Classification of Remotely Sensed Data by an Artificial Neural Network: Issues Related to Training Data Characteristics

Giles M. Foody, Mary B. McCulloch, and William B. Yates

Abstract

Artificial neural networks have considerable potential for the classification of remotely sensed data. In this paper a feed-forward artificial neural network using a variant of the back-propagation learning algorithm was used to classify agricultural crops from synthetic aperture radar data. The performance of the classification, in terms of classification accuracy, was assessed relative to a conventional statistical classifier, a discriminant analysis. Classifications of training data sets showed that the artificial neural network appears able to characterize classes better than the discriminant analysis, with accuracies of up to 98 percent observed. This better characterization of the training data need not, however, translate into a significantly more accurate classification of an independent testing set. The results of a series of classifications are presented which show that in general markedly higher classification accuracies may be obtained from the artificial neural network, except when a priori *information on class occurrence is incorporated into the discriminant analysis, when the classification performance was similar to that of the artificial neural network. These and other issues were analyzed further with reference to classifications of synthetic data sets. The results illustrate the dependency of the two classification techniques on representative training samples and normally distributed data.*

Introduction

Supervised image classification is one of the most widely used procedures in the analysis of digital remotely sensed data. A wide range of supervised classifiers are available but all share a common objective, to allocate each case of unknown class membership to a pre-defined class on the basis of its spectral properties. The accuracy with which a remotely sensed image may be classified is dependent on many factors. In addition to the properties of the remotely sensed data set, the latter include the characteristics of the training data (Swain, 1978; Schneider, 1980; Campbell, 1981; Foody, 1988) and the nature of the classifier (Mather, 1987a; Thomas *et al.*, 1987; Schalkoff, 1992). These latter issues are often important because widely used conventional statistical classifiers are based on a range of assumptions about the data. For instance, the maximum-likelihood classification, one of the most widely used image classification techniques, assumes that the data for each class display a Gaussian normal distribution. This assumption is often untenable with remotely sensed data where classes may display a range of distributions. Furthermore, because class membership is unknown—it is the objective of the classification to define it—correction for non-normality is impossible. Even if the assumptions regarding the data distribution are satisfied, the performance of the classification will be dependent on the quality of the training data used. In order to form a representative training set for an individual class, it is generally recommended that the size of the training sample should be at least 30 times the number of features (e.g., wavebands) for each class (Swain, 1978; Mather, 1987a), yet such a training set is frequently not acquired.

To avoid some of the problems associated with the maximum-likelihood classification, users may adopt more robust techniques. For instance, discriminant analysis is in many ways similar to the maximum-likelihood classification but is less sensitive to deviations from normality and other assumed conditions (Tom and Miller, 1984). Other alternative procedures which make no assumption about the statistical distribution of the data may also be used. A wide range of techniques are available, including non-parametric classifiers (Skidmore and Turner, 1988) and fuzzy classifiers (Kent and Mardia, 1988; Key *et al.*, 1989; Wang, 1990). Recently, attention has focused on artificial neural network techniques as a distribution-free approach to image classification (e.g., Benediktsson *et al.*, 1990; Liu and Xiao, 1991).

Artificial neural networks have a wide range of applications (Davalo and Naïm, 1991) and have been used to classify remotely sensed data to accuracies that are generally comparable to or higher than those derived from conventional statistical classifications (Hepner *et al.* 1990; Medina and Vasquez, 1991; Short, 1991). In essence, an artificial neural network may be considered to comprise a relatively

---

G. M. Foody and M. B. McCulloch are with the Department of Geography, University College of Swansea, Singleton Park, Swansea SA2 8PP, United Kingdom.
W. B. Yates is with the Department of Mathematics and Computer Science, University College of Swansea, Singleton Park, Swansea SA2 8PP, United Kingdom.

Photogrammetric Engineering & Remote Sensing,
Vol. 61, No. 4, April 1995, pp. 391–401.

0099-1112/95/6104-391$3.00/0

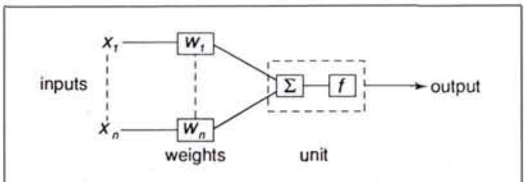


Figure 1. A typical artificial neural network unit. The output of the unit is a function of the net input to it which is a composite of the input strengths (x_i) and the connecting weights (w_i).

large number of simple interconnected neurons or units that work in parallel to categorise input data into output classes (Hepner *et al.*, 1990; Schalkoff, 1992). The interconnections between the units are weighted and, with the input data, these weights determine the level of activation of a unit in the network, which in turn influences the level of activation of other units in the network and ultimately determines the network outputs (Figure 1). The magnitude of the weights is determined by an iterative training procedure in which the network repeatedly tries to learn the correct output for each of the training samples. The procedure involves modifying the weights between units until the artificial neural network is able to characterize the training data accurately. Conventionally, a backpropagation learning algorithm has been employed in which the error between the network output and desired output is minimized (Davalo and Naïm, 1991; Schalkoff, 1992). Once trained, the artificial neural network may then be used to determine class membership for other data. The network architecture parameters influence the performance of an artificial neural network (Benediktsson *et al.*, 1990; Hepner *et al.*, 1990; Lee *et al.*, 1990) but may be difficult to define and are typically determined subjectively. Once defined, however, the artificial neural network is more robust than conventional statistical classifiers (Hepner *et al.*, 1990). Artificial neural networks are also more tolerant to noise and missing data (Hepner *et al.*, 1990), can adapt over time (Short, 1991), weight the importance of the data in the classification (Benediktsson *et al.*, 1990), and, once trained, can be more efficient computationally than conventional classifiers, especially if run on a parallel processing system (Lee et al., 1990).

As with the conventional statistical classifications, the characteristics of the training data will influence classification performance. The size of the training set, for instance, may be important in determining the relative accuracy of classifications derived from artificial neural networks and conventional statistical image classifications (Benediktsson *et al.*, 1990; Hepner *et al.*, 1990). It may be possible for artificial neural networks to classify a remotely sensed data set accurately with a smaller training set than is required for a conventional classification (Hepner *et al.*, 1990) and, if so, then substantial advantages are afforded to the user; the ability to use small training sets and the absence of data distribution assumptions. This is important because, if the remotely sensed data satisfy the assumptions of conventional statistical classifiers, then no significant difference in classification accuracy is usually observed between conventional and neural network classifications (Ryan *et al.*, 1991). Without any further advantage over conventional statistical classifiers, artificial neural networks would therefore only be attractive for the classification of data which showed markedly non-normal distributions. In this paper we evaluate the performance of an artificial neural network against a discriminant analysis with particular attention to the characteristics of the training data used in the classification.

Data

The classifiers were evaluated using two data sets. First, X-band HH-polarized synthetic aperture radar (SAR) data for a region of flat agricultural land centered on Feltwell, United Kingdom (Figure 2) and second, a simulated data set. The latter data set will be discussed later in the paper.

The SAR data were acquired on four dates during the 1986 growing season for an approximately 100km^2 test site by the VARAN-S SAR as part of the European AgriSAR campaign (Figure 3). These data suffer from a range of radiometric distortions (Quegan *et al.*, 1991) and, consequently, required preprocessing. The major radiometric distortions were reduced by deleting duplicated lines and radiometric balancing of the imagery, and the tone of the images was standardized by a simple inter-image regression (Foody *et al.*, 1989). Because these data were to be generalized to produce a nominal level land-cover map, and the temporal dimension of the data was not being utilized further, more detailed corrections were not considered necessary. The analyses aimed to classify crop type on a per-field basis (Pedley and Curran, 1991) using image tone (DN) as the discriminating variable. The field mean DN from each of the four images was estimated from a sample of at least 10,000 pixels, which were located away from the field edges to

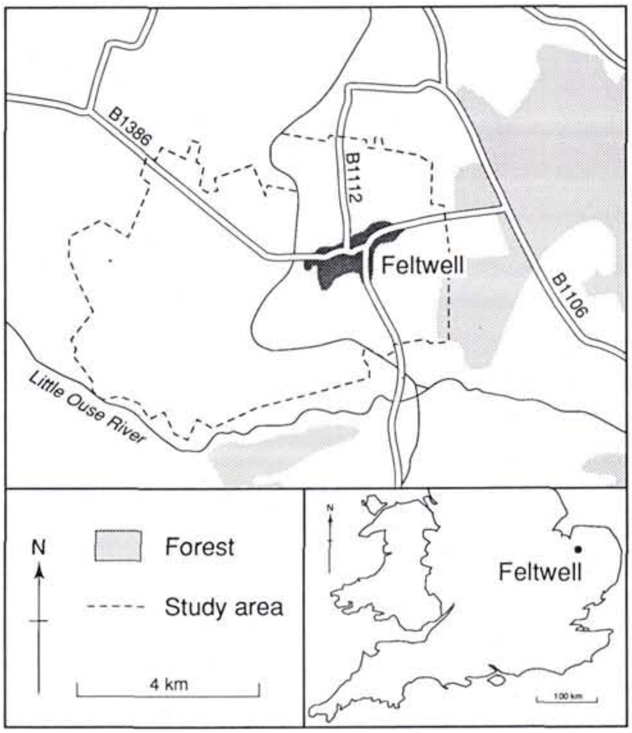


Figure 2. Location of the Feltwell test site in the United Kingdom.

Figure 3. SAR image (uncorrected) of the test site recorded on 28 June 1986.

avoid boundary effects, for a total of 158 fields (Foody *et al.*, 1989). These fields had been planted to one of seven crop classes; winter wheat (42), spring barley (42), sugar beet (42), potatoes (12), carrots (7), spring wheat (5), and grass (8).

The Classifiers

The two classification techniques used in the analyses were a discriminant analysis (Klecka, 1980) and a feedforward artificial neural network using a variant of the backpropagation learning algorithm (Yates *et al.*, 1993). The discriminant analysis, like the maximum-likelihood classification, allocates each case to the class with which it has the highest *a posteriori* probability of membership (Klecka, 1980; Tom and Miller, 1984), and has been widely used in the classification of remotely sensed data. Although the discriminant analysis is a relatively robust technique (Tom and Miller, 1984), an artificial neural network may be a more attractive approach to classification when the data show a marked deviation from normality.

An artificial neural network is constructed from a set of processing units interconnected by weighted channels according to some architecture. Each unit consists of a number of input channels, an activation function, and an output channel. Signals impinging on a unit's inputs are multiplied by the channel's weight and are summed to derive the net input to that unit. The net input (net) is then transformed by the activation function (f) to produce an output for the unit (Figure 1). This may be expressed as

$$\text{net} = \sum_{i}^{n} x_i w_i$$

$$\text{output} = \text{f(net)}$$

where x_i is the magnitude of the i^{th} input and w_i is the weight of the interconnection channel.

In order to apply the backpropagation learning algorithm, the activation function must be continuous, differentiatable, and non-linear. Frequently, a sigmoid function is employed (Rumelhart *et al.*, 1986): i.e.,

$$\text{f}:\mathbf{R} \rightarrow [0,1]$$

$$\text{f(net)} = \frac{1}{1 + e^{-\lambda net}}$$

where λ is a gain parameter which is often, and is in this paper, set to 1 (Schalkoff, 1992). The activation function employed here is symmetric, and such functions have been shown to improve learning algorithm performance over the sigmoid function (Stornetta and Huberman, 1987). Specifically, the hyperbolic tangent function, tanh, was used. In this paper, a fully connected layered feedforward architecture with one hidden layer and an external bias unit was used (Figure 4). The units of the network are grouped into layers, with the output of a layer feeding forward to the next layer. The output of the network as a whole is simply the activation of the units in the final layer.

The backpropagation learning algorithm (Rumelhart *et al.*, 1986) has been widely used in pattern recognition applications of artificial neural networks; it iteratively minimizes an error function over the network outputs and a set of target outputs, taken from a training data set. The process continues until the error value converges to a (possibly local) minima. Conventionally, the error function is given as

$$E = 1/2 \sum (T_i - O_i)^2$$

where T_i is the target output vector for the training set $(T_1, \ldots, T_n)$ and O_i is the output vector from the network for the given training set. On each iteration backpropagation recursively computes the gradient or change in error with respect to each weight in the network, $\partial E/\partial w$, and these values are used to modify the weights. Adding a fraction of the negative gradient to each weight is equivalent to performing a steepest descent minimization of the error function with respect to each weight in the network (Rumelhart *et al.*, 1986). Because standard backpropagation tends to be slow and does not scale up to larger problems well, faster variants of backpropagation are widely used. The algorithm employed in this paper is Fahlman's (1988) Quickprop algorithm with Almeida and Silva's (1990) adaptive learning rate. In essence, Quickprop is a second-order gradient descent technique based on Newton's method (Fahlman, 1988). The algorithm recursively computes the error derivative with respect to each weight in the network as in conventional backpropagation (Rumelhart *et al.*, 1986). However, for each weight, Quickprop uses a copy of the previous gradient and weight update, in conjunction with the current error gradient, to determine a parabola of error versus weight strength. By employing the quadratic update formula,

$$\Delta w(t) = \frac{\dfrac{\partial E}{\partial w}}{\dfrac{\partial E}{\partial w(t-1)} - \dfrac{\partial \mathbf{E}}{\partial \mathbf{w}(\mathbf{t})}} \cdot \Delta \mathbf{w}(t-1)$$

where t indicates the iteration number, and other heuristics, Quickprop tries to jump to the minima of the parabola. Although the derived parabola is only an approximation to the true error surface seen by each weight, the procedure when applied iteratively has been shown to be more effective than

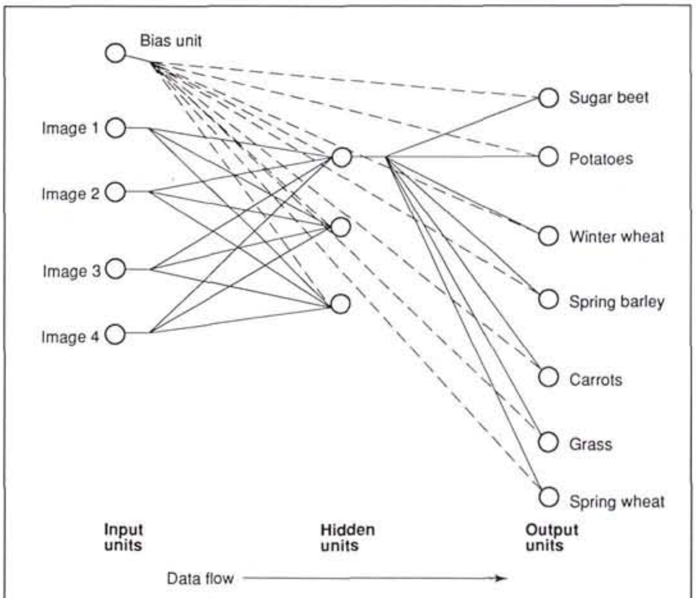


Figure 4. Basic structure of the 4:3:7 artificial neural network used. For clarity, some of the network's connections have been omitted.

standard backpropagation (Fahlman, 1988). An extension proposed by Almeida and Silva (1990) employs an individual learning rate for each weight in the network. These learning rates are adapted during training according to the heuristic that, if the previous and current gradients have the same sign, that is, direction, the local learning rate is increased. If the gradients have different signs, the learning rate is decreased. Such a heuristic allows each learning rate to tune itself during training and has been shown to reduce zigzagging on the error surface and therefore learning times. (Almeida and Silva, 1990). The network architecture and learning algorithm parameters used were tuned following a number of learning trials. The architecture selected contained as few weights as possible while still allowing significant learning to occur. Such an approach not only decreases learning algorithm run times but also increases the artificial neural networks' potential for generalization (Baum, 1990). A full exposition of the algorithm employed here is beyond the scope of this article but may be found in Yates *et al.* (1994).

Classification Accuracy Assessment and Comparison

A statement of classification accuracy is required for the evaluation of classification performance. There are many approaches for the assessment of image classification accuracy and, because there are no general rules which can be followed to define an appropriate classification accuracy technique (Congalton, 1991), one simple, easy to interpret, and widely used measure, the percentage correct allocation, was used. However, the classification confusion matrices are also presented, enabling the calculation of other measures of classification accuracy if desired (Story and Congalton, 1986; Foody, 1992). For clarity, the main diagonal of the classification confusion matrices, which illustrates the correct class allocations, are highlighted.

The significance of the difference in classification accuracy observed between the artificial neural network and discriminant analysis classifications was assessed, on the assumption that the samples may be considered independent, with the difference between proportions test (Freund and Williams, 1959; Clark and Hosking, 1986).

Results and Discussion

For the classifications of the SAR data, a three-layer artificial neural network architecture was used, comprising four input units, three hidden units, seven output units, and a bias unit (Figure 4). The SAR data were divided into training and testing sets for a series of classifications. For the initial set of analyses, the training set comprised 122 fields drawn from all seven classes. The testing set, however, comprised 36 fields of only the three most predominant crop types found in this region; spring barley, sugar beet, and winter wheat. To assess the level of inter-class separability and illustrate the relative ability of the artificial neural network and discriminant analysis to identify or "learn" the characteristic appearance of a class, the accuracy with which both could classify the training data was assessed. The results showed that a significantly higher classification accuracy was derived from the artificial neural network classification than from the discriminant analysis (Figure 5). This result indicates that the artificial neural network is able to learn class appearance more accurately than does the discriminant analysis.

Once trained, both classifiers were then used to classify the independent testing set. As expected, the accuracy of the classifications of the independent testing data sets were lower than those derived from the training classifications (Swain, 1978). As with the training classification, the artificial neural network was able to classify the data to a significantly higher accuracy than the discriminant analysis (Figure 5).

The results above were derived under the assumption that each class had an equal *a priori* probability of occurrence. Often this is not the case in remote sensing and, for the data set under investigation, there was a marked range in class occurrence, with three classes covering most of the test site. If this information on class occurrence is known before the analysis, