An Evaluation of the Potential for Fuzzy Classification of Multispectral Data Using Artificial Neural Networks

Timothy A. Warner and Michael Shank

Abstract

Fuzzy classification, or pixel unmixing, is the estimation of the proportion of the cover types from the composite spectrum of a mixed pixel. In this paper, we evaluate how the separation between class means, the covariance matrix of each class, and the relative location of the class means in the spectral space limit the fuzzy representation of mixtures. The influence of these factors is illustrated with a fuzzy classification using a back-propagation artificial neural net. Experiments using simulated data indicate that a fuzzy classification with an average error of less than 10 percent requires a Bhattacharyya Distance between classes of at least 9. The error in the fuzzy representation using a neural net also varies as the proportions of the classes changes, with a peak error when one class comprises approximately 0.20 to 0.25 of the mixed pixel. Back-propagation neural networks are not necessarily good at spectral unmixing. The back-propagation neural network produces spectral-space partitions between the classes that are generally steep, and that are not necessarily midway between the classes. The partitions tend to be simple, and somewhat linear. In addition, the output on nodes does not have to sum to 1.0, which may result in situations where high values are predicted for two classes simultaneously. Two methods of improving neural network behavior for fuzzy classification include use of a compound linear-sigmoid activation function, and training using synthetic mixed pixels.

Introduction

Mixed pixels occur when the instantaneous field-of-view of a sensor falls on an area that includes more than one spectral cover type. Such pixels are a problem in classification because the radiance recorded at the sensor is a composite of the individual classes. Changing the spatial resolution so that smaller objects can be discerned does not always mitigate this problem, as there are typically optimal scales of observation (Woodcock and Strahler, 1987). Therefore, a finer spatial resolution may change the nature of the classes that are mixed, rather than reduce the problem of mixed pixels. Thus, for example, mixtures arise from different forest stands in low resolution imagery, from different trees in high resolution imagery, and from different leaves in extremely high resolution data. In many cases, mixing takes place at such a fine scale, such as the case with minerals in a rock, that this problem will probably never be eliminated. Finally, it is worth considering that it may sometimes actually be desirable to produce a fuzzy classification. For example, to remotely sense a forest community response to an environmental gradient, it may be necessary to quantify a subtle change in the ratio of certain species, rather than to map a change from one pure class to another (Warner *et al.*, 1994).

If a pixel has a spectral radiance that is intermediate between two classification cover classes, one interpretation is that this is a mixed pixel. Both classes are assumed to be present in the pixel, and the pixel spectral radiance represents the sum of the spectra of the individual classes, weighted by the proportion of each class in the instantaneous field-of-view. (For a recent discussion of non-linear mixing of vegetation and soils, see Ray and Murray (1996).) An alternative interpretation, however, is that only one cover class is present. The fact that the cover class has a spectral reflectance intermediate between the two previously chosen classification classes may be a result of chance. Often, however, the pixel cover class is itself transitional between the two previously chosen classes; for example, if two previously chosen classification classes are, respectively, Mature and Young Forest, an intermediate class might be Intermediate Age Forest. Intermediate classes can be described by a fuzzy membership function, which can also be used in a classification scheme. However, in this case there is no *a priori* reason that there will necessarily be a linear relationship between the change in spectral reflectance and the change in the class itself. Therefore, in the rest of this paper we limit our discussion to spectral unmixing.

Spectral unmixing has become particularly important with the advent of hyperspectral imaging because the tremendous number of bands facilitates the simultaneous mapping of multiple classes (Fox *et al.*, 1990; Boardman, 1994; Resmini *et al.*, 1996). Drawing on a linear mixing assumption, models have been developed to identify component end-members in the spectral space (for example, Adams *et al.* (1989), Smith *et al.* (1990), and Foody and Cox (1994)). In a different approach, more traditional parametric classification methods have been adapted to relate classifier output to class proportion, rather than to probability (Wang, 1990; Foody *et al.*, 1992; Maselli *et al.*, 1994). Recently, it has been suggested that the output of artificial neural networks can be related to the mixtures present within a pixel (Civco and Wang, 1994; Foody, 1996). Artificial neural networks are discussed in greater detail in the next section.

T.A. Warner is with the Department of Geology and Geography, West Virginia University, Morgantown, WV 26506-6300 (twarner2@wvu.edu).

M. Shank is with the West Virginia Department of Environmental Protection, 10 McJunckin Road, Nitro, WV 25143-2506.

Photogrammetric Engineering & Remote Sensing, Vol. 63, No. 11, November 1997, pp. 1285–1294.

0099-1112/97/6311-1285$3.00/0

With such interest in pixel unmixing, it is important that the factors that limit the potential for fuzzy classification be understood. Clearly, the factors involved will be interrelated in a complex manner, and must include both the sensor and the scene. The sensor characteristics that are most important include the number, spectral width, and spectral region of the sensor bands; the sensor quantization, bias, and gain; and the sensor spatial resolution. The most important scene characteristics are the number of spectral classes in the scene, the covariance matrices of the classes, and the relative distribution of the classes in the spectral space, including both the relative separation and the relative spatial arrangement of the classes.

In this paper we investigate how the spectral classes in the scene determine the potential for fuzzy classification. The role of sensor characteristics in fuzzy classification goes beyond the scope of this paper. We use simulated data to investigate the basic principles, and then apply those principles to real data. Although a back-propagation neural network is used to estimate the fuzzy membership function, many of the issues raised are of significance to pixel unmixing methods in general.

Neural Networks

Neural Networks and Image Classification

Artificial neural networks (ANNs) — also known as connectionist models, parallel distributed processing models, or simply neural networks — developed out of investigations of human cognition, and the belief that they could be useful in understanding aspects of perception, learning, memory, and language. Neural networks are constructed as a set of nodes connected by weighted, directed edges. This structure is thought to be a mathematical analog of the interaction of the brain's neurons through a web of axons (Schalkoff, 1992).

The most common network structure used for classification problems organizes nodes into several layers, in which each node in one layer is connected to every node in the succeeding layer, while nodes within a layer are not connected. These *feed-forward* networks usually consist of three or more layers. During operation, the data vector is presented to an input layer and then propagated through one or more hidden layers to an output layer. Each node in the hidden and output layers calculates a weighted sum of all its inputs and applies the result to an activation function to produce the node's output. Neural networks develop functional relationships by adjusting the values of these weighted connections between nodes, using an appropriate learning algorithm. The back-propagation learning algorithm, developed by Rumelhart *et al.* (1986), is commonly used with feed-forward networks. This algorithm incrementally reduces error between actual and desired outputs during iterative presentations of a training set until a preset error threshold has been reached. At that point the network is considered ready to process real data.

Neural networks form one branch of artificial intelligence research. However, unlike more common expert systems, neural networks develop functional relationships between evidence and conclusions automatically from examples, without the need for constructing an extensive and complex set of rules. One disadvantage with this "black box" approach is that it does not generally facilitate an understanding of the relationship between the input data and output categories. Mapping the partition boundaries, a technique employed in this paper, is one way of overcoming this limitation.

Neural networks have been shown to compare favorably with more conventional classification methods for classification of remotely sensed data (Key *et al.*, 1989; Decatur, 1989; Hepner *et al.*, 1990; Benediktsson *et al.*, 1990; Lee *et al.*, 1990; Heerman and Khazenie, 1992; Bischof *et al.*, 1992). Because the algorithm is non-parametric (Lippmann, 1987, Foody *et al.*, 1995), neural nets are particularly useful for integrating disparate data types (Key *et al.*, 1989) and ancillary data, such as topographic information (Benediktsson *et al.*, 1990), measures of texture (Lee *et al.*, 1990; Bischof *et al.*, 1992; Civco and Wang, 1994), and property ownership data (Zhuang *et al.*, 1991). Neural networks also are able to "generalize" inputs, producing correct outputs for new inputs that the neural network has not been trained to recognize (Rich and Knight, 1994).

Artificial Neural Networks and Fuzzy Image Classifications

Neural networks are commonly trained to produce a set of output values, usually in the range of [0,1], with a high value on the node corresponding to the desired class, and a low value for all others. The output of a neural network with sufficient hidden units has been found to approximate the *a posteriori* probabilities of the training classes, (Ruck *et al.*, 1990), and thus has greater similarities with statistical classification techniques than is often assumed. The interpretation of neural network output, however, has usually conformed to the convention of a "hard" classification by assuming that only one actual cover class is possible for a particular pixel. Recently, an alternative approach has been suggested in which the continuous range in output values of neural networks is interpreted as the fuzzy membership values (Bezdek, 1993; Foody, 1996). In the following discussion we discuss the use of back-propagation neural networks and fuzzy classification, although it is important to recognize that there are other fuzzy neural network classifiers that do fuzzy classification efficiently (see, for example, Carpenter *et al.* (1992) and Kosko (1992)).

In fuzzy classification approaches utilizing neural networks, the difference between the two highest output values has been related to classification confidence (Bischoff *et al.*, 1992) and used as an indicator of the likelihood of a mixed pixel (Civco and Wang, 1994). Foody (1996) related the node output of a neural network classification to the proportion of each cover type present in mixed pixels. The partitions between the pure classes were found to be rather steep, and thus unsuitable for fuzzy classification. However, these partitions were made more gradational by replacing the conventional sigmoid activation function used in training the neural network with a linear activation function for the final classification. The fuzzy classification was applied to two case studies in which synthetic low-resolution mixed pixels were produced by combining adjacent pixels of high-resolution imagery.

A recurrent problem in pixel unmixing methods is that in a group of mixed pixels there is variability contributed by both the normal variation in the pure spectral classes, as well as from the varying proportions of the mixed classes. As will be shown, the covariance matrix of the pure classes is therefore very important in limiting the potential accuracy of the estimation of class proportions because, as the variability of the pure classes increases, likewise the spectral variability of each specific mixture also becomes greater.

Methods

Fuzzy Classification of Simulated Data

A C program was used to develop simulated, bivariate, normally distributed classes with user-specified class means, standard deviation, and sample size. Each class is assumed to represent a "pure" spectral class. Mixed pixels were pro-

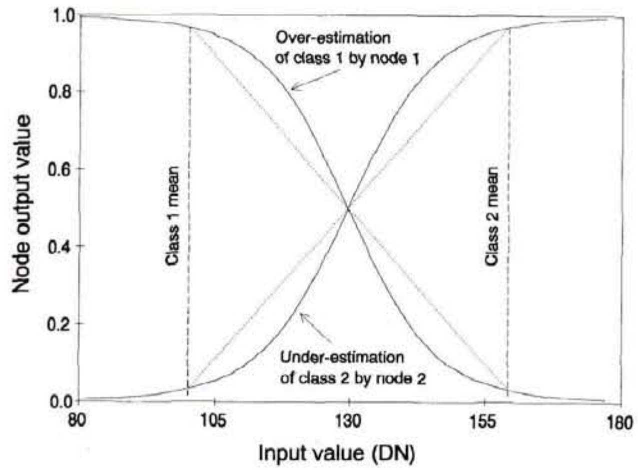

Figure 1. Shape of the fuzzy partition produced by the neural network using a sigmoidal activation function. Assuming the linear mixing indicated by the dotted line, the curved shape of the partition causes over- and under-estimation of the proportions of the two classes occurring in simulated mixed pixels.

duced by averaging varying proportions of randomly selected pixels from the pure classes. A simulated scene comprised 1,000 pixels, with 100 in each of the two pure classes and eight mixtures of varying proportions. For each experiment, ten trials were carried out by repeating the exercise ten times.

The neural network used in the classification was based on the back-propagation algorithm of Rumelhart *et al.* (1986), described in the previous section. The neural net consisted of two input nodes, four hidden nodes, and two output nodes. Each input was associated with a simulated spectral band, while each output node was associated with a cover class. The number of hidden nodes was determined empirically by determining the minimum number of nodes for which the network produced repeatable and accurate classification. The neural network was trained to produce a scaled output of 1.0 at the node corresponding to the correct cover class, and an output of 0.0 for the other node, using a 5 percent random sample from each pure class. Various combinations of learning rates and momentum terms were tried during the experiments. It was found that increasing the learning rate improved training times until the network began to oscillate, and failed to converge. Optimal learning rates were in the range of 0.2 to 0.3, and momentum terms varied between 0.6 and 0.9. Within those ranges, varying these terms did not appear to affect the final results. Because of the varying spectral separability of the training data, it was not practical to specify an absolute error level for training to end. Instead, a threshold level of a 1×10^{-6} change in error between iterations provided a more consistent point to exit training.

One concern in neural network classification is overtraining. This is where the neural network produces a partition of the classification space that is so narrowly defined around the training pixels, that the partition becomes only valid for the specific pixels used in training, and, consequently, the neural network's ability to generalize is compromised. Where such a problem occurs, the solution is to halt the training early, once a good general partition has been learned. This research emphasizes the visualization of the classification space, and, therefore, particularly facilitated our ability to check for this phenomenon. In general, overtraining was not a problem in the classification of either the simulated data or the real data discussed in the next section.

After training, the neural network classified the entire data set of both pure and mixed pixels. It was assumed that node outputs should correspond directly to the proportion of each class used to create the mixed pixel. Error was measured as the average of the absolute values of the difference between actual class proportions (p_{actual}) and scaled neural network outputs ($p_{estimated}$) for both classes, over the n trials.

Foody (1996) found that the standard sigmoid activation function provided a rather poor estimation of fuzzy membership. In our experiments, we mapped the partition of the spectral space (see Figure 1) and found a pattern very similar to the sigmoid activation function used in training the neural net. The sigmoid pattern results in errors concentrated on the shoulders of the curve, where one class dominates another. Errors are minimal for pure classes, or those that are mixed in equal proportions. This suggested the use of a compound linear-sigmoid function (Figure 2), which can be used for both training and classification.

To facilitate the application of these experiments to real data, we need to express the separation between the classes in terms of a general separability index (Swain and Davis, 1978; Richards, 1993). One index that takes into account both the distance between the means, and the covariance matrices of the classes, is the Bhattacharyya Distance (B_{ij}): i.e.,

$$B_{ij} = \frac{1}{8}(\boldsymbol{\mu}_i - \boldsymbol{\mu}_j)^T \left(\frac{\boldsymbol{\Sigma}_i + \boldsymbol{\Sigma}_j}{2}\right)^{-1} (\boldsymbol{\mu}_i - \boldsymbol{\mu}_j) + \frac{1}{2}\mathrm{Ln}\frac{\det\left(\frac{1}{2}(\boldsymbol{\Sigma}_i + \boldsymbol{\Sigma}_j)\right)}{\{\det(\boldsymbol{\Sigma}_i)\det(\boldsymbol{\Sigma}_j)\}^{1/2}}$$

where

i and j are the classes,
$\boldsymbol{\mu}_i$ and $\boldsymbol{\mu}_j$ are the mean vectors of classes i and j,
$\boldsymbol{\Sigma}_i$ and $\boldsymbol{\Sigma}_j$ are the covariance matrices for classes i and j,
$(\,)^{-1}$ is the inverse of the matrix,
det means determinant of the matrix, and
^{T} denotes the transpose of the matrix.

Simulated data with two classes were produced with

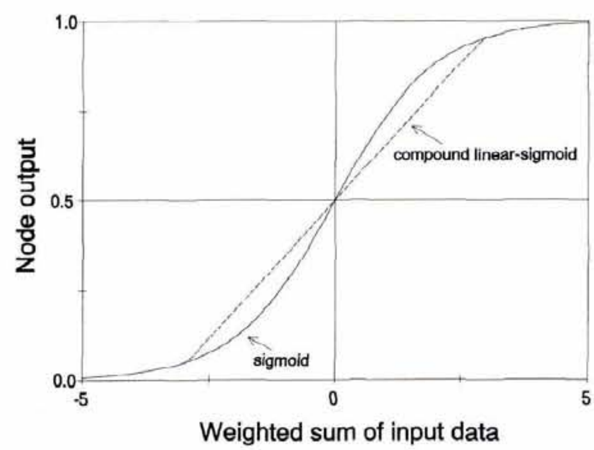

Figure 2. Activation functions. The sigmoid activation function is plotted as a solid line, and the linear section of the compound linear-sigmoid activation function is plotted as a dotted line.

Figure 3. SPOT HRV band 2 (0.61 to 0.68 μm) image of Morgantown, West Virginia, acquired on 16 August 1987 (© SPOT image Copyright 1987 CNES).

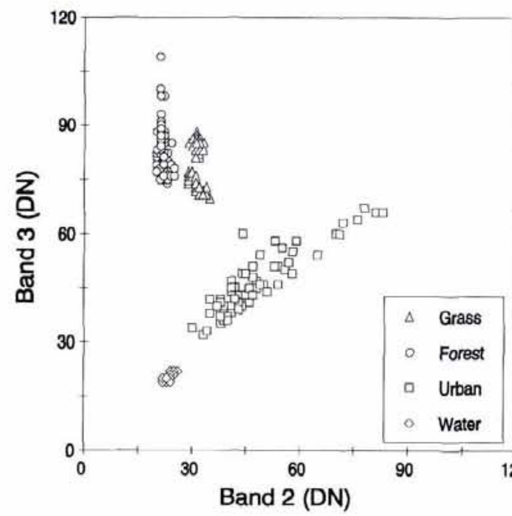

Figure 4. Spectral distribution of training samples for classifying the SPOT HRV image.

separations between the means of 120, 90, 60, 30, and 15 and standard deviations of each class of 1, 2, 3, and 4, for a total of 20 test cases. These 20 test cases have Bhattacharyya Distance values that vary from 1800 to 1.8, with the larger number indicating the greatest separability. Ten repetitions, with entirely new data, were run for each test case to evaluate the variability of results.

A simple linear unmixing model was used to benchmark the accuracy of the neural network classification. The unmixing model estimates class proportion by assuming that all variation between the class means is a consequence of mixing of pure classes equivalent to the class means. Thus, the model determines the proportion of each cover type by calculating for each pixel the Euclidean distance to the class means, and comparing that to the total distance between the means. The model has the obvious constraints that the proportions have to sum to 1.0, and that no single contribution can be less than 0.0 or greater than 1.0. Because the simulated data were created by assuming linear mixing in proportion to the area of the cover type in the mixed pixel, this unmixing model provides an estimate of the best possible accuracy obtainable.

Real Data

The simulated data were particularly useful for illustrating how the class spectral separability constrains the accuracy of the derived fuzzy classification. We used real data, consisting of SPOT HRV imagery of Morgantown, West Virginia (Figure 3), to investigate how the distribution of classes in the spectral space influences fuzzy accuracy. The image is from scene K/J 616/270, acquired on 16 August 1987. Four major cover types dominate the scene: (1) Urban areas, including buildings, roads, and the airport; (2) Water, primarily consisting of the Monongahela River; (3) Forest; and (4) Grass, including pasture and residential lawns. To facilitate visual representation of the neural network behavior in partitioning the spectral space, we only used two of the SPOT HRV bands, namely, bands 2 (0.61 to 0.68 μm) and 3 (0.79 to 0.89 μm). We recognize that incorporating the third SPOT HRV band might have improved our results, but the aim was to investigate fuzzy classification, not to investigate the potential of SPOT data. The SPOT HRV data were classified using the same neural network algorithm as that used for the simulated data. However, for this classification, there were two input nodes for the two SPOT HRV bands, ten hidden nodes, and four output nodes for the four cover classes. The increased number of hidden nodes was required for the real data compared to the simulated data, because of the increased number of classes. The actual number chosen represented a compromise between providing enough input-output connections to associate input and output patterns without significantly compromising processing time. The learning rates, momentum terms, and iterative error reduction threshold were the same as used in the experiments with simulated data. The compound activation function was used, because the experiments with simulated data indicated the superiority of that approach over a sigmoid activation function.

The neural network was trained using two methods. In the first method, the original pure training pixels chosen from the scene (Figure 4) were used, and it was assumed that the neural network would automatically produce fuzzy outputs for mixed pixels. The pure training pixels were chosen based on detailed ground knowledge of the area. They were selected from large homogeneous regions of a single cover type, so as to minimize the chance of a mixed pixel. The second training method used synthetic mixed pixels of various proportions of each cover class, created by applying a linear mixing model to the means of the pure classes. Thus, the synthetic mixed pixel model ignores within class variability, because only the means of the classes are used. These synthetic mixed pixels were used to train the neural network to produce the fuzzy coefficients representing the known proportion of each class, rather than the usual [0,1] output discussed earlier. This approach therefore forces the neural network to partition the spectral space much as the linear unmixing model does.

However, it is apparent from Figure 4 that, in this two-dimensional spectral space, the fuzzy relationship between the Urban and Forest classes is indeterminable. This is because mixtures of those two classes will fall in the same spectral region as pure pixels of the Grass class. It was therefore decided to form two three-class fuzzy training sets describing fuzzy relationships between Water-Urban-Grass and Water-Grass-Forest, as is shown in Figure 5.

Results

Fuzzy Classification of the Simulated Data

A suitable fuzzy partition of the spectral space requires that the partition be in the correct location, and have the appropriate linear shape, as was illustrated in Figure 1. If the shape is inappropriate, error will not be constant as the proportions of the two mixed classes change. Figure 6 shows

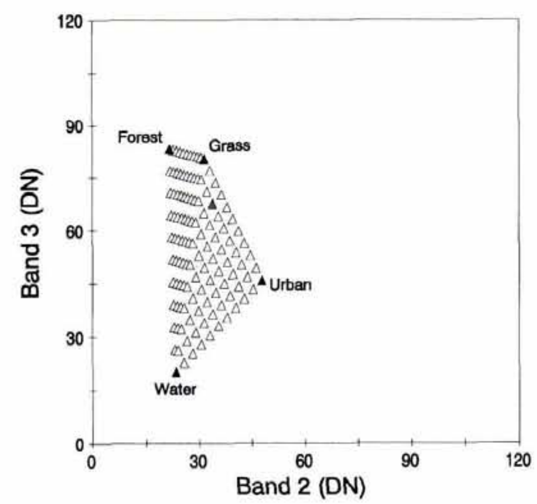

Figure 5. Spectral distribution of class means (solid triangles) and synthetic mixed pixels (open triangles) used in the fuzzy training method. Each location is associated with fuzzy memberships for combinations of Grass, Forest, and Water and combinations of Grass, Water, and Urban. For example, the intermediate gray triangle represents a mixed pixel of the proportion Grass 0.7, Urban 0.2, and Water 0.1.

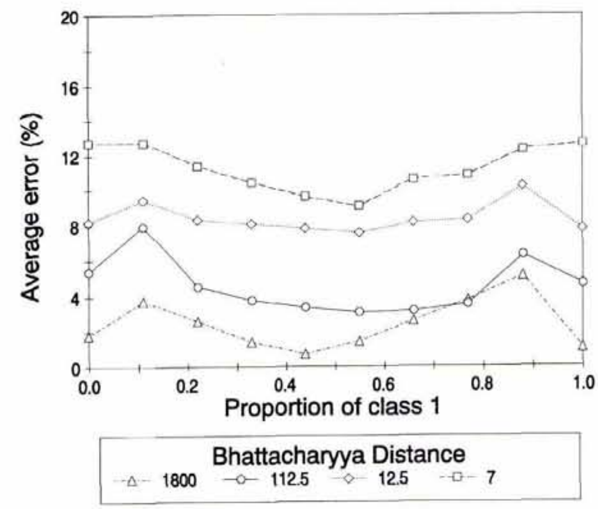

Figure 7. Average error predicting class proportions using a compound linear-sigmoid activation function to classify simulated mixed pixels.

that, for the fuzzy classification using the sigmoid activation function, error peaks where one of the classes comprises 0.20 to 0.25 of the total mixture, giving a distinctive "m" shape to the curves. A comparison of this pattern to Figure 1 suggests that the neural network is generally placing the fuzzy boundary in the appropriate place within the spectral space, but that slope shape is not sufficiently linear.

Another interesting aspect of the curves is that there is a slight tendency for an asymmetric distribution of error between the two classes, especially for the smallest Bhattacharyya Distance values. This is not entirely surprising, because the algorithm is not constrained to produce a symmetric pattern, nor does the output on the nodes have to sum to 1.0. In this case, when the two classes have a large variance relative to the separation of the means (i.e., a low Bhattacharyya Distance), the neural net is responding with a boundary that is no longer placed midway between the two classes.

The classification using the compound linear-sigmoid function (Figure 7) produced a much lower overall error, and much flatter curves compared to the sigmoid activation function. This suggests that the compound activation function more correctly approximates the desired pattern. However, the error is still not uniform across the mixtures, suggesting that further modification of the activation function may be required, particularly on the shoulders of the original sigmoid curve. There is, however, a contradiction in the attempt to reduce error for all proportions. This is because

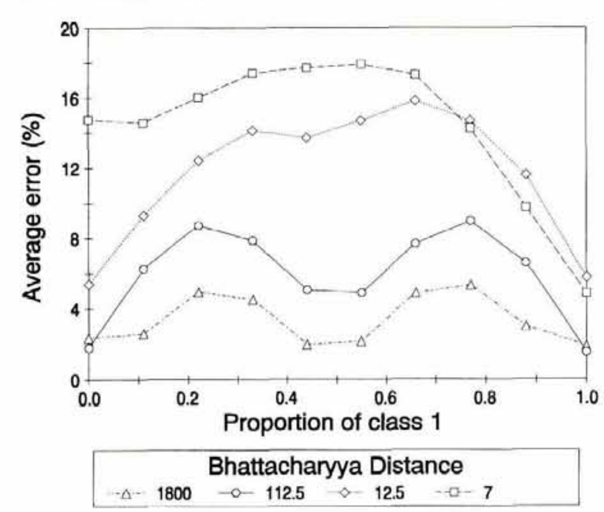

Figure 6. Average error predicting class proportions using a sigmoid activation function to classify simulated mixed pixels.

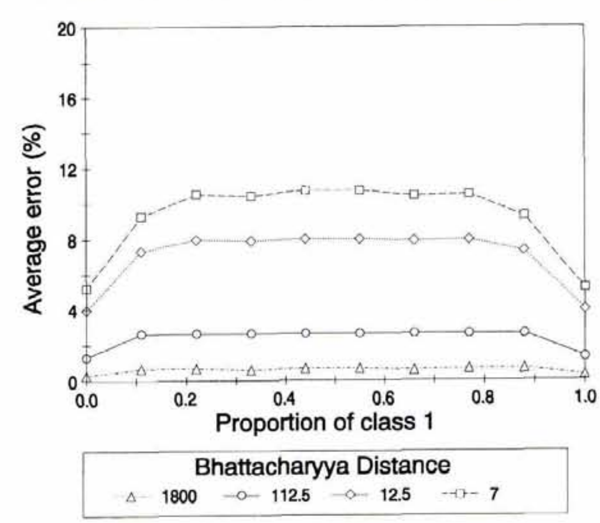

Figure 8. Average error predicting class proportions using a simple linear unmixing model to classify simulated mixed pixels.

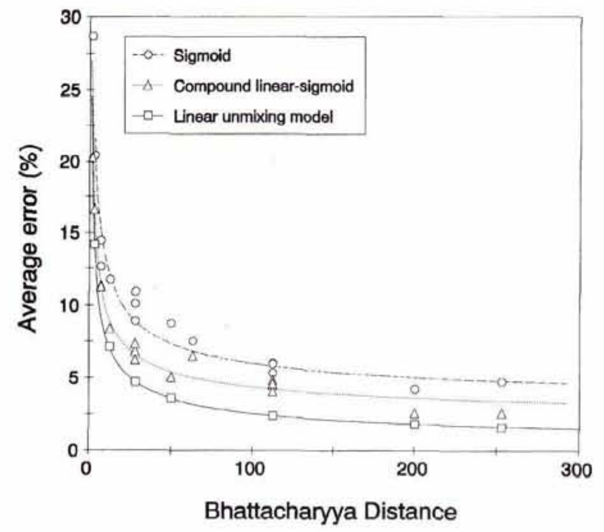

Figure 9. Relationship between overall error and Bhattacharyya Distance between classes.

inevitably, as the partition between the class means becomes more linear, a greater proportion of the inherent variation of *pure* classes will be classified as a result of a *mixture* of cover classes. By comparison, the sigmoid partition assigns a greater region of the classification space around the spectral means to the pure classes, and not to mixtures.

The linear unmixing model (Figure 8) has minimum error for pixels that comprise only one class, or are dominated by one class. This is because the model only allows mixed class assignment for those pixels that fall between the class means. Furthermore, the constraint that one class many not comprise a proportion greater than 1.0, or less than 0.0, limits the total error possible for pixels that fall close to class means. The linear unmixing model produces a lower error on average for all Bhattacharyya Distances compared to the neural network, but the difference is greatest when the classes are well separated. For example, when the Bhattacharyya Distance is 1800, the linear unmixing model produces almost negligible error, whereas the neural network error varies from less than 1 percent to greater than 4 percent.

Figure 9 shows a plot of the average total error against the Bhattacharyya Distance for the experimental data for both the neural network fuzzy classification and the linear unmixing model. Regression analysis suggested that the average error is inversely proportional to the square root of the Bhattacharyya Distance: i.e.,

$$\% \text{ error}_{\text{sigmoid activation function}} = 32.8\ B_{ij}^{-0.5} + 2.7$$
$$\% \text{ error}_{\text{compound activation function}} = 25.1\ B_{ij}^{-0.5} + 1.8$$
$$\% \text{ error}_{\text{linear unmixing model}} = 24.4\ B_{ij}^{-0.5}$$

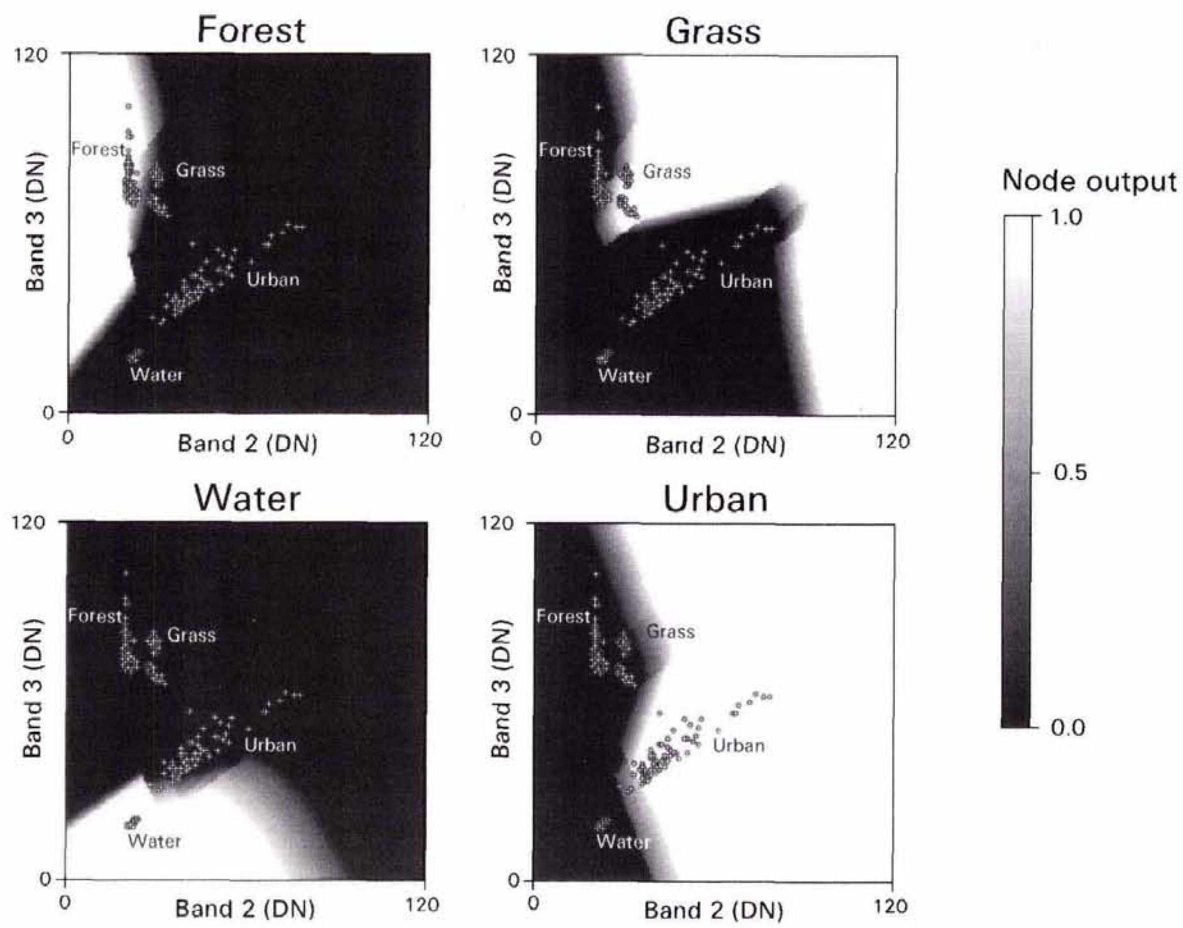

Figure 10. Spectral-space partitions produced by the neural network trained using pixels that represent pure classes.

The r^2 for these relationships was 0.98. These regression equations suggest that the advantage of using the compound function is less than 1.5 percent when the Bhattacharyya Distance is greater than 120. However, for situations in which the classes are moderately to poorly separated, the compound activation function gives a much lower error, almost approaching that of the linear unmixing model. For example, producing a fuzzy relationship with no more than 10 percent average error requires a Bhattacharyya Distance of at least 20 using the sigmoidal activation function, but only 9 for the compound linear-sigmoid function, and 6 for the linear unmixing model.

Real Data

The SPOT HRV image of Morgantown is dominated by four classes, and fuzzy relationships between four classes cannot be represented in two-dimensional data. Indeed, for *n* fuzzy classes, one requires at a minimum $n-1$ bands *and* an optimal arrangement of the mean vectors in the spectral space such that there are no colinear arrangements of three or more spectral classes (Bateson and Curtiss, 1996). However, this geometric argument is a slight simplification because, as the number of bands increases, classes with the same mean vector become increasingly separable, so long as the classes have different covariance matrices (Lee and Landgrebe, 1993). This raises the possibility of fuzzy representation, even in the presence of colinear classes.

An examination of the spectral distribution of the training data for Morgantown (Figure 4) shows other complexities that arise due to the covariance between bands. Compared to the other classes, the Urban class has a large variance in each band, and a high degree of covariance. Furthermore, the Water class lies on the axis of elongation of the Urban cluster, whereas the Grass class lies along the shortest axis of the Urban cluster. Consequently, despite the fact that the means of the Grass and Water classes are approximately equidistant from the Urban class mean (Figure 5) and that the Water class has a small variance (Figure 4), the spectral separability of Urban from Water is likely to be much lower compared to that of Urban from Grass. This is confirmed by the value of the Bhattacharyya Distances of 4 and 12 between these respective pairs of classes. Therefore, it is essential to consider second-order statistics in evaluating class separability and, thus, the potential of fuzzy representation of mixed pixels.

Figure 10 shows the spectral-space partitions produced by the neural network trained using the original pure classes. The images in Figure 10 were created by mapping the output on each of the four nodes for all possible input combinations. Dark tones are low values on that node, and light tones correspond to high values. Thus, for example, the map of the node output for Forest (upper left plot) shows that unknown pixels of very low Band 2 values, and moderate to high Band 3 values (i.e., falling in the bright triangle in the upper left corner) will be classified as Forest.

The mapped output of the nodes demonstrates the ability of a neural network to produce partitions of the spectral space that separate the training data. However, it is also clear that these partitions tend to be simple and relatively linear. Another interesting characteristic of the map of neural network node output is that the spectral space is almost completely partitioned between the various nodes. Sometimes the partitions overlap (for example, between Grass and Urban, and also between Urban and Water), but only a small part of the spectral space between the classes is left unassigned. This is presumably one of the reasons why neural networks are regarded as good generalizers — the boundaries are not narrowly drawn around each class.

Although the partitions separating the four classes might be suitable for indicating a condition of uncertainty (as suggested by Bischof *et al.* (1992) and Civco and Wang (1994)), they are less than optimal for producing useful fuzzy memberships. The intermediate gray areas in Figure 10 are the fuzzy partitions of the spectral space, and it is clear that these fuzzy boundaries are limited in extent, and generally rather too steep.

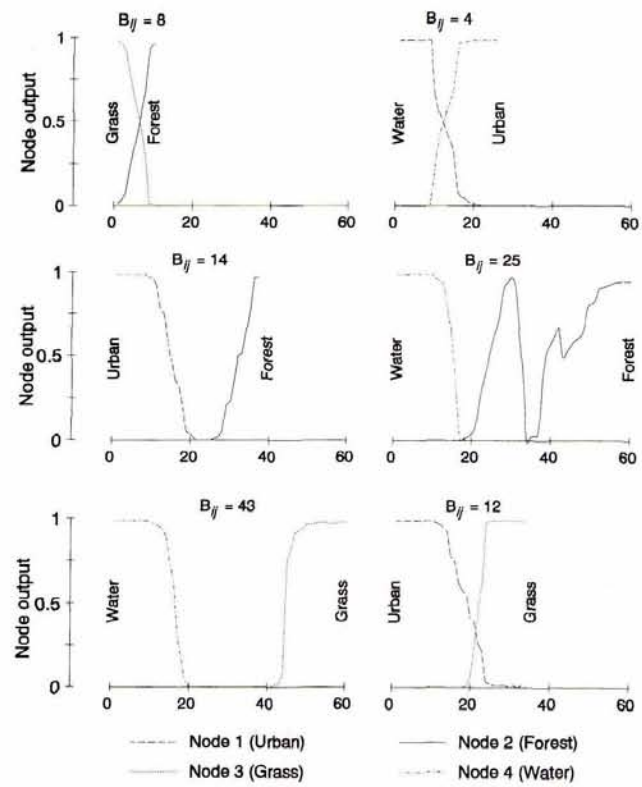

Figure 11. Cross-sections of spectral-space partitions created by the neural network along profile lines drawn between class means, after training using pixels that represent pure classes.

The comparative relationship of node output along mixing lines is explored in Figure 11, which shows the cross-sectional shapes of the partitions obtained by sampling neural network outputs along profile lines drawn between class means. In each case, the Bhattacharyya distance between class means is also indicated on the cross section. The Grass-Forest and Water-Urban partitions approach the general "X-shaped" form considered desirable for characterizing a two-class fuzzy relationship. Notice, however, that the gradients are both very steep due to the low Bhattacharyya Distance between the class means. From the relationships between Bhattacharyya Distance and average error established using the simulated data described in the previous section, the error in the fuzzy representation of Grass-Forest is likely to be at least 10 percent, compared to 14 percent for estimates of the mixtures of Urban and Water classes.

The valley in the profile of Urban-Forest in Figure 11 is expected because of the presence of the Grass cluster in between these two classes. However, for the two profiles from Water (Water-Forest and Water-Grass), the observed valley features are apparently an artifact of the neural network algorithm. Note that the Bhattacharyya Distance between Water and Grass is the largest of any pairs of class means (B_{ij} = 43), with an expected error of only 6 percent. Therefore, although small divergence values indicate low potential for fuzzy esti-

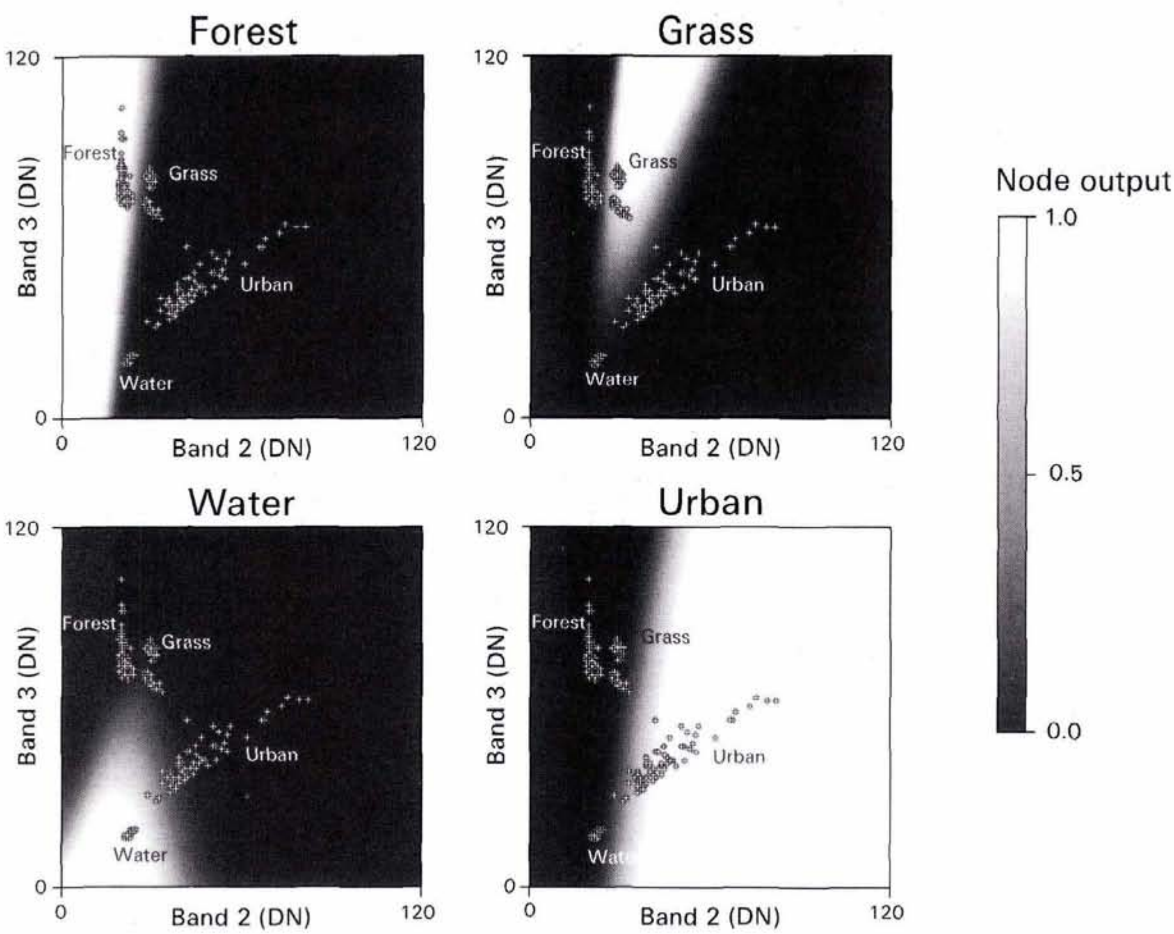

Figure 12. Spectral-space partitions produced by the neural network using the synthetic mixed pixels.

mates of class composition, large divergence values do not necessarily mean that a fuzzy classification is possible, or that the neural net will create an optimal partition between the classes. The complex Water-Forest profile is partly a function of the fact the profile between the class means approximately follows the boundary of the Forest class, as well as the erratic nature of the boundary itself. Again, it is instructive to note that the Bhattacharyya Distance of 25 for this pair of classes is comparatively large.

To overcome these limitations of the neural-network partition of the spectral space, the neural network was then trained with synthetic mixed pixels as described in the Methods section. The mapped output of the nodes, shown in Figure 12, is much smoother than that of Figure 10, and the boundaries are optimally located, as required by the synthetic training data. Notice, however, that the problem of zones of multiple high output on nodes is still present: for example, Urban overlaps with both Grass and Water, and Water also overlaps with Forest. Generally, the profiles of the output of two nodes (Figure 13) now produce a more gentle and symmetric pattern for all class combinations, except, of course, for the Urban-Forest relationship, which is constrained by the presence of the intermediate Grass cluster.

There is an inevitable problem, however, with the fuzzy partition because of the limitations of the data distribution, which has already been raised in the section discussing the synthetic data. In a linear mixing model based on class means, all variance that lies *between* the class means is assumed to arise due to mixing. In comparing the fuzzy partitions in Figure 13 to the original training data of Figures 4 and 5, it can be seen that many of the pure pixels will now incorrectly be mapped as mixed pixels. Only the Water class, with a very tight grouping of pixels, does not suffer much from this problem.

Conclusions

Any attempt to conduct a fuzzy classification should begin by evaluating the sensor and scene constraints that limit the possible fuzzy representation. In this paper, we have illustrated how the separability and arrangement of those classes in the spectral space determine those limits.

Error is also not necessarily uniform for different proportions of the classes within the mixed pixel change. In the case of the neural net fuzzy classification, the error was generally greatest when one of the classes comprises 0.20 to 0.25 of the proportion of the composite pixel, although the pattern tended to vary with class separability.

The plot of overall error against the Bhattacharyya Distance suggests that, to achieve error rates less than 10 percent, a Bhattacharyya Distance value of at least 9 is required. Unfortunately, however, applying this to real data is rather difficult, because separability measures can only be calculated for two classes at a time. Consequently, as the number of classes that are simultaneously unmixed grows, the difficulty of evaluating the separability of all class combinations increases.

For the SPOT HRV data of Morgantown, the large variance in the Urban class, and the position of the Water class along

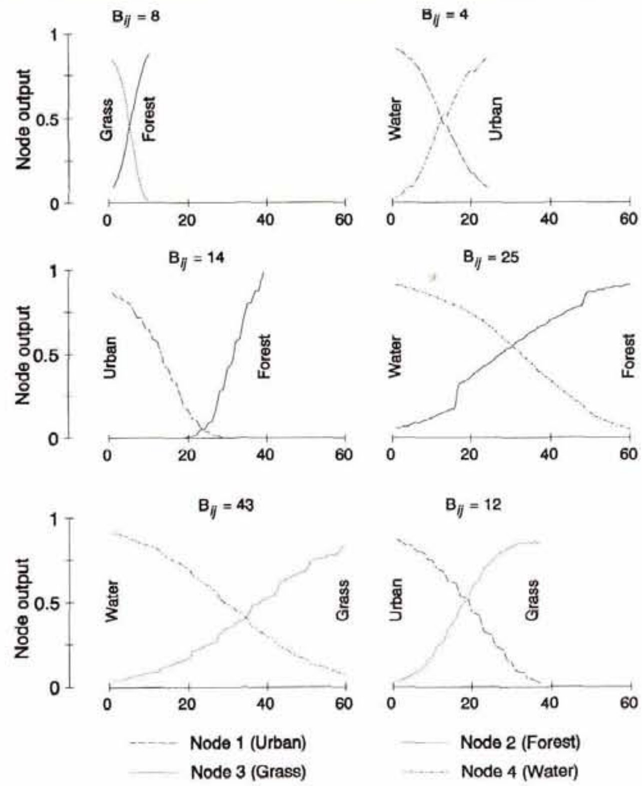

Figure 13. Cross-sections of spectral-space partitions created by the neural network along profile lines drawn between class means after training using synthetic mixed pixels.

the axis of the maximum variance of the Urban class, limits the potential of a fuzzy Water-Urban classification compared to the potential of a fuzzy Urban-Grass classification. Thus, in determining the potential for fuzzy classification, the variance, covariance, and relative positions of class means need to be considered. Even classes with high variance and a relatively small distance between the means may have a high potential for fuzzy classification as long as the classes are appropriately positioned in the spectral space, and there is a high correlation between the bands.

The number of classes, and the arrangement of the class clusters in the spectral space, is also very important in determining what classes can be unmixed. Because there were two more classes than the number of bands, it was not possible to unmix all the classes simultaneously in the SPOT HRV data, and one fuzzy relationship (Forest-Urban) could not be represented. Note that this occurs despite the perfect separation possible between all the training pixels, and the relatively high Bhattacharyya Distance value of 14. Thus, although small separabilities mean that fuzzy representation error is likely to be large, large separabilities do not necessarily mean that highly accurate unmixing is possible. In our example, we were able to evaluate the spectral distribution of classes because we limited the data to only two bands. With higher dimensional data, such as the seven bands of Landsat Thematic Mapper imagery, or the 224 bands of Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) data, such a visual approach will be of much less value! Therefore, for high-dimensional data, it will be necessary to develop mathematical or visual tools that can rapidly screen all classes to see whether any one class can be produced with a linear mixing of any combination of the other classes (Bateson and Curtiss, 1996).

This study also illuminates the potential of neural networks to act as fuzzy classifiers. A neural network tends to produce relatively simple partitions of the spectral space, with steep sides. Even if the classes are very well separated, with high Bhattacharyya Distance values, the partitions are not necessarily midway between the class means, nor are they necessarily symmetric. Although the neural net is trained to produce high values for only one node at a time, there is no constraint that it do this for other combinations of input values not used in the training. Thus, for example, in the Morgantown SPOT HRV data, the spectral partitions for the Grass and Urban classes overlap.

Two methods to improve neural networks for fuzzy classification include modification of the activation function and development of synthetic mixed pixels to train the classifier. A compound linear-sigmoid activation function allows the same activation function to be used both in training the net and in classifying the unknown data. This compound function reduced the overall error, and tended to make the error more uniform between various proportions of mixed classes. The compound activation function was most useful for classes that are moderately to poorly separated. The use of synthetic mixed pixels in training the net tended to produce partitions very similar to that of the linear unmixing model, with linear slopes between the classes and with symmetric shapes. General application of the use of the synthetic mixed pixels for classification will, however, require that the user first check that all classes can be unmixed simultaneously. If the positions of the class clusters in the spectral space do not allow all classes to be unmixed simultaneously, the user will need to specify which combinations should be represented in the synthetic data.

Acknowledgments

The authors would like to thank Dr. David Landgrebe of Purdue University for advice on class separability measures. The authors, however, take full responsibility for the way the class separability measures are used. The helpful suggestions of three anonymous reviewers are gratefully acknowledged.

References

Adams, J.B., M.O. Smith, and A.R. Gillespie, 1989. Simple models for complex natural surfaces: A strategy for the hyperspectral era of remote sensing, *Proceedings of the IEEE International Geosciences and Remote Sensing Symposium 1989*, pp. 16–21.

Bateson, A., and B. Curtiss, 1996. A method for manual endmember selection and spectral unmixing, *Remote Sensing of Environment*, 55:229–243.

Benediktsson, J., P. Swain, and O. Ersoy, 1990. Neural network approaches versus statistical methods in classification of multisource remote sensing data, *IEEE Transactions on Geoscience and Remote Sensing*, 28(4):540–551.

Bezdek, J., 1993. Fuzzy models-what are they, and why? *IEEE Transactions on Fuzzy Systems*, 1(1):1–5.

Bischof, H., W. Schneider, and A. Pinz, 1992. Multispectral classification of Landsat images using neural networks, *IEEE Transactions on Geoscience and Remote Sensing*, 30(3):482–490.

Boardman, J., 1994. Geometric mixture analysis of imaging spectrometry data, *Proceedings of the IEEE International Geosciences and Remote Sensing Symposium 1994*, pp. 2369–2371.

Carpenter, G.A., S. Grossberg, N. Markuzon, J.H. Reynolds, and D.B. Rosen, 1992. Fuzzy ART: A neural network architecture for incremental supervised learning of analog multidimensional maps, *IEEE Transactions on Neural Networks*, 3:598–713.

Civco, D., and Y. Wang, 1994. Classification of multispectral, multitemporal, multisource spatial data using artificial neural networks, *ASPRS/ACSM Annual Convention & Exposition Technical Papers*, Reno, Nevada, 1:123–133.

Decatur, S., 1989. Application of neural networks to terrain classification, *Proceedings of the 1989 International Joint Conference on Neural Networks*, Washington, D.C., 1:283–288.

Foody, G.M., 1996. Relating the land-cover composition of mixed pixels to artificial neural network classification output, *Photogrammetric Engineering & Remote Sensing*, 62(5):491–499.

Foody, G.M., N.A. Campbell, N.M. Trodd, and T.F. Wood, 1992. Derivation and applications of probabilistic measures of class membership from the maximum likelihood classification, *Photogrammetric Engineering & Remote Sensing*, 58(9):1335–1341.

Foody, G.M., and D.P. Cox, 1994. Sub-pixel land cover composition estimation using a linear mixture model and fuzzy membership functions, *International Journal of Remote Sensing*, 15(3):619–631.

Foody, G.M., M.B. McCulloch, and W.B. Yates, 1995. Classification of remotely sensed data by an artificial neural network: Issues related to training data characteristics, *Photogrammetric Engineering & Remote Sensing*, 61(4):391–401.

Fox, L., A. Fischer, A. Gillespie, and M. Smith, 1990. Using spectral mixture analysis of AVIRIS high dimensional data for distinguishing soil chronosequences, *Proceedings of the Second AVIRIS Workshop*, JPL Publication 90–54, pp. 94–98.

Hepner, G., T. Logan, N. Ritter, and N. Bryant, 1990. Artificial neural network classification using a minimal training set: Comparison to conventional supervised classification, *Photogrammetric Engineering & Remote Sensing*, 56(4):469–473.

Heerman, P., and N. Khazenie, 1992. Classification of multispectral remote sensing data using a back-propagation neural network, *IEEE Transactions on Geoscience and Remote Sensing*, 30(1): 540–551.

Key, J., J. Maslanik, and A. Schweiger, 1989. Classification of merged AVHRR and SMMR Arctic data with neural networks, *Photogrammetric Engineering & Remote Sensing*, 55(9):1331–1338.

Kosko, B., 1992. *Neural Networks and Fuzzy Systems: A Dynamical Systems Approach to Machine Intelligence*, Prentice Hall, Englewood Cliffs, New Jersey, 452 p.

Lee, C., and D.A. Landgrebe, 1993. Analyzing high-dimensional multispectral data, *IEEE Transactions on Geoscience and Remote Sensing*, 31(4):792–800.

Lee, J., R. Weger, S. Sengupta, and R. Welch, 1990. A neural network approach to cloud classification, *IEEE Transactions on Geoscience and Remote Sensing*, 28(5):846–855.

Lippmann, R., 1987. An introduction to computing with neural nets, *IEEE ASSP Magazine*, (April):4–22.

Maselli, F., C. Conese, and L. Petkov, 1994. Use of probability entropy for the estimation and graphical representation of the accuracy of maximum likelihood classifications, *ISPRS Journal of Photogrammetry and Remote Sensing*, 49(2):13–20.

Ray, T.W., and B.C. Murray, 1996. Nonlinear spectral mixing in desert vegetation, *Remote Sensing of Environment*, 55:59–64.

Richards, J., 1993. *Remote Sensing Image Data Analysis*, Springer Verlag, New York, 340 p.

Resmini, R.G., M.E. Kappus, W.S. Aldrich, J.C. Harsanyi, and M. Anderson, 1996. Use of Hyperspectral Digital Imagery Collection Experiment (HYDICE) sensor data for quantitative mineral mapping at Cuprite, Nevada, *Eleventh Thematic Conference on Applied Geologic Remote Sensing*, Las Vegas, Nevada, 27–29 February, 1:48–65.

Rich, E., and K. Knight, 1994. *Artificial Intelligence*, McGraw-Hill, New York, 740 p.

Ruck, D.W., S.K. Rogers, M. Kabrisky, M. Oxley, and B. Suter, 1990. The multilayer perceptron as an approximation to a Bayes discriminant function, *IEEE Transactions on Neural Networks*, 1(4): 296–298.

Rumelhart, D., G. Hinton, and R. Williams, 1986. Learning internal representations by error propagation, *Parallel Distributed Processing, Volume 1: Foundations* (D. Rumelhart and J. McClelland, editors), MIT Press, Cambridge, Massachusetts, pp. 318–362.

Schalkoff, R., 1992. *Pattern Recognition: Statistical, Structural and Neural Approaches*, Wiley, New York, 364 p.

Smith, M.O., S.L. Ustin, J.B. Adams, and A.R. Gillespie, 1990. Vegetation in deserts: I. A regional measure of abundance from multispectral images, *Remote Sensing of Environment*, 31:1–26.

Swain, P., and S. Davis, 1978. *Remote Sensing: The Quantitative Approach*, McGraw Hill, New York, 369 p.

Wang, F., 1990. Fuzzy supervised classification of remote sensing images, *IEEE Transactions on Geoscience and Remote Sensing*, 28(2):194–201.

Warner, T.A., D.W. Levandowski, R. Bell, and H. Cetin, 1994. Rule-based geobotanical classification of topographic, aeromagnetic and remotely sensed vegetation community data, *Remote Sensing of Environment*, 50:41–51.

Woodcock, C.E., and A.H. Strahler, 1987. The factor of scale in remote sensing, *Remote Sensing of Environment*, 21:311–332.

Zhuang, X., B. Engel, M. Baumgardner, and P. Swain, 1991. Improving classification of crop residues using digital land ownership data and Landsat TM imagery, *Photogrammetric Engineering & Remote Sensing*, 57(11):1487–1492.

(Received 11 July 1996; accepted 9 December 1996; revised 23 January 1997)