

in the perceived hues and intensities due to enhancement. This makes interpretation of the enhanced image straightforward, with the decorrelated information exaggerated primarily in terms of saturation. Previously pastel hues become much more saturated.

Because decorrelation stretching is based on principal component analysis, it is readily extended to any number of image channels. Recall that the IHS procedure is applied to only three channels at a time.

7.7 IMAGE CLASSIFICATION

The overall objective of image classification procedures is to automatically categorize all pixels in an image into land cover classes or themes. Often this is done using *spectral patterns*; that is, pixels that share similar combinations of spectral reflectance or emissivity are grouped together in classes that are assumed to represent particular categories of surface features. No attention is paid to the neighbors or surroundings of the pixel being classified. The term *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.

This approach can be extended to make use of many other types of pixel-level data. For example, polarimetric radar imagery could be classified using polarization pattern recognition; multitemporal imagery of any type could be classified using temporal pattern recognition; multi-angle imagery such as that from MISR could be classified based on bidirectional reflectance patterns; and so forth. What all these examples have in common is that each pixel is classified, individually, based on some statistical or deterministic model using the values from multiple data layers (spectral bands, polarization bands, temporal bands, ...) for that one pixel.

Spatial pattern recognition is a very different approach, involving 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.

These two types of image classifiers may be used in combination in a hybrid mode. For example, *object-based* image analysis (OBIA) involves combined use of both spectral and spatial pattern recognition. It is important to emphasize that 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.

We begin our discussion of image classification with treatment of spectrally oriented procedures for land cover mapping. Historically, spectral approaches have

formed the backbone of multispectral classification activities (although with the current wide-scale availability of high resolution data, there is increased use of spatially oriented procedures). 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.

Building upon the preceding methods, we then 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 also look at other specialized topics, such as the classification of mixed pixels (Section 7.13), object-based classification using both spatial and spectral information (Section 7.15), and the use of neural networks for classification (Section 7.16), before concluding this extensive portion of the chapter with a discussion of methods for assessing and reporting the accuracy of image classifications (Section 7.17).

Note that hyperspectral images present their own issues for classification, and as a result some specialized procedures have been developed for classification of these data. We defer coverage of such methods to Section 7.21, along with other analytical procedures for use with hyperspectral data.

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 sensor data. (The identical procedures would apply to Landsat, SPOT, WorldView-2, or virtually any other source of

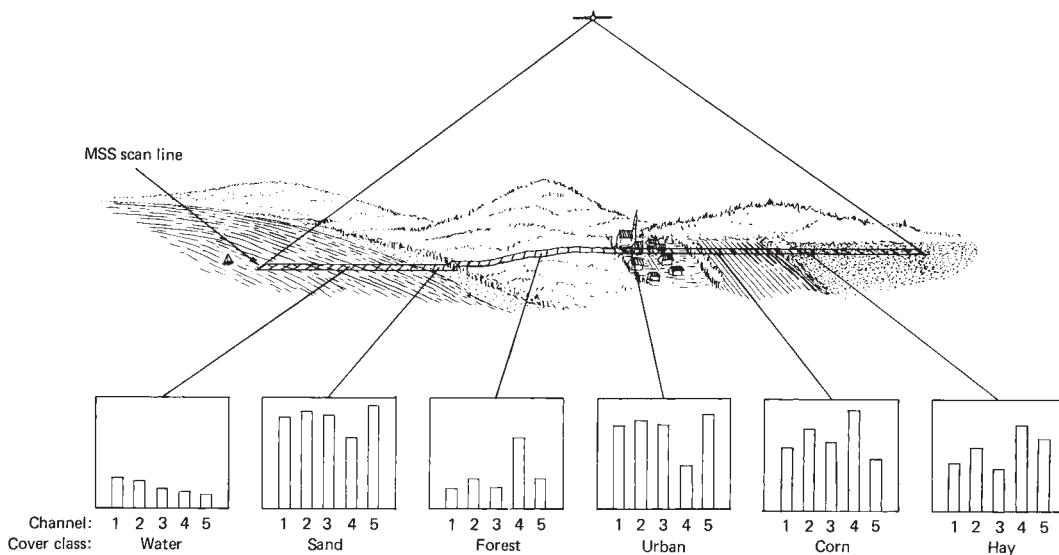


Figure 7.34 Selected multispectral sensor measurements made along one scan line. Sensor covers the following spectral bands: 1, blue; 2, green; 3, red; 4, near infrared; 5, thermal infrared.

multispectral data.) Figure 7.34 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 sensor 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 7.35 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 class it most closely resembles. If the pixel is insufficiently similar to any training data set, it is usually labeled “unknown.” After all pixels in the input image have 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 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.”

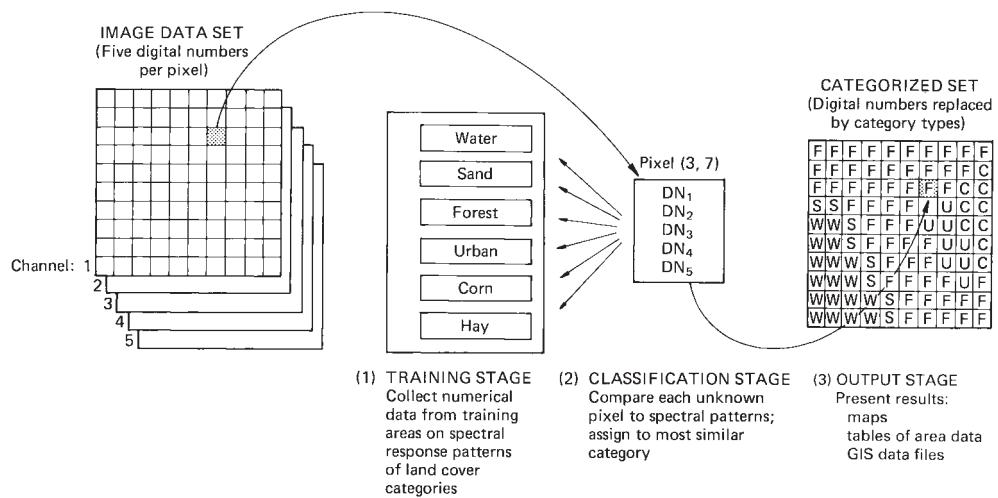


Figure 7.35 Basic steps in supervised classification.

We begin with a discussion of the *classification* stage because it is the heart of the supervised classification process—during this stage a computer-based evaluation of the spectral patterns is made 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. Our discussion only scratches the surface of this topic.

We illustrate the various classification approaches with a two-channel (bands 3 and 4) subset of our hypothetical five-channel multispectral sensor 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 7.36. 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

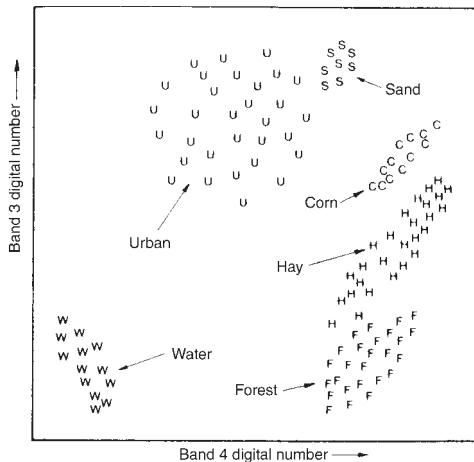


Figure 7.36 Pixel observations from selected training sites plotted on scatter diagram.

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.¹

Let us also assume that the pixel observations shown in Figure 7.36 are from areas of known cover type (that is, from selected training sites). Each pixel 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 7.37 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 + symbols in Figure 7.37. By considering the two-channel pixel values as positional coordinates (as they are portrayed in the scatter

¹ Pattern recognition literature frequently refers to individual bands of data as *features* and scatterplots of data as *feature space plots*.

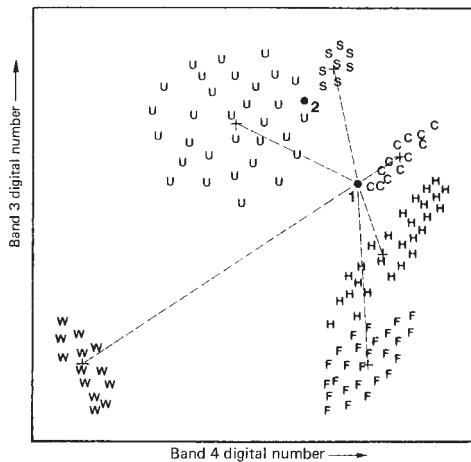


Figure 7.37 Minimum distance to means classification strategy.

diagram), a pixel of unknown identity may be classified by computing the *distance* between the value of the unknown pixel and each of the category means. In Figure 7.37, 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 7.37, 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 7.38. 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 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 with low values in band 4. The “urban” class shows a lack of covariance, resulting in a nearly circular distribution on the scatter diagram.

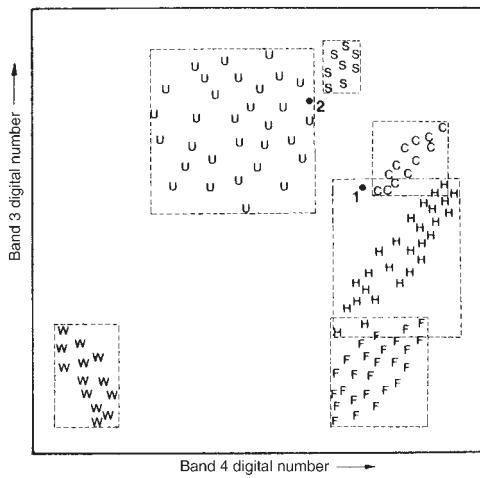


Figure 7.38 Parallelepiped classification strategy.

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.

Gaussian Maximum Likelihood Classifier

The maximum likelihood classifier quantitatively evaluates both the variance and covariance of the category spectral response patterns when classifying an 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 7.39 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.

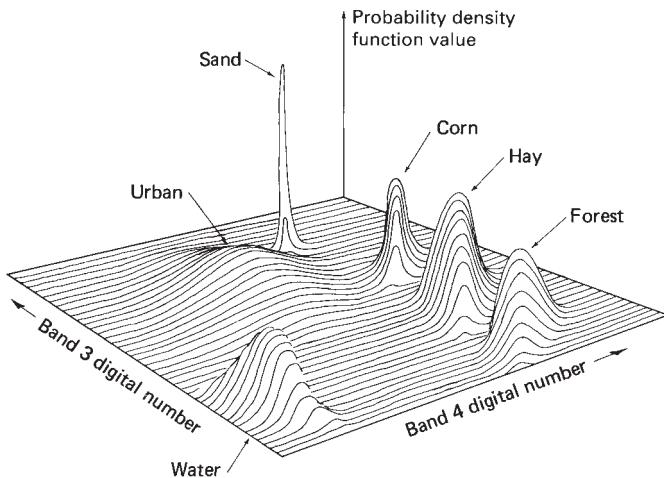


Figure 7.39 Probability density functions defined by a maximum likelihood classifier.

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.

In essence, the maximum likelihood classifier delineates ellipsoidal "equiprobability contours" in the scatter diagram. These decision regions are shown in Figure 7.40. 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.

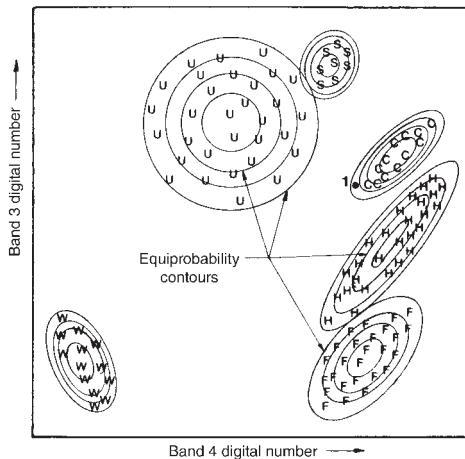


Figure 7.40 Equiprobability contours defined by a maximum likelihood classifier.

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. However, with the rapid increase in computational power over the past several decades, this computational complexity is no longer a major consideration for most applications. If an increase in speed is required, one approach is to reduce the dimensionality of the data set used to perform the classification (thereby reducing the complexity of the required computations). As discussed in Section 7.6, 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.

Figure 7.41 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 and to avoid any areas that are visually “rough” in the imagery. The row and column coordinates of the vertices for these polygons are used as the basis for extracting



Figure 7.41 Training area polygons delineated on a computer monitor. (Author-prepared figure.)

(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).

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 $10n$ 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.

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 pairs of training sets is studied. The analyst confirms that all data sets are unimodal and reasonably close to a Gaussian distribution. Training areas that are bimodal or whose distributions are highly skewed may include more than one spectral class and should be deleted or split. Likewise, extraneous pixels may be deleted from some of the training 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 7.42a shows a hypothetical histogram for one of the “hay” category training sites in our five-channel data set. (A similar display would be available for all training areas.) Histogram output is particularly important when a maximum likelihood classifier is used, because 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 7.42b. 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 gray bars). Such plots indicate the overlap between category response patterns. For example, Figure 7.42b 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 7.36 to 7.38 and 7.40) provide better representations of the spectral response pattern distributions.

The utility of scatter diagrams (or scatter plots) is further illustrated in Figures 7.43 to 7.45. Shown in Figure 7.43 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 7.44 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 7.45 shows the histograms and the scatter diagram for bands 2 and 3. In contrast to Figure 7.44, the scatter diagram in Figure 7.45

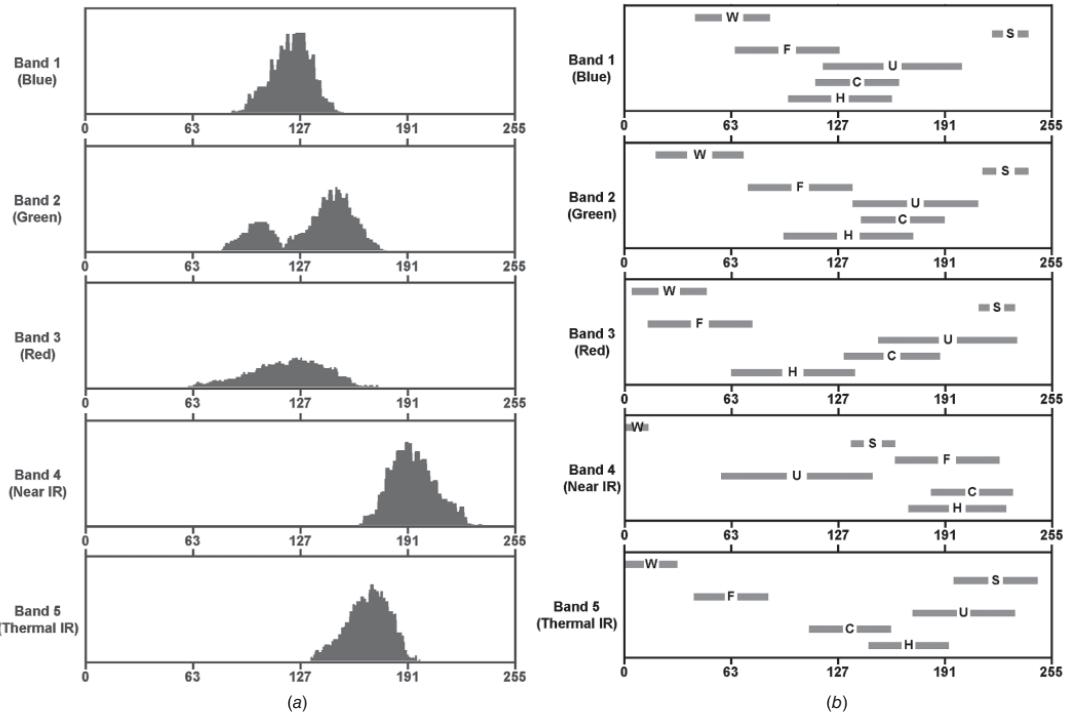


Figure 7.42 Visualization of training plot data. (a) Sample histograms for data points included in the training areas for cover type “hay.” (b) Coincident spectral plots for training data obtained in five bands for six cover types.

