

early and late in the growing season. In early season images in the upper Midwest, bare soils often appear that later will probably be planted in such crops as corn or soybeans. At the same time, early season images might show perennial alfalfa or winter wheat in an advanced state of maturity. In the late season images, substantial changes in the appearance of the crops present in the scene are typical. Merging various combinations of bands from the two dates to create color composites can aid the interpreter in discriminating the various crop types present.

Plate 31 illustrates two examples of multitemporal NDVI data merging. Shown in (*a*) is the use of this technique to aid in mapping invasive plant species, in this case reed canary grass (*Phalaris arundinacea* L.). This part of the plate consists of a multitemporal color composite of NDVI values derived from Landsat-7 ETM+ images of southern Wisconsin on March 7 (blue), April 24 (green), and October 15 (red). The reed canary grass, which invades native wetland communities, tends to have a relatively high NDVI in the fall (October) compared to the native species and hence appears bright red to pink in the multitemporal composite. It can also be seen that features such as certain agricultural crops also manifest such tones. To eliminate the interpretation of such areas as "false positives" (identifying the areas as reed canary grass when they are not of this cover type) a GIS-derived wetland boundary layer (shown in yellow) has been overlain on the image. In this manner, the image analyst can readily focus solely upon those pink to red areas known to be included in wetlands.

Plate 31*b* represents a multitemporal color composite that depicts the three northernmost lakes shown in 31*a*. In this case, a slightly different set of dates and color assignments have been used to produce the color composite of NDVI values. These include April 24 (blue), October 31 (green), and October 15 (red). It so happened that the timing of the October 31 image corresponded with the occurrence of algal blooms in two of the three lakes shown. These blooms appear as the bright green features within the lakes.

As one would suspect, automated land cover classification is often enhanced through the use of multidate data sets. In fact, in many applications the use of multitemporal data is required to obtain satisfactory cover type discrimination. The extent to which use of multitemporal data improves classification accuracy and/or categorized detail is clearly a function of the particular cover types involved and both the number and timing of the various dates of imagery used.

Various strategies can be employed to combine multitemporal data in automatic land cover classification. One approach is to simply register all spectral bands from all dates of imaging into one master data set for classification. For example, the 6 reflectance (nonthermal) bands of a TM or ETM+ image from one date might be combined with the same 6 bands for an image acquired on another date, resulting in a 12-band data set to be used in the

classification. Alternatively, principal components analysis can be used to reduce the dimensionality of the combined data set prior to classification. For example, the first three principal components for each image could be computed separately and then merged to create a final 6-band data set for classification. The 6-band image can be stored, manipulated, and classified with much greater efficiency than the original 12-band image.

Another means of dealing with multitemporal data for crop classification is the *multitemporal profile* approach. In this approach, classification is based on physical modeling of the time behavior of each crop's spectral response pattern. It has been found that the time behavior of the greenness of annual crops is sigmoidal (Figure 7.54), whereas the greenness of the soils (G_0) in a given region is nearly constant. Thus, the greenness at any time t can be modeled in terms of the peak greenness G_m, the time of peak greenness t_p, and the width σ of the profile between its two inflection points. (The inflection points, t_1 and t_2, are related to the rates of change in greenness early in the growing season and at the onset of senescence.) The features G_m, t_p, and σ account for more than 95 percent of the information in the original data and can therefore be used for classification instead of the original spectral response patterns. These three features are important because they not only reduce the

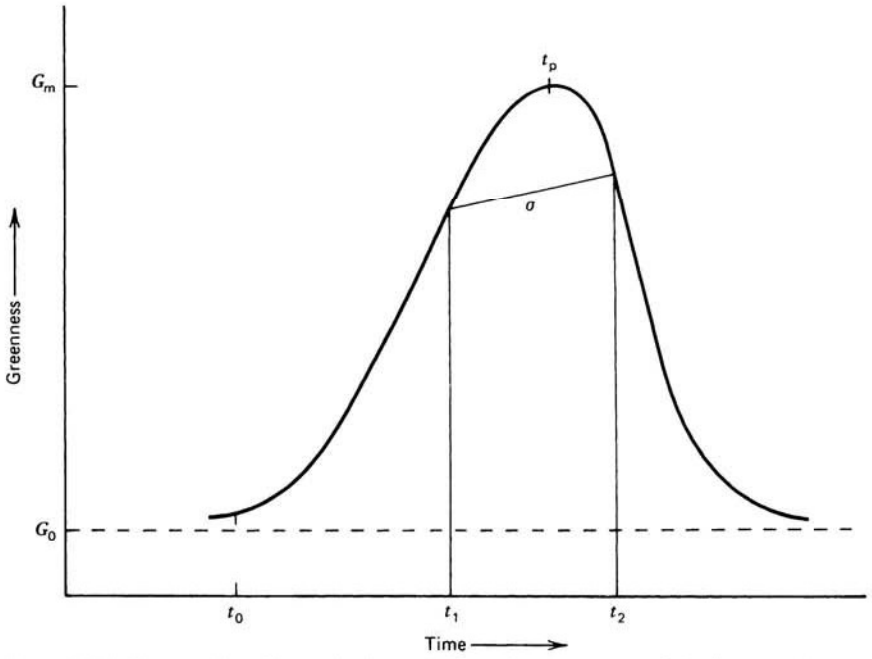

Figure 7.54 Temporal profile model for greenness. Key parameters include spectral emergence date (t_0), time (t_p) of peak greenness (G_m), and width of the profile (σ). (Adapted from Bauer, 1985, after Badhwar, 1985.)

dimensionality of the original data but also provide variables directly relatable to agrophysical parameters.

Change Detection Procedures

Change detection involves the use of multitemporal data sets to discriminate areas of land cover change between dates of imaging. The types of changes that might be of interest can range from short term phenomena such as snow cover or floodwater to long term phenomena such as urban fringe development or desertification. Ideally, change detection procedures should involve data acquired by the same (or similar) sensor and be recorded using the same spatial resolution, viewing geometry, spectral bands, radiometric resolution, and time of day. Often *anniversary dates* are used to minimize sun angle and seasonal differences. Accurate spatial registration of the various dates of imagery is also a requirement for effective change detection. Registration to within $\frac{1}{4}$ to $\frac{1}{2}$ pixel is generally required. Clearly, when misregistration is greater than one pixel, numerous errors will result when comparing the images.

The reliability of the change detection process may also be strongly influenced by various environmental factors that might change between image dates. In addition to atmospheric effects, such factors as lake level, tidal stage, wind, or soil moisture condition might also be important. Even with the use of anniversary dates of imagery, such influences as different planting dates and season-to-season changes in plant phenology must be considered.

One way of discriminating changes between two dates of imaging is to employ *postclassification comparison*. In this approach, two dates of imagery are independently classified and registered. Then an algorithm can be employed to determine those pixels with a change in classification between dates. In addition, statistics (and change maps) can be compiled to express the specific nature of the changes between the dates of imagery. Obviously, the accuracy of such procedures depends upon the accuracy of each of the independent classifications used in the analysis. The errors present in each of the initial classifications are compounded in the change detection process.

Another approach to change detection using spectral pattern recognition is simply the *classification of multitemporal data sets*. In this alternative, a single classification is performed on a combined data set for the two dates of interest. Supervised or unsupervised classification is used to categorize the land cover classes in the combined image. The success of such efforts depends upon the extent to which "change classes" are significantly different spectrally from the "nonchange" classes. Also, the dimensionality and complexity of the classification can be quite great, and if all bands from each date are used, there may be substantial redundancy in their information content.

Principal components analysis is sometimes used to analyze multidate image composites for change detection purposes. In this approach, two (or

more) images are registered to form a new multiband image containing various bands from each date. Several of the uncorrelated principal components computed from the combined data set can often be related to areas of change. One disadvantage to this process is that it is often difficult to interpret and identify the specific nature of the changes involved.

Plate 32 illustrates the application of multidate principal components analysis to the process of assessing tornado damage from "before" and "after" images of the tornado's path of destruction. The "before" image shown in (*a*), is a Landsat-7 ETM+ composite of bands, 1, 2, and 5 shown as blue, green, and red, respectively. The "after" image, shown in (*b*), was acquired 32 days later than the image shown in (*a*), on the day immediately following the tornado. While the damage path from the tornado is fairly discernable in (*b*), it is most distinct in the principal component image shown in (*c*). This image depicts the second principal component image computed from the 6-band composite formed by registering bands, 1, 2, and 5 from both the "before" and "after" images.

Temporal image differencing is yet another common approach to change detection. In the image differencing procedure, DNs from one date are simply subtracted from those of the other. The difference in areas of no change will be very small (approaching zero), and areas of change will manifest larger negative or positive values. If 8-bit images are used, the possible range of values for the difference image is −255 to +255, so normally a constant (e.g., 255) is added to each difference image value for display purposes.

Temporal image ratioing involves computing the ratio of the data from two dates of imaging. Ratios for areas of no change tend toward 1 and areas of change will have higher or lower ratio values. Again, the ratioed data are normally scaled for display purposes (Section 7.6). One of the advantages to the ratioing technique is that it tends to normalize the data for changes in such extraneous factors as sun angle and shadows.

Whether image differencing or ratioing is employed, the analyst must find a meaningful "change–no change threshold" within the data. This can be done by compiling a histogram for the differenced or ratioed image data and noting that the change areas will reside within the tails of the distribution. A variance from the mean can then be chosen and tested empirically to determine if it represents a reasonable threshold. The threshold can also be varied interactively in most image analysis systems so the analyst can obtain immediate visual feedback on the suitability of a given threshold.

In lieu of using raw DNs to prepare temporal difference or ratio images, it is often desirable to correct for illumination and atmospheric effects and to transform the image data into physically meaningful quantities such as radiances or reflectances (Section 7.2). Also, the images may be prepared using spatial filtering or transformations such as principal components or vegetation components. Likewise, linear regression procedures may be used to compare the two dates of imagery. In this approach a linear regression model is

applied to predict data values for date 2 based on those of date 1. Again, the analyst must set a threshold for detecting meaningful change in land cover between the dates of imaging.

Change vector analysis is a change detection procedure that is a conceptual extension of image differencing. Figure 7.55 illustrates the basis for this approach in two dimensions. Two spectral variables (e.g., data from two bands, two vegetation components) are plotted at dates 1 and 2 for a given pixel. The vector connecting these two data sets describes both the magnitude and direction of spectral change between dates. A threshold on the magnitude can be established as the basis for determining areas of change, and the direction of the spectral change vector often relates to the type of change. For example, Figure 7.55*b* illustrates the differing directions of the spectral change vector for vegetated areas that have been recently cleared versus those that have experienced regrowth between images.

One of the more efficient approaches to delineating change in multidate imagery is use of a *change-versus-no-change binary mask to guide multidate classification*. This method begins with a traditional classification of one image as a reference (time 1). Then, one of the spectral bands from this date is registered to the same band in a second date (time 2). This two-band data set is then analyzed using one of the earlier described algebraic operations (e.g., image differencing or ratioing). A threshold is then set to separate areas that have changed between dates from those that have not. This forms the basis for creating a binary mask of change-versus-no-change areas. This mask is then applied to the multiband image acquired at time 2 and only the areas of change are then classified for time 2. A traditional postclassifica-

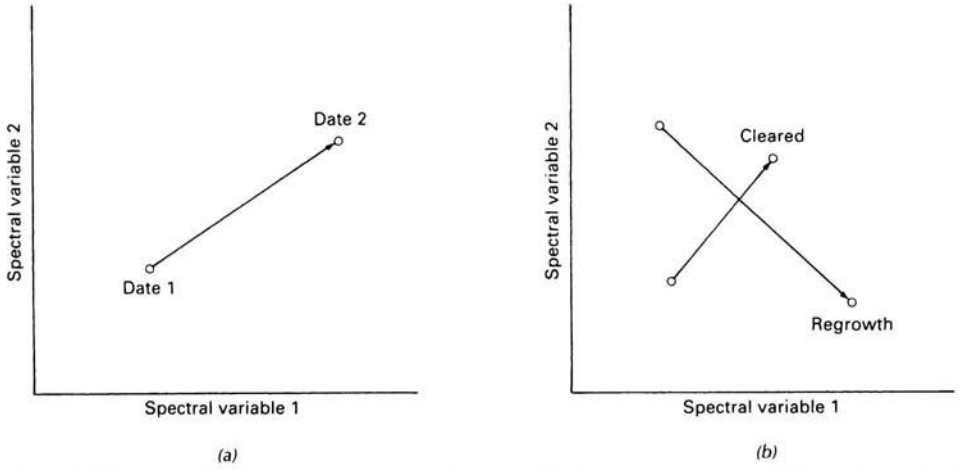

Figure 7.55 Spectral change vector analysis: (*a*) spectral change vector observed for a single land cover type; (*b*) length and direction of spectral change vectors for hypothetical "cleared" and "regrowth" areas.

tion comparison is then performed in the areas known to have changed between dates.

Research continues on the development of software tools to assist in multidate change detection. Representative of such tools is a technique proposed by Walkey (1997) that consists of a set of interactive, iterative steps to aid an analyst in delineating true land cover changes from incidental scene-to-scene changes. The technique is based on the simple notion that a two-dimensional scatter plot of a band at time 1 versus the same band at time 2 would result in an elongated ellipse oriented at 45° in the spectral measurement space (Figure 7.56*b*). The ellipse, rather than a straight line (Figure 7.56*a*), results from the natural variability of the landscape even in the absence of land cover change. Also, extraneous effects such as changing atmospheric conditions or sensor response drift can cause the "stable spectral space ellipse" to depart from an exact 45° orientation (Figures 7.56*c* and *d*).

Figure 7.56*e* illustrates a situation where certain features in the two-date data set have experienced a land cover change. The small ellipse above the stable spectral space ellipse represents those pixels that have gotten brighter between the two dates, and the ellipse below the stable spectral space ellipse corresponds to those areas that have gotten darker between dates. By displaying such a scatter plot, it is possible to interactively describe new *delta transformation* axes with a transformed band 1 aligned in the direction of the stable spectral space ellipse and a transformed band 2 located at 90° from band 1 (Figure 7.56*f*). The transformed band 2 axis then defines a change axis that the analyst can use to set a pair of change-versus-no-change thresholds—one in the positive (brighter) direction and one in the negative (darker) direction. Hence, the analyst can then display separately images of the positive-change component, the negative-change component, and the stable component.

Delta transformations can be established for each spectral band available, such that the two change images plus the stable component of each band pair can be generated. The options for display of the images are numerous. For example, the various change and stable components can be displayed individually in black and white or in the blue, green, and red planes of a color monitor in various combinations.

It should also be emphasized that the scatter plots shown in Figure 7.56 have been greatly simplified in order to describe the delta transformation process. The scatter plots for actual image data can be more complex and the analyst must be very careful in delineating the direction of "stable spectral space." Again, this can be done interactively and iteratively until an acceptable change-versus-no-change axis is determined.

Figure 7.57 illustrates the application of the delta transformation process to images acquired over a portion of the Nicolet National Forest in northeastern Wisconsin. Shown in (*a*) is a Landsat TM band 5 (mid-IR) image acquired in 1984, and its counterpart acquired in 1993 is shown in (*b*). Note that between dates several areas of the forest have become brighter due to

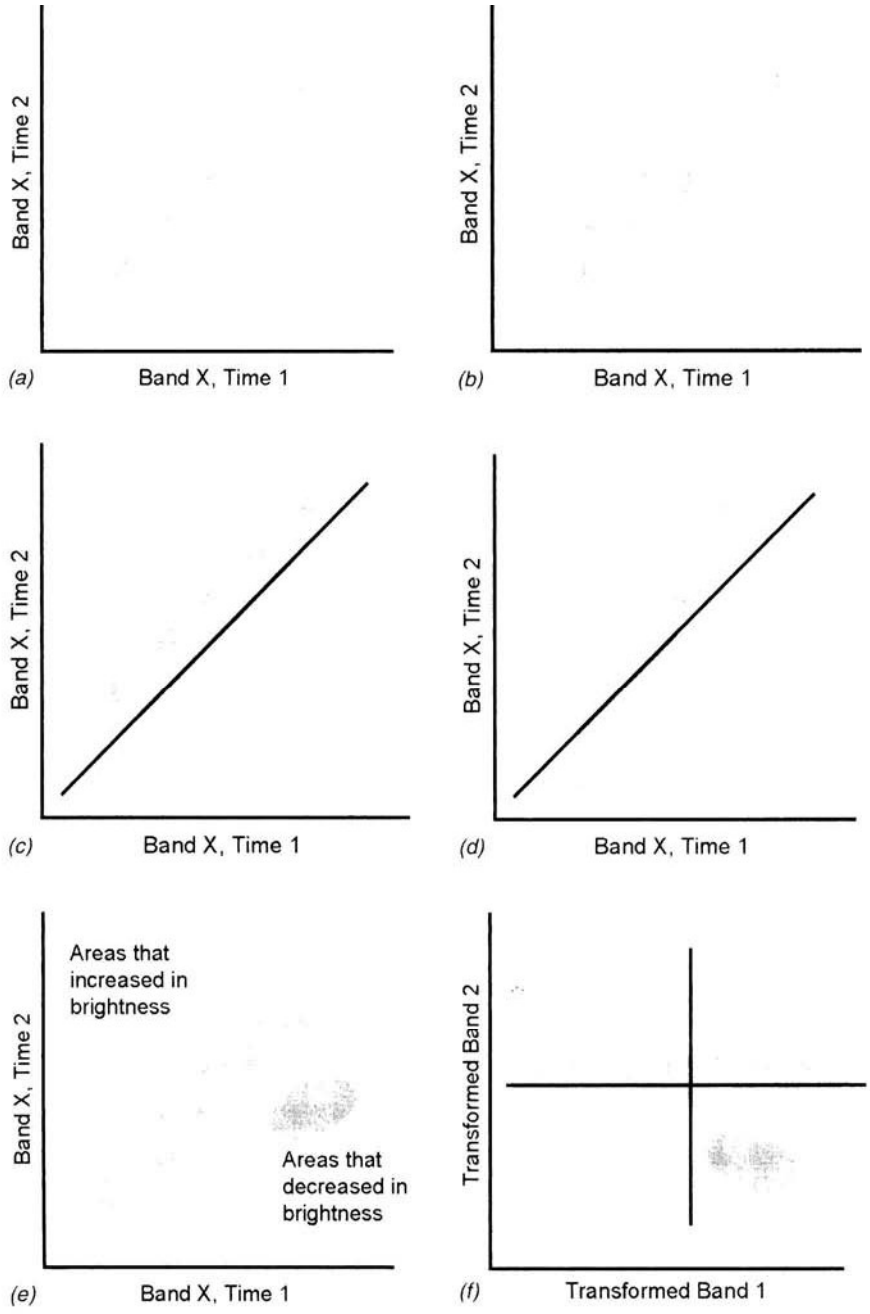

Figure 7.56 Conceptual basis for the delta transformation: (*a*) plot of a band vs. itself with no spectral change whatsoever; (*b*) stable spectral space ellipse showing natural variability in the landscape between dates; (*c*) effect of uniform atmospheric haze differences between dates; (*d*) effect of sensor drift between dates; (*e*) pixels that appear brighter (above) or darker (below) than the stable spectral space ellipse due to land cover change; (*f*) delta transformation with transformed band 2 defining the image change axis. (After Walkey, 1997.)

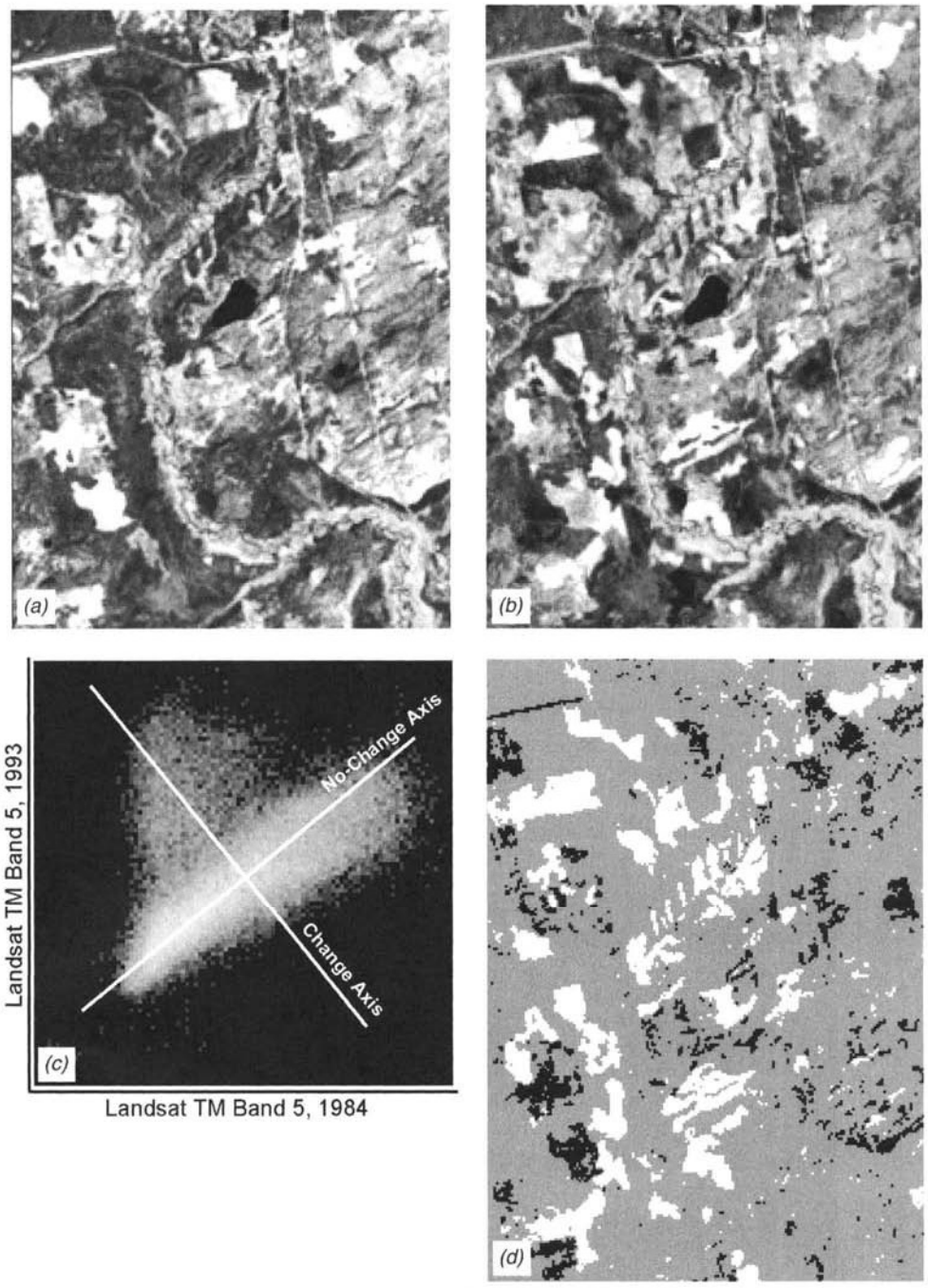

Figure 7.57 Application of delta transformation techniques to Landsat TM band 5 (mid-IR) images of the Nicolet National Forest: (*a*) 1984 image; (*b*) 1993 image; (*c*) two-date scatter plot, including "no-change" and "change" axes; (*d*) areas of forest cutting between image dates and forest regeneration between image dates shown as light and dark areas, respectively.

forest cutting activity between the dates, and several areas appear darker due to forest regeneration over the same time period.

Figure 7.57*c* depicts the two-date scatter plot for the images shown in (*a*) and (*b*). The direction of the stable spectral space and the land cover change axis are also shown. Figure 7.57*d* illustrates those areas that have changed between dates due to forest cutting and regeneration as lighter and darker areas, respectively.

Multisensor Image Merging

Plate 33 illustrates yet another form of data merging, namely, the combination of image data from more than one type of sensor. This image depicts an agricultural area in southcentral Wisconsin. The following sources of image data comprise this IHS-enhanced composite image:

1. SPOT HRV band 2 (red), shown in blue.
2. SPOT HRV band 3 (near IR), shown in green.
3. Landsat TM band 5 (mid IR), shown in red.
4. Digital orthophotograph, used in the intensity component of the IHS transformation.
5. GIS overlay of the boundaries of selected farm fields.

This image demonstrates how multisensor image merging often results in a composite image product that offers greater interpretability than an image from any one sensor alone. For example, Plate 33 affords the spatial information content of the digital orthophoto data (2 m) and the red and near-infrared spectral data from the SPOT HRV, as well as the mid-infrared information content of the Landsat TM (band 5). The composite image is aimed at exploiting the advantages of each of the data sources used in the merger process.

In addition to the merger of digital photographic and multispectral scanner data, multisensor image merging has been extensively used to combine multispectral scanner and radar image data. Such combinations take advantage of the spectral resolution of the multispectral scanner data in the optical wavelengths and the radiometric and "sidelighting" characteristics of the radar data.

Merging of Image Data with Ancillary Data

Probably one of the most important forms of data merger employed in digital image processing is the registration of image data with "nonimage," or ancillary, data sets. This latter type of data set can vary, ranging from soil type to

elevation data to assessed property valuation. The only requirement is that the ancillary data be amenable to accurate geocoding so that they can be registered with the image data to a common geographic base. Usually, although not a necessity, the merger is made in a GIS environment.

Digital elevation models (DEMs) have been combined with image data for a number of different purposes. Figure 7.58 illustrates the merger of DEM and image data to produce *synthetic stereoscopic images*. Shown in this figure is a synthetic stereopair generated by introducing simulated parallax into a Landsat MSS image. Whereas standard Landsat images exhibit only a fixed, weak stereoscopic effect in the relatively small areas of overlap between orbit passes, the synthetic image can be viewed in stereo over its entirety and with

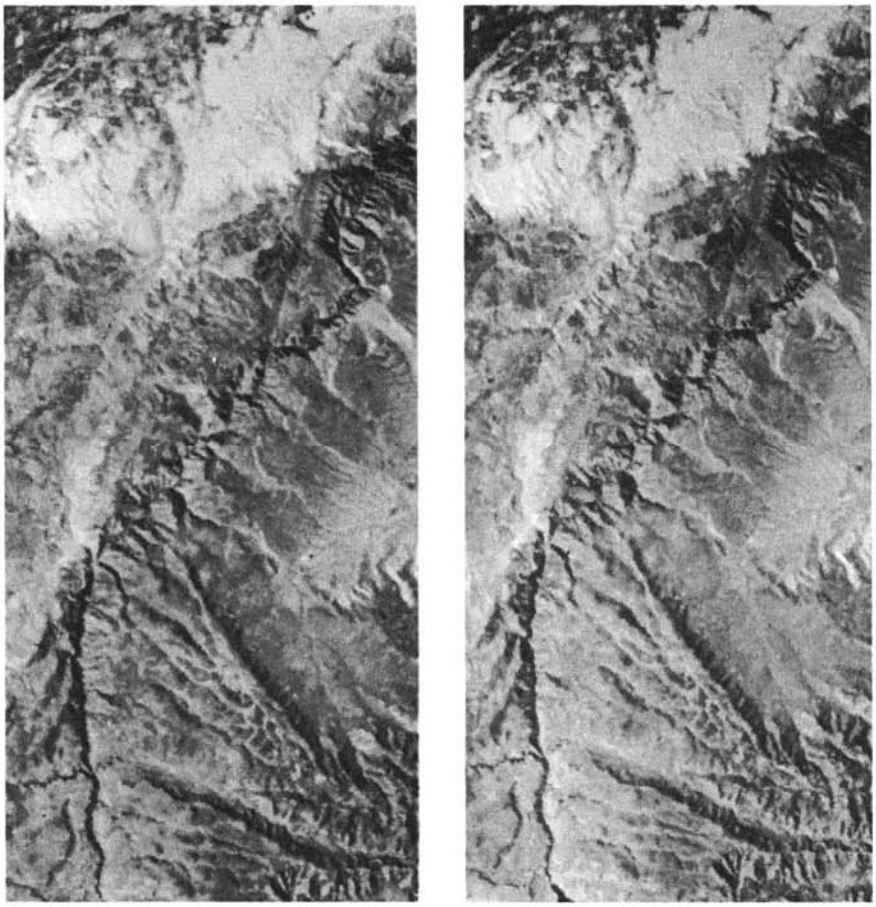

Figure 7.58 Synthetic stereopair generated from a single Landsat MSS image and a digital elevation model, Black Canyon of the Gunnison, CO. Maximum canyon depth is 850 m. Scale 1 : 400,000. (Courtesy USGS.)

an analyst-specified degree of vertical exaggeration. These images are produced in a manner similar to the process used to produce stereomates for orthophotographs (Section 3.9). That is, the elevation at each pixel position is used to offset the pixel according to its relative elevation. When this distorted image is viewed stereoscopically with the original scene, a three-dimensional effect is perceived. Such images are particularly valuable in applications where landform analysis is central to the interpretation process. The technique is also useful for restoring the topographic information lost in the preparation of spectral ratio images.

Another common use of DEM data is in the production of perspective-view images, such as Figure 7.59, a merger of Landsat TM and DEM data, which shows Mount Fuji, the highest peak in Japan.

Merging topographic information and image data is often useful in image classification. For example, topographic information is often important in forest-type mapping in mountainous regions. In such situations, species that have very similar spectral characteristics might occupy quite different elevation ranges, slopes, or aspects. Thus, the topographic information might serve as another "channel" of data in the classification directly or as a postclassification basis upon which to discriminate between only the spectrally similar classes in an image. In either case, the key to improving the classification is being able to define and model the various associations between the cover types present in a scene and their habitats.

Incorporating GIS Data in Automated Land Cover Classification

Obviously, topographic information is not the only type of ancillary data that might be resident in a GIS and useful as an aid in image classification. For example, data as varied as soil types, census statistics, ownership boundaries, and zoning districts have been used extensively in the classification process. The basic premise of any such operation is that the accuracy and/or the categorical detail of a classification based on image *and* ancillary data will be an improvement over a classification based on either data source alone.

Ancillary data are often used to perform *geographic stratification* of an image prior to classification. As with the use of topographic data, the aim of this process is to subdivide an image into a series of relatively homogeneous geographic areas (strata) that are then classified separately. The basis of stratification need not be a single variable (e.g., upland versus wetland, urban versus rural) but can also be such factors as landscape units or ecoregions that combine several interrelated variables (e.g., local climate, soil type, vegetation, landform).

There are an unlimited number of data sources and ways of combining them in the classification process. Similarly, the ancillary data can be used either prior to, during, or after the image classification process (or even some

Figure 7.59 Perspective-view image of Mount Fuji, Japan, produced by combining Landsat TM data (bands 2, 3, and 4, shown in black and white) and digital elevation model (DEM) data. (Copyright © RESTEC/NIHON University, 1991.) (Figure 8.45 shows a radar image of this area as well.)

combination of these choices might be employed in a given application). The particular sources of data used and how and when they are employed are normally determined through the formulation of *multisource image classification decision rules* developed by the image analyst. These rules are most often formulated on a case-by-case basis through careful consideration of the form, quality, and logical interrelationship among the data sources available. For example, the "roads" in a land cover classification might be extracted from current digital line graph (DLG) data rather than from an image source. Similarly, a particular spectral class might be labeled "alfalfa" or "grass" in different locations of a classification depending upon whether it occurs within an area zoned as agricultural or residential.