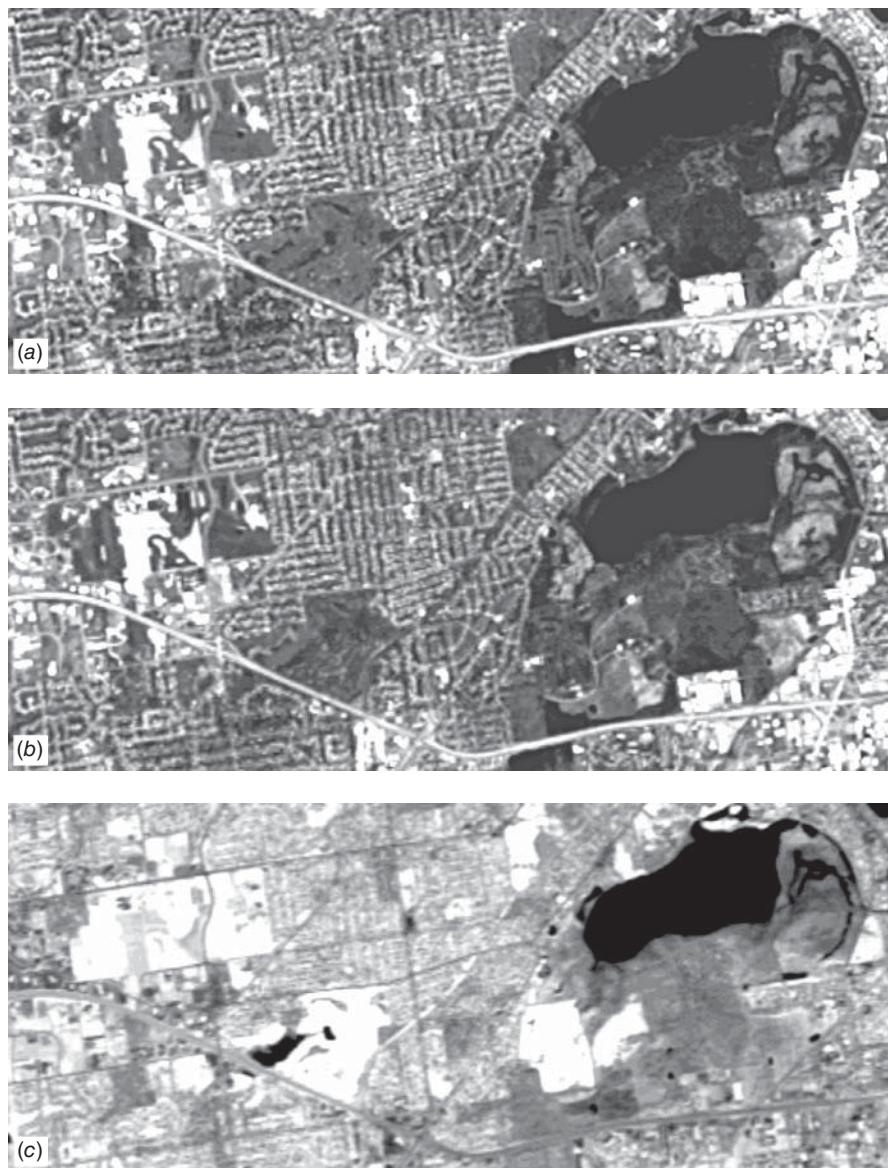


Figure 7.43 SPOT HRV multispectral images of Madison, WI: (a) band 1 (green); (b) band 2 (red); (c) band 3 (near IR). (Author-prepared figure.)

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.

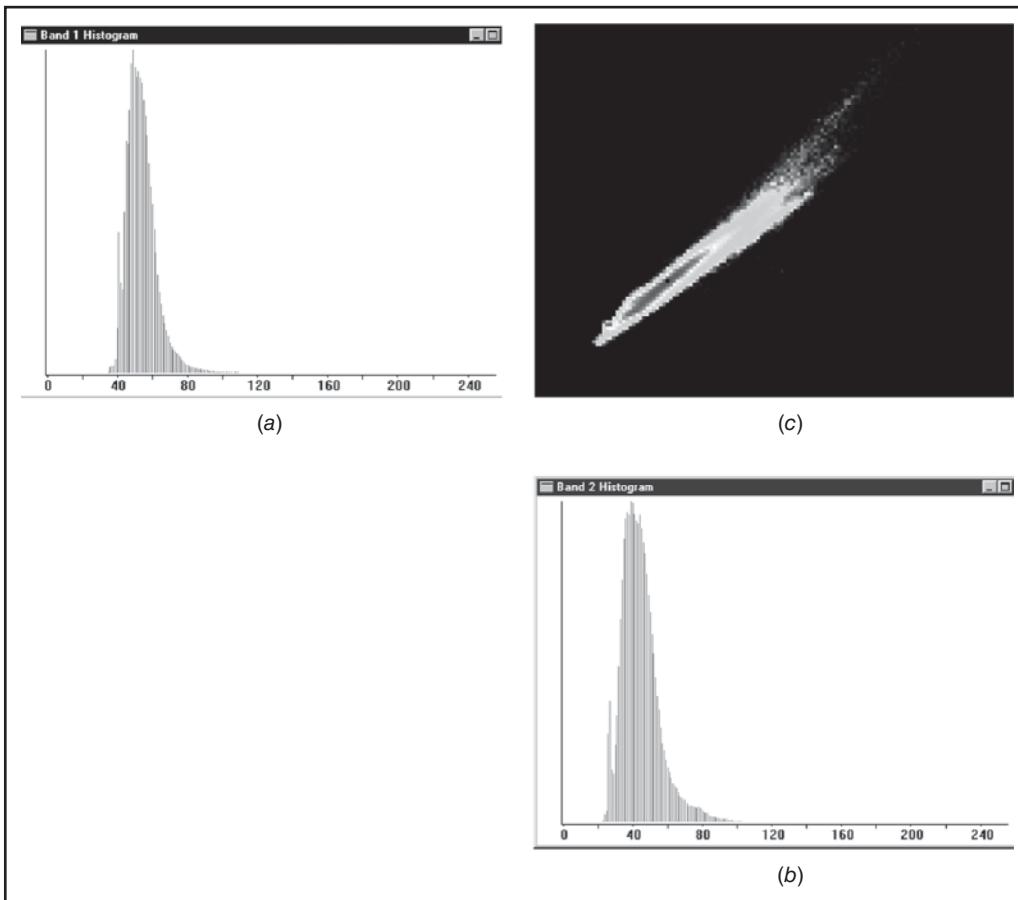


Figure 7.44 Histograms and two-dimensional scatter diagram for the images shown in Figures 7.43a and b: (a) band 1 (green) histogram; (b) band 2 (red) histogram; (c) scatter diagram plotting band 1 (vertical axis) versus band 2 (horizontal axis). Note the high correlation between these two visible bands. (Author-prepared figure.)

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 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

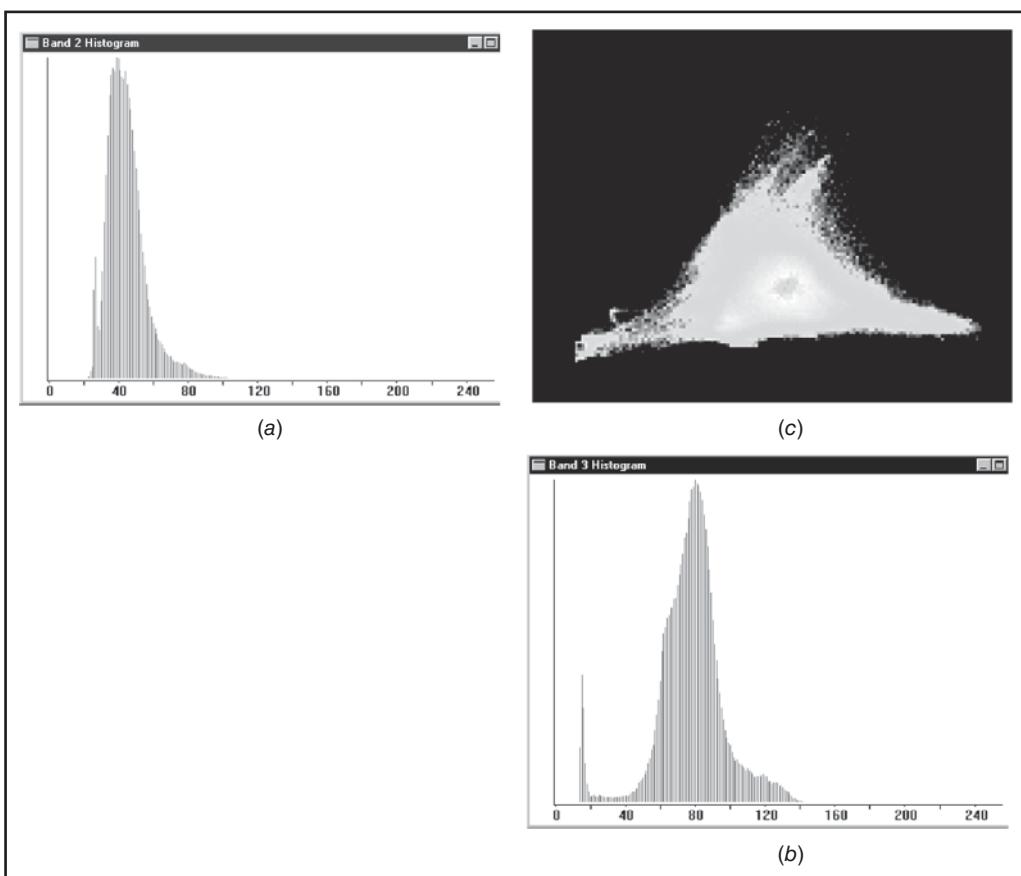


Figure 7.45 Histograms and two-dimensional scatter diagram for the images shown in Figures 7.43b and c: (a) band 2 (red) histogram; (b) band 3 (near-IR) histogram; (c) scatter diagram plotting band 2 (vertical axis) versus band 3 (horizontal axis). Note the relative lack of correlation between these visible and near-IR bands. (Author-prepared figure.)

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

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
:	:								

^aW, water; C, corn; H, hay.

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 can be presented in the form of an *error matrix* (to be described in Section 7.17).

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 entire classification process is complete (as discussed in Section 7.17).

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 30, which shows a partially completed classification of a subset of the data included in Figures 7.43 and 7.45 (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 should not be rushed. 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”).

When multiple images need to be classified—either for a wider area or to study changes in land cover over time—the traditional approach is to extract spectral training data from each image individually. This is necessary because each image will typically have variations in atmospheric conditions, sun angle,

and other factors that are sufficient to render the spectral “signatures” extracted from one image unrepresentative of the same classes in the other image(s). If there is a need to classify multiple images with the actual spectral training data from a single image, one approach is to use relative radiometric normalization (Figure 7.5; Section 7.2) to adjust the radiometric characteristics of all the images to match a single “base” image prior to classification. Obviously, the success of this approach is dependent on the accuracy with which the secondary images can be matched to the base image. In addition, this process works best for images of non-vegetated areas, or for images acquired at the same point in the phenological cycle.

One final note to be made here is that training set refinement is essential 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 incorporating additional imagery from other sensors, using data resident in a GIS, or performing a visual interpretation, 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 7.46 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

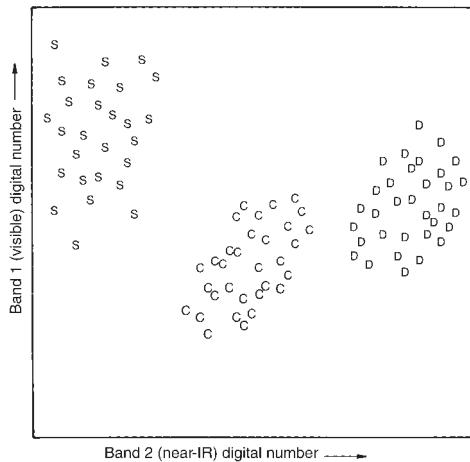


Figure 7.46 Spectral classes in two-channel image data.

one to stressed trees of both types (indicated by D, C, and S in Figure 7.46). 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 measurement 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 Techniques A*, or *ISODATA* (Ball and Hall, 1965).² This algorithm permits the number of clusters to change from one iteration to the next, by merging, splitting, and deleting clusters.

²In the words of its originators, “the A was added to make *ISODATA* pronounceable.”

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 other variants of 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.

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 supervised/unsupervised approaches in Section 7.12.) 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.

Often a multistage approach is used with unsupervised classification to improve the representation of certain classes that are imperfectly differentiated in the initial classification. In this approach, two or more clusterings are used to narrow the focus on a particular class of interest. The general sequence is shown in Figure 7.47:

1. **Initial unsupervised classification.** In the first classification, one spectral class is over-broad; it includes pixels that should belong in other spectral classes.
2. **Masking of problem class.** A new image is created in which only pixels from the “problem” class are retained; all others are set to values of no-data.
3. **Second-stage classification of problem class.** A second unsupervised classification is performed, using only pixels from the “problem” class.
4. **Recode output from second-stage classification.** The spectral subclasses from step 3 are reassigned to existing classes from the initial classification in step 1, or to new classes.
5. **Merger of classification results.** The results from the recoded output of the second-stage classification are inserted back into the output image from the initial classification.

The result of this procedure is a modified classification that is identical to the original, except that one over-broad spectral class has been split into two or more other classes. If necessary, more than one “problem” class can be split this way, either simultaneously or in series. Note that this approach is conceptually similar

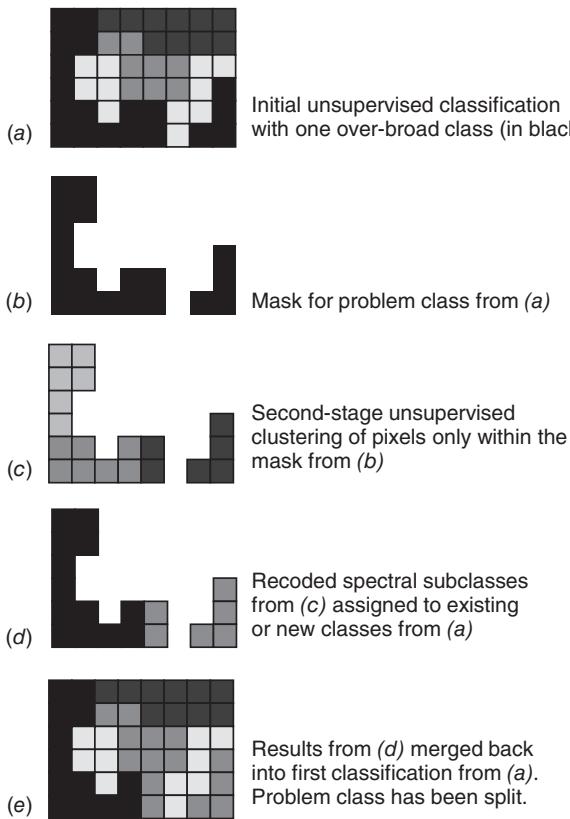


Figure 7.47 The multistage unsupervised classification process.

to the supervised decision-tree (stratified or layered) classifier discussed at the end of Section 7.9, but using unsupervised clustering at each stage rather than a supervised procedure.

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. 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

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	Water
3	Sunlit conifers	Coniferous trees
4	Shaded hillside conifers	Coniferous trees
5	Upland deciduous	Deciduous trees
6	Lowland deciduous	Deciduous trees
7	Brushland	Brushland
Possible Outcome 3		
1	Turbid water	Water
2	Clear water	Water
3	Coniferous trees	Coniferous trees
4	Mixed coniferous/deciduous	Coniferous trees Deciduous trees
5	Deciduous trees	Deciduous trees
6	Deciduous/brushland	Deciduous trees Brushland

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.

7.12 HYBRID CLASSIFICATION

Various forms of hybrid supervised/unsupervised 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 supervised/unsupervised 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 (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 (Bauer et al., 1994; Lillesand et al., 1998; Reese et al., 2002; Chipman et al., 2011).

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, \dots, 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 (a) display of a given class on the raw image, (b) multidimensional histogram analysis for each cluster, and (c) 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.

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 1.5 and 4.2), mixed pixels result when a sensor's instantaneous field of view (IFOV) includes more than one land cover type or feature on the ground. This can occur for several reasons. First, it is common for two or more different feature types to be widely intermingled across the landscape (e.g., every ground resolution cell within an agricultural field may include both corn plants and bare soil). Second, even when a landscape is made up of more or less homogeneous areas, the ground resolution cells (or image pixels) falling along the boundaries between these homogeneous patches will contain some mixture of the two classes. In either case, these mixed pixels present a difficult problem for image classification, because 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.

Subpixel classification is also frequently referred to as “soft” classification, in that more than one possible land cover class can be associated with each pixel, and information on the proportion of each class might be determined. Figure 7.48

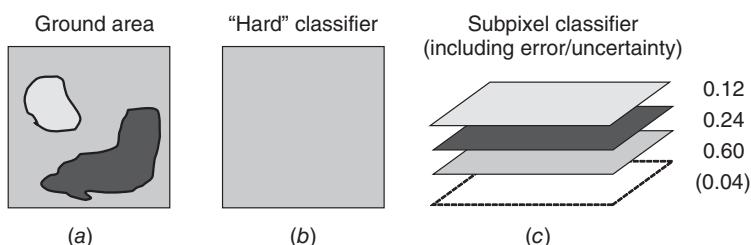


Figure 7.48 Principle of subpixel classification. (a) Ground area contained within a single pixel of the image, including patches of three different land cover types. (b) Output from a traditional “hard” classifier, representing only a single class. (c) Output from a subpixel classifier, including estimated fraction of each class plus error.

contrasts traditional “hard” classification with this subpixel approach. When an image pixel covers a ground area with multiple land cover classes at the subpixel scale, the traditional classifier assigns a single class. As shown in (b), this could be the numerically dominant class within the pixel, but it could also be a class that represents a smaller fraction of the area but has a particularly high reflectance. Alternatively, and perhaps worse still, the classifier could interpret the mixture of these three classes as being most similar to some other class that is not present at all in the pixel! (One common real-world example of this problem occurs at the boundaries of water bodies. Because water has very low near-IR reflectance while deciduous forest can have very high reflectance in the near IR, image pixels spanning the border between the two may have an intermediate near IR reflectance. Based on this intermediate value, the “hard” classifier often will conclude that this pixel is some other class entirely, such as evergreen forest or impervious surfaces.)

As shown in Figure 7.48c, a subpixel classifier would produce multiple output layers—one for each class, plus (in some cases) an additional estimate of error or uncertainty. For a given pixel, the value in each layer represents the fraction of the pixel that consists of that layer’s class. In principle, these fractions should add up to 1.0 (or 100%); depending on the software implementation and on methodological preferences, the analyst may choose to force this “unit sum” constraint, or to allow the sum of all class fractions to depart from 1.0. In the latter case, this “error” may represent the presence of additional land cover types not represented in the set of classes used by the classifier or imperfections in the statistical representation of the existing set of classes.

Figure 7.49 illustrates two class layers (soil and vegetation) resulting from a subpixel classification. In both cases, the brighter the pixel value, the higher the fractional area of that class at the subpixel level. Areas that appear dark in both (a) and (b) represent landscape units that have low fractions of both soil and vegetation. (Many of those pixels are in water bodies, a third class not shown here.)

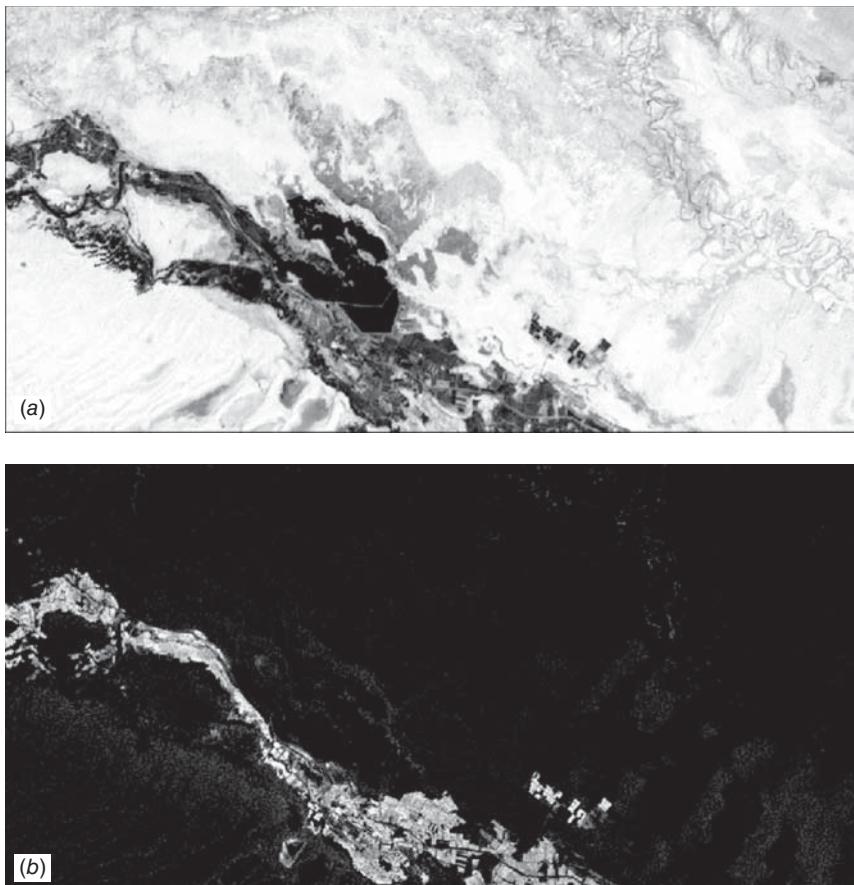


Figure 7.49 Examples of subpixel classification. (a) Soil fraction. (b) Vegetation fraction. Lighter tones represent higher fractional coverage at the subpixel level. (Author-prepared figure.)

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. For example, Figure 7.50 shows spectra for two “pure” feature types in an EO-1 Hyperion hyperspectral image of Kilauea Volcano, Hawaii. One of the pure spectra is for a bare, unvegetated lava flow (less than 30 years old), and the other is for a nearby tropical rainforest canopy. The third spectrum, located between the other two in Figure 7.50, is for a mixed pixel located on the boundary between a

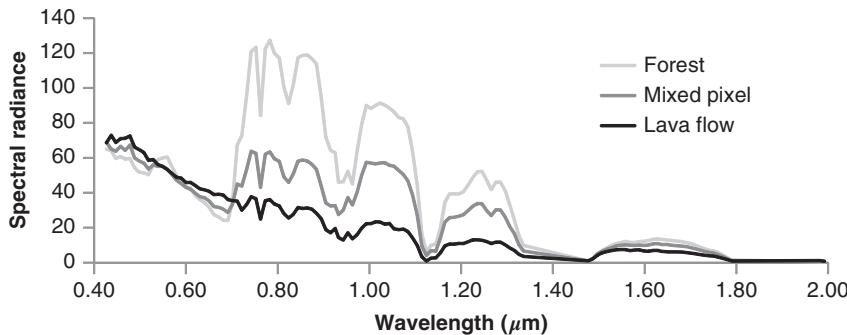


Figure 7.50 Linear mixture of spectra from an EO-1 Hyperion image of tropical rainforest and lava flows in Hawaii.

lava flow and undisturbed forest. From a numerical analysis using the assumption of linear mixing (see below) this mixed pixel's area is estimated to consist of approximately 55% forest and 45% lava flow.

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, because it is based on a physical model of the mixture of discrete spectral response patterns. It provides useful information at the subpixel level, because 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% 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,

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

where F_1, F_2, \dots, 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

$$\text{DN}_\lambda = F_1 \text{DN}_{\lambda,1} + F_2 \text{DN}_{\lambda,2} + \dots + F_N \text{DN}_{\lambda,N} + E_\lambda \quad (7.11)$$

where DN_λ is the composite digital number actually observed in band λ ; F_1, \dots, F_N equal the fractions of the pixel actually occupied by each of the N endmembers; $\text{DN}_{\lambda,1}, \dots, \text{DN}_{\lambda,N}$ equal the digital numbers that would be observed 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, \dots, 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-6 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 7.51 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 7.51a shows a single band (TM band 4, near IR), while 7.51b through 7.51d 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

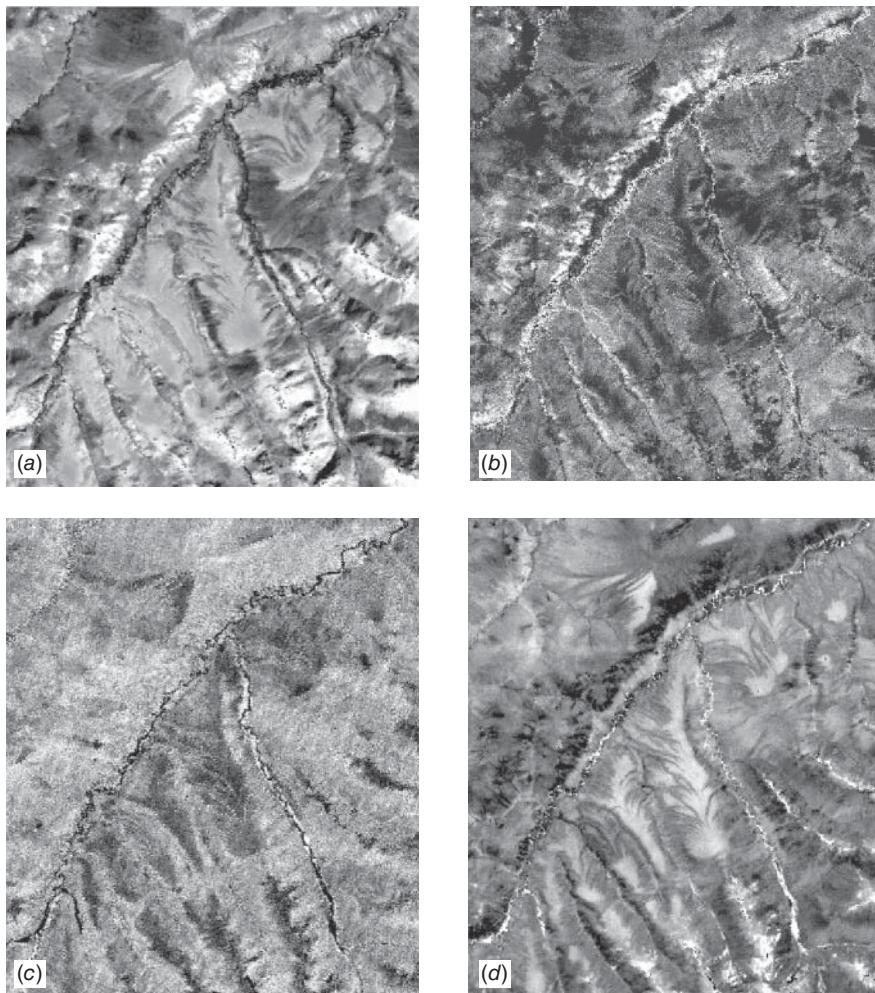


Figure 7.51 Linear spectral mixture analysis of a Landsat TM image including the Steese National Conservation Area of central Alaska: (a) band 4 (near IR) of original image; fractional cover images for trees (b), shrubs (c), and herbaceous plants (d). Brighter pixels represent higher fractional cover. (Courtesy Bureau of Land Management–Alaska and Ducks Unlimited, Inc.)

include additional radiance reflected multiple times between scene components such as leaves and the soil surface. In this situation, a more sophisticated *non-linear* spectral mixture model may be required (Somers et al., 2009). Artificial neural networks (Section 7.16) 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 (Plaza et al., 2011).

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 membership in more than one category (Jensen, 2005; Schowengerdt, 2006). 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 AND POSTCLASSIFICATION SMOOTHING

The utility of any image classification is ultimately dependent on the production of output maps, tables, and geospatial data that effectively convey the interpreted information to its end user. Here the boundaries between remote sensing, digital cartography, geovisualization, and GIS management become blurred. Plate 2a shows the output from a land cover classification that is being used as input to a model of soil erosion potential within a GIS. For internal use within the GIS, nothing more is needed from the classification output beyond its inherent raster grid of pixel values. On the other hand, as shown in Plate 2a, for visualization purposes choices about scale, color selection, shading, and other cartographic design topics become important. In this example, the colors associated with each land cover class are modified with a light hillshading algorithm to convey the shape of local topography, to aid the viewer in their interpretation of the landscape.

One issue that often arises during the output stage is the need for smoothing of classification images to remove isolated misclassified pixels. 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 7.52a). For example, in an agricultural area, several pixels scattered throughout a cornfield 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 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 7.52b was prepared in this manner, applying a 3×3 -pixel majority filter to the data shown in Figure 7.52a. Figure 7.52c 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

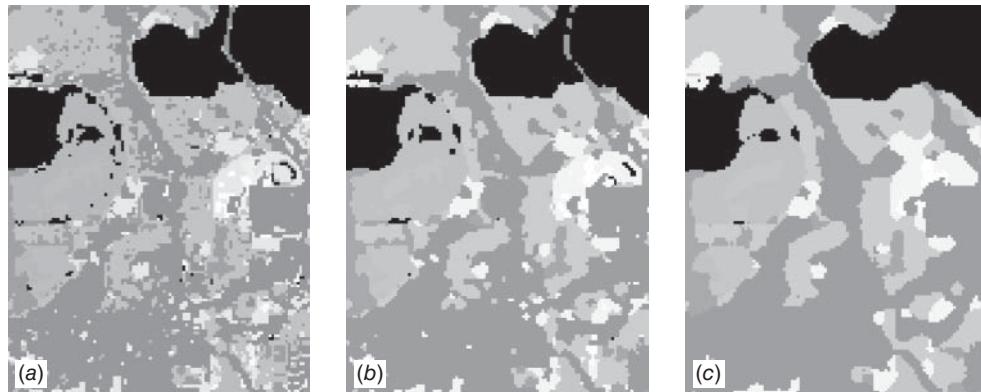


Figure 7.52 Postclassification smoothing: (a) original classification; (b) smoothed using a 3×3 pixel-majority filter; (c) smoothed using a 5×5 -pixel majority filter. (Author-prepared figure.)

user-specified minimum area of any given land cover type that will be maintained in the smoothed output.

Another way of obtaining smoother classifications is to integrate the types of logical operations described above directly into the classification process. Object-based classification, the subject of the following section, affords this capability (and much more).

7.15 OBJECT-BASED CLASSIFICATION

All of the classification algorithms discussed to this point have been based solely on the analysis of the spectral characteristics of individual pixels. That is, these *per pixel* procedures generally use spectrally based decision logic that is applied to each pixel in an image individually and in isolation. In contrast, object-based classifiers use both spectral and spatial patterns for image classification (Blaschke, 2010). This is a two-step process involving (1) segmentation of the imagery into discrete objects, followed by (2) classification of those objects. The basic assumption is that the image being classified is made up of relatively homogeneous “patches” that are larger in size than individual pixels. This approach is similar to human visual interpretation of digital images, which works at multiple scales simultaneously and uses color, shape, size, texture, pattern, and context information to group pixels into meaningful objects.

The scale of objects is one of the key variables influencing the image segmentation step in this process. For example, in the case of a forested landscape, at a fine scale the objects being classified might represent individual tree crowns. Segmentation at an intermediate scale would produce objects that correspond to stands of trees of similar species and sizes, while at a still coarser scale, large areas of the forest would be aggregated into a single object. Clearly, the actual scale parameter used for an object-based classification will depend on a number of factors, including the resolution of the sensor and the general scale of the features on the landscape that the analyst is seeking to identify.

Once an image has been segmented, there are many characteristics that can be used to describe (and classify) the objects. Broadly speaking, these characteristics fall into two groups. One set of characteristics is intrinsic to each object—its spectral properties, its texture, its shape, for example. Other characteristics describe the relationships among objects, including their connectivity, their proximity to objects of the same or other types, and so forth. For example, an object that has a linear shape, a spectral signature similar to asphalt, a smooth texture, and topological connectivity to other road objects would probably be itself a road.

Figure 7.53 illustrates object-based segmentation and classification. Figure 7.53a is a grayscale rendering of a color-IR composite of Landsat-5 TM data from bands 2, 3, and 4 (originally depicted in blue, green, and red, respectively). This image was acquired over western Vilas County, located in north-central Wisconsin. The major land cover types present in this area include water,

deciduous forest, evergreen forest, cranberry bog, wetlands, and clear-cut/barren areas. Figures 7.53b–d are a series of segmentations that were produced from the TM data holding most parameters constant (e.g., spectral band weighting, shape factor, smoothness factor), but varying the scale factor.

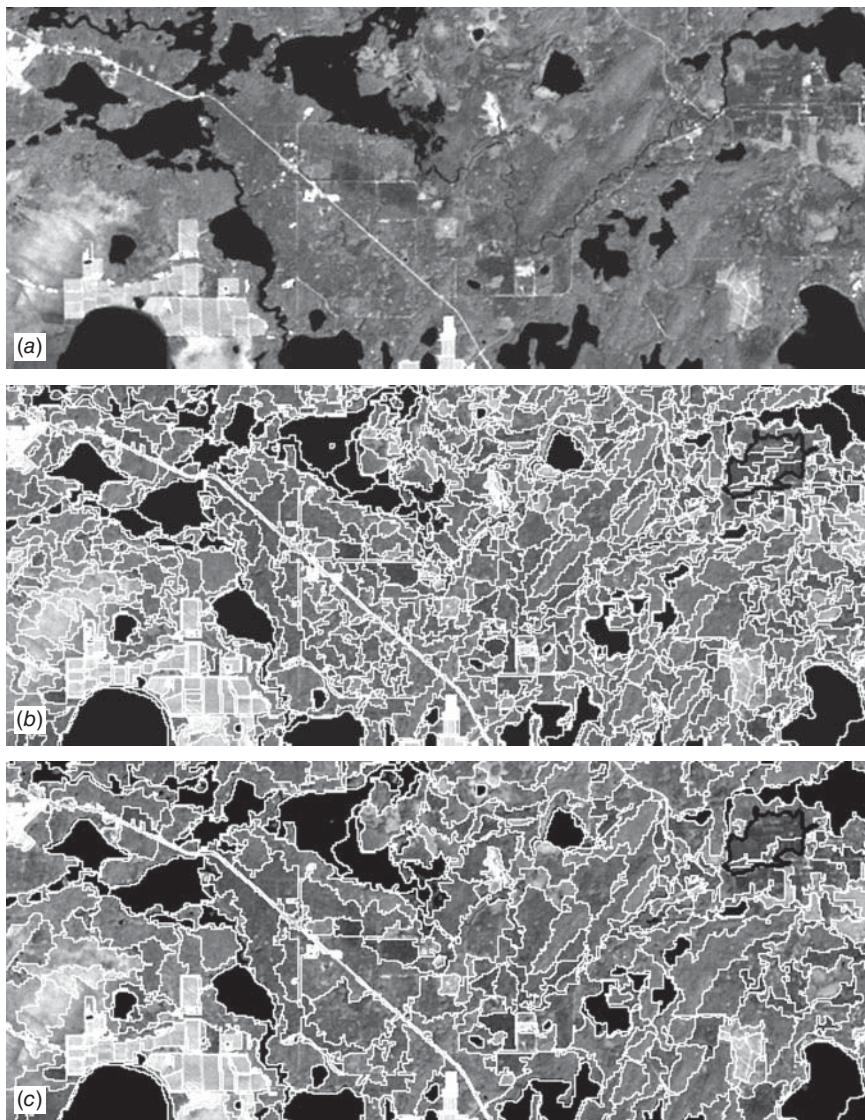


Figure 7.53 Object-oriented image segmentation and classification. (a) Landsat-5 TM image obtained over Vilas County, WI. (b) Image segmentation using a fine scale factor. (c) Image segmentation using an intermediate scale factor. (d) Image segmentation using a coarse scale factor. (e) Land cover classification using image segmented with a fine scale factor. (f) Traditional unsupervised “per pixel” land cover classification of nonsegmented image data. (Author-prepared figure.)

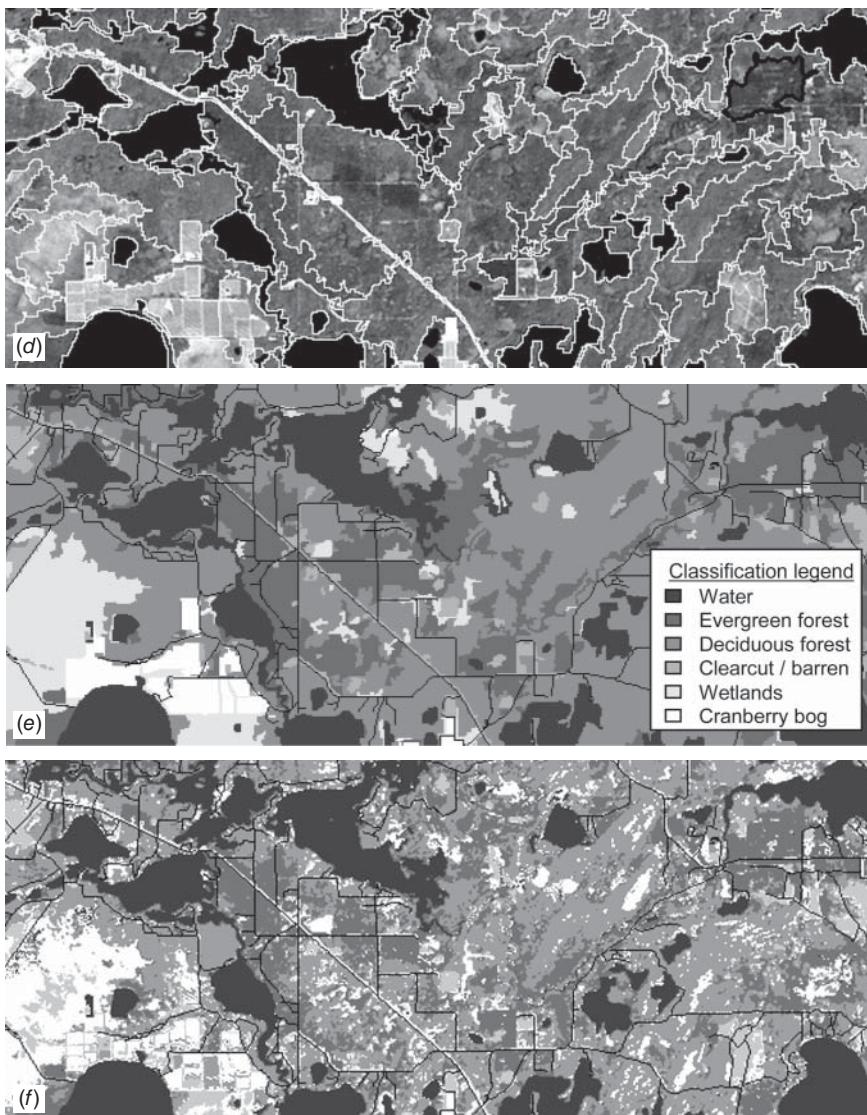


Figure 7.53 (Continued)

At a fine scale factor, shown in (b), the landscape is divided into small objects corresponding to features such as groups of trees of uniform species composition, sunlit or shady sides of hills, small wetlands, and individual cranberry bog cultivation beds. These objects appear as white polygons superimposed on a backdrop of the original (unclassified) Landsat imagery. The large polygon outlined in black in the upper right portion of the image represents a tree plantation with stands of

conifers (red pine and jack pine) of varying ages and sizes. Note that the polygon has been subdivided into multiple objects corresponding to individual, homogeneous stands.

At an intermediate scale factor, shown in (c), the objects are larger, and at a coarse scale factor, shown in (d), the entire plantation outlined in black is now represented by a single object. Note that the objects at different scales “nest” in a hierarchical fashion, so that all the objects at a fine scale fit perfectly within the boundaries of aggregated objects at a coarser scale.

Once the boundaries of the objects have been established, they can then be classified based on the various characteristics described above in a wide variety of ways. The classification shown in Figure 7.53e was produced by first segmenting the image using a fine scale factor and then aggregating the resulting objects into the six land cover information classes listed above. (The roads shown in this image were “burned in” from a transportation GIS layer.) A traditional (non-segmented) unsupervised classification of the image is presented in Figure 7.53f. Compared to these per-pixel classifier results, the object-based approach produces a classification that is much smoother in appearance, with relatively homogeneous regions. Note also how the segmentation aids in the discrimination of linear features, such as the road trending from the upper left side to the middle of the lower edge of the study area.

Object-based analysis can also be used to facilitate land cover change in that the approach is capable of preserving the “parent–child” relationships among objects. For example, a large agricultural field containing only one crop type at an early time period might be classified as a single object. If multiple crop types are present in the field at a later date, the parent field will be split into numerous child objects. Spectral data could then be used to classify the identity of the new child objects.

7.16 NEURAL NETWORK CLASSIFICATION

Another increasingly common method for image classification makes use of *artificial neural networks*. Initially inspired by the biological model of neurons in the brain, the use of artificial neural networks was pioneered within computer science’s subfields of artificial intelligence, machine learning, and related areas. In later years, implementation of artificial neural networks in computer science focused less on replicating actual biological structures and processes, and instead developed its own statistical and computational logic. Nonetheless, the terminology (including the use of the term “neurons” within these networks) persists, as a legacy of the method’s biologically inspired origins.

Neural networks 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 or for more complex operations such as spectral mixture analysis. For image classification, neural

networks do not require that the training 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. (The use of ancillary data from other sources in combination with remotely sensed imagery for classification purposes is discussed in Section 7.20.) In addition, once they have been fully trained, neural networks can perform image classification relatively rapidly, 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 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 7.54 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 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.

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

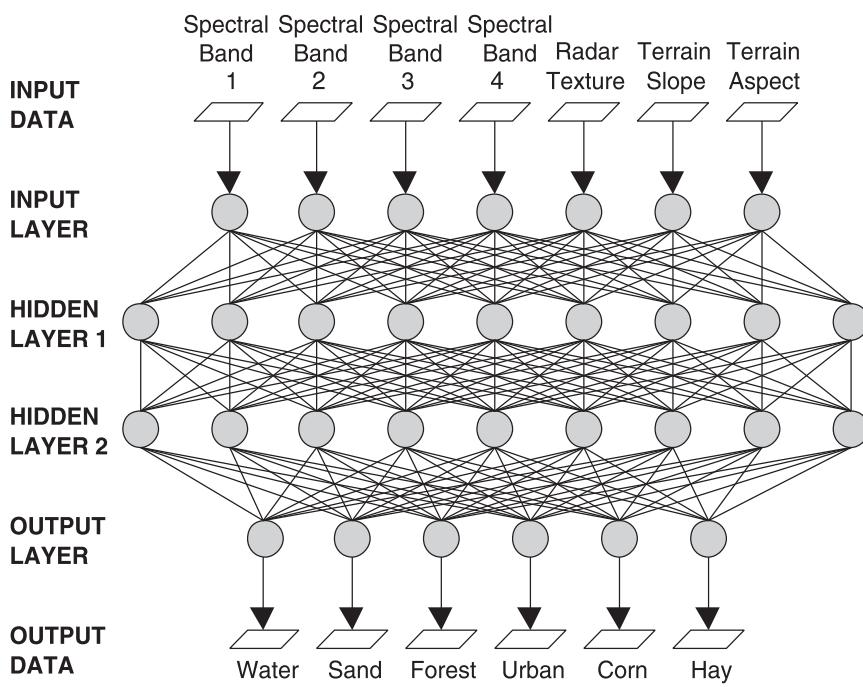


Figure 7.54 Example of an artificial neural network with one input layer, two hidden layers, and one output layer.

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. A variety of strategies have been proposed to help push neural networks out of these pitfalls and enable them to continue development toward the absolute minimum error.

7.17 CLASSIFICATION ACCURACY ASSESSMENT

The process of producing a land cover classification from remotely sensed data can be lengthy. At the end of the process, when the classification algorithm has done its work, any problematic classes have been split or refined, and the analyst

has recoded the larger set of spectral classes into a reduced set of information classes, it is tempting to publish the resulting land cover map and then move on to other work. However, before moving on to other endeavors, it is essential to document the *categorical accuracy* of the classification results. This need is embodied in the expression “A classification is not complete until its accuracy is assessed.”

Congalton and Green (2009) 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 set of test pixels whose “true” classes were determined from *ground truth* data. This ground truth provides information about the actual land cover types present on the ground at the test pixel locations. (Sampling procedures for collecting these ground truth data will be discussed later in this section.) The matrix in Table 7.3 stems from classifying the test pixels and listing the known (“true”) cover types from ground truth (columns) versus the mapped cover type from the classifier (rows).

Many 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 test 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., seven pixels that should have been classified as “water” were omitted from that category—three were misclassified as “forest,” two as “urban,” and one each as “corn” and “hay”). Commission errors are represented by nondiagonal row elements (e.g., 92 “urban” pixels plus one “corn” pixel were improperly included in the “sand” 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

TABLE 7.3 Error Matrix Resulting from Classifying Test Pixels

	Reference Data ^a						Row Total
	W	S	F	U	C	H	
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%							

^aW, water; S, sand; F, forest; U, urban; C, corn; H, hay.

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 test set pixels used for that category (the column total). This figure indicates how well test 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 65%. However, producer's accuracies range from just 42% ("urban") to 97% ("water"), and user's accuracies vary from, again, 42% ("corn") to 94% ("water").

From a statistical perspective, it is essential to consider the independence of the test data used for assessing the final accuracy of the classification. Generally

speaking, these data should not have been used previously in any stage of the classification process. In some cases, analysts will compute an error matrix based on the same data used for training the classifier (in a supervised classification process) or based on the same data used as guidance for the analyst when assigning spectral clusters to information classes in an unsupervised classification. *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. Examination of such an error matrix (created using the training data, rather than independent test data) can aid in the training set refinement process, but it indicates little about how the classifier performs elsewhere. 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!)

Evaluating Classification Error Matrices

A number of features are readily apparent from inspection of the error matrix included in Table 7.3. 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%. 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%). This would potentially lead one to the conclusion that although the overall accuracy of the classification was poor (65%), 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%. That is, even though 84% of the forested areas have been correctly identified as "forest," only 60% of the areas identified as "forest" within the classification are truly of that category. Accordingly, although the producer of the classification can reasonably claim that 84% of the time an area that was forested was identified as such, a user of this classification would find that only 60% 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 some correct values in the error matrix. In fact, if the number of classes is small, such a random assignment could result in a surprisingly good apparent classification result—a two-category classification could be expected to be 50% correct solely due to random chance. The \hat{k} ("kappa" or "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

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

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% 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.

The \hat{k} statistic is computed as

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

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 kappa for the error matrix included in Table 7.3,

$$\begin{aligned} \sum_{i=1}^r x_{ii} &= 226 + 216 + 360 + 397 + 190 + 219 = 1608 \\ \sum_{i=1}^r (x_{i+} \cdot x_{+i}) &= (239 \cdot 233) + (309 \cdot 328) + (599 \cdot 429) \\ &\quad + (521 \cdot 945) + (453 \cdot 238) + (359 \cdot 307) = 1,124,382 \\ \hat{k} &= \frac{2480(1608) - 1,124,382}{(2480)^2 - 1,124,382} = 0.57 \end{aligned}$$

Note that the \hat{k} 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, \hat{k} 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 \hat{k} 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, different classification techniques, or different individuals performing the classification. Such tests are based on computing an estimate of the variance of \hat{k} , 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 each other. This is a somewhat complicated process because the calculation for the variance of \hat{k} depends on the sampling design used to collect the test data. The interested reader is encouraged to consult the literature for details; for example, Stehman (1996) provides the formula for the variance of \hat{k} under stratified random sampling.

Sampling Considerations

Test areas are areas of representative, uniform land cover that are different from and (usually) more numerous than training areas. They are often located during the training stage of supervised classification by intentionally collecting ground truth for more sites than are actually needed for training data. 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. In addition, training data are often collected using an “ad-hoc” sampling design that is neither systematic nor random, instead emphasizing the opportunistic acquisition of ground truth at sites that are readily accessible on the ground, readily interpretable in higher-resolution imagery, or otherwise conducive to use in the training process. Such a sampling scheme may be inadequate for the statistical validity that is desired in the accuracy assessment process.

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), but any such “interventions” will again affect the interpretation of the resulting accuracy statistics.

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.

Other sampling designs have also been described, such as those 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. One frequent mistake is to use test areas consisting of multiple pixels and treat each pixel as an independent entry in the error matrix. This is inappropriate, due to the extremely high spatial autocorrelation (local similarity) among adjacent pixels in most remotely sensed data sets.

Sample size must also weigh heavily in the development and interpretation of classification accuracy figures. As a broad guideline, it has been suggested that a minimum of 50 samples of each land cover category be included in the error matrix. Classifications covering large areas (over a million acres) or with more than 12 land cover categories require more thorough sampling, typically 75 to 100 samples per category (Congalton and Green, 2009, p. 75). 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). As noted earlier, however, all decisions about sampling design must include consideration of the statistical validity and tractability of the accuracy assessment process.

Final Thoughts on Accuracy Assessment

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, image 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 unacceptable 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. The propagation of errors through the multiple layers of information in a GIS is beyond the scope of this work; interested readers may consult the literature on GIS and spatial analysis.

7.18 CHANGE DETECTION

One of the most powerful advantages of remote sensing images is their ability to capture and preserve a record of conditions at different points in time, to enable the identification and characterization of changes over time. This process is referred to as *change detection* and is among the most common uses of digital image analysis. The types of changes that might be of interest can range from nearly instantaneous (movement of a vehicle or animal) 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. For detecting changes at time scales greater than one year, *anniversary dates* are preferred to minimize sun angle and seasonal differences. Accurate spatial registration of the various dates of imagery