Space limits us from providing numerous examples of how user-defined classification decision rules are designed and implemented. In lieu of such a discussion, Plate 34 is presented to illustrate the general principles involved. Shown in Plate 34 is a "composite" (lower right) land cover classification performed in the vicinity of Fox Lake, Wisconsin, which is located in the east central portion of the state. The composite classification was produced from five separate data sources:

1. A supervised classification of the scene using a TM image acquired in early May (upper left).
2. A supervised classification of the scene using a TM image acquired in late June (upper right).
3. A supervised classification of both dates combined using a principal components analysis (not shown).
4. A wetlands GIS layer prepared by the Wisconsin Department of Natural Resources (DNR) (lower left).
5. A DNR-supplied road DLG (lower left).

Used alone, none of the above data sources provided the classification accuracy or detail needed for the purpose of monitoring the wildlife habitat characteristics of the area depicted in the plate. However, when all of the data were integrated in a GIS, the data analyst was able to develop a series of post-classification decision rules utilizing the various data sources in combination. Simply put, these decision rules were based on the premise that certain cover types were better classified in one classification than the others. In such cases, the optimal classification for that category was used for assigning that cover type to the composite classification. For example, the water class was classified with nearly 100 percent accuracy in the May classification. Therefore, all pixels having a classification of water in the May scene were assigned to that category in the composite classification.

Other categories in the above example were assigned to the composite classification using other decision rules. For example, early attempts to automatically discriminate roads in any of the TM classifications suggested that

this class would be very poorly represented on the basis of the satellite data. Accordingly, the road class was dropped from the training process and the roads were simply included as a DLG overlay to the composite classification.

The wetland GIS layer was used in yet another way, namely to aid in the discrimination between deciduous upland and deciduous wetland vegetation. None of TM classifications could adequately distinguish between these two classes. Accordingly, any pixel categorized as deciduous in the May TM data was assigned either to the upland or wetland class in the composite classification based on whether that pixel was outside or within a wetland according to the wetland GIS layer. A similar procedure was used to discriminate between the grazed upland and grazed wetland classes in the principal components classification.

Several of the land cover categories in this example were discriminated using rules that involved comparison among the various classifications for a given pixel in each of the three preliminary classifications. For example, hay was classified well in the May scene and the principal components image but with some errors of commission into the grazed upland, cool season grass, and old field categories. Accordingly, a pixel was classified as hay if it was hay in the May or principal components classification but at the same time was not classified as grazed upland or cool season grass in the principal components classification or as old field in the June classification. Similarly, pixels were assigned to the oats class in the composite classification if they were originally classified as oats, corn, peas, or beans in the May scene but oats in the June scene.

Table 7.5 lists a representative sample of the various decision rules used in the composite classification depicted in Plate 34. They are presented to illustrate the basic manner in which GIS data and spatial analysis techniques can be combined with digital image processing to improve the accuracy and categorical detail of land cover classifications. The integration of remote sensing, GIS, and "expert system" techniques for such purposes is an active area of current research. Indeed, these combined technologies are resulting in the development of increasingly "intelligent" information systems.

Another area of current research is the use of *artificial neural networks* in image classification. Such systems are "self-training" in that they adaptively construct linkages between a given pattern of input data and particular outputs. Neural networks can be used to perform traditional image classification tasks (Foody et al., 1995) and are also increasingly used for more complex operations such as spectral mixture analysis (Moody et al., 1996). For image classification, neural networks do not require that the training class data have a Gaussian statistical distribution, a requirement that is held by maximum likelihood algorithms. This allows neural networks to be used with a much wider range of types of input data than could be used in a traditional maximum likelihood classification process. In addition, once they have been fully trained, neural networks can perform image classification relatively rapidly,

TABLE 7.5 Basis for Sample Decision Rules Used for Composite Classification Shown in Plate 34

Sample Class in Composite Classification	Classification			GIS Data	
	May	June	PC	Wetlands	Roads
Water	Yes				
Roads					Yes
Deciduous upland	Yes			Outside	
Deciduous wetland	Yes			Inside	
Grazed upland			Yes	Outside	
Grazed wetland			Yes	Inside	
Hay	Yes		Yes		
		Old field—no	Grazed upland—no Cool season grass—no		
Oats	Oats—yes Corn—yes Peas—yes Beans—yes	Oats—yes			
Peas	Oats—yes Corn—yes Peas—yes Beans—yes		Peas—yes		
Beans	Oats—yes Corn—yes Peas—yes Beans—yes	Beans—yes			
Reed canary grass			Yes		
Warm season grass	Yes				
Cool season grass			Yes		

although the training process itself can be quite time consuming. In the following discussion we will focus on back-propagation neural networks, the type most widely used in remote sensing applications, although other types of neural networks have been described.

A neural network consists of a set of three or more layers, each made up of multiple nodes. These nodes are somewhat analogous to the neurons in a biological neural network and thus are sometimes referred to as neurons. The network's layers include an input layer, an output layer, and one or more hidden layers. The nodes in the input layer represent variables used as input to the neural network. Typically, these might include spectral bands from a remotely sensed image, textural features or other intermediate products derived from such images, or ancillary data describing the region to be analyzed. The nodes in the output layer represent the range of possible output categories to

be produced by the network. If the network is being used for image classification, there will be one output node for each class in the classification system.

Between the input and output layers are one or more hidden layers. These consist of multiple nodes, each linked to many nodes in the preceding layer and to many nodes in the following layer. These linkages between nodes are represented by weights, which guide the flow of information through the network. The number of hidden layers used in a neural network is arbitrary. Generally speaking, an increase in the number of hidden layers permits the network to be used for more complex problems but reduces the network's ability to generalize and increases the time required for training.

Figure 7.60 shows an example of a neural network that is used to classify land cover based on a combination of spectral, textural, and topographic information. There are seven nodes in the input layer, as follows: nodes 1 to 4 correspond to the four spectral bands of a multispectral scanner image, node 5 corresponds to a textural feature that is calculated from a radar image, and nodes 6 and 7 correspond to terrain slope and aspect, calculated from a digital elevation model. After the input layer, there are two hidden layers, each with nine nodes. Finally, the output layer consists of six nodes, each corresponding to a land cover class (water, sand, forest, urban, corn, and hay). When given any combination of input data, the network will produce the output class that is most likely to result from that set of inputs, based on the network's analysis of previously supplied training data.

If multiple sources of input data are used, the range of values in each data set may differ. In the example shown in Figure 7.60, the four multispectral bands might have digital number values ranging from 0 to 255, while the terrain slope might be expressed as a percentage and the terrain aspect might be in degrees. Before these various data sets can be provided to the neural network, they each must be rescaled to fit within the same range. Generally, all input data are scaled to fit within the range from 0 to 1.

Applying a neural network to image classification makes use of an iterative training procedure in which the network is provided with matching sets of input and output data. Each set of input data represents an example of a pattern to be learned, and each corresponding set of output data represents the desired output that should be produced in response to the input. During the training process, the network autonomously modifies the weights on the linkages between each pair of nodes in such a way as to reduce the discrepancy between the desired output and the actual output.

It should be noted that a back-propagation neural network is not guaranteed to find the ideal solution to any particular problem. During the training process the network may develop in such a way that it becomes caught in a "local minimum" in the output error field, rather than reaching the absolute minimum error. Alternatively, the network may begin to oscillate between two slightly different states, each of which results in approximately equal error (Paola and Schowengerdt, 1995). A variety of strategies have been proposed to

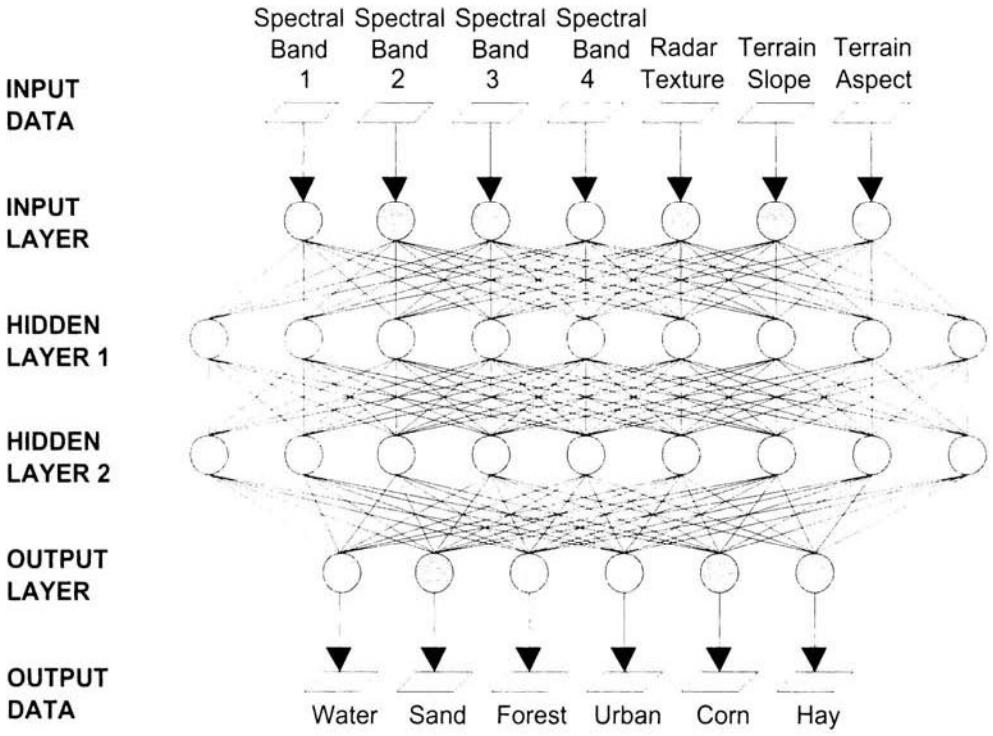

Figure 7.60 Example of an artificial neural network with one input layer, two hidden layers, and one output layer.

help push neural networks out of these pitfalls and enable them to continue development toward the absolute minimum error. Again, the use of artificial neural networks in image classification is the subject of continuing research.

7.18 HYPERSPECTRAL IMAGE ANALYSIS

The hyperspectral sensors discussed in Sections 5.14 and 6.15 differ from other optical sensors in that they typically produce contiguous, high resolution radiance spectra rather than discrete measurements of average radiance over isolated, wide spectral bands. As a result, these sensors can potentially provide vast amounts of information about the physical and chemical composition of the surface under observation as well as insight into the characteristics of the atmosphere between the sensor and the surface. While most multispectral sensors merely *discriminate* among various earth surface features, hyperspectral sensors afford the opportunity to *identify* and *determine many characteristics about* such features. However, these sensors have their disadvantages as well, including an increase in the volume of data to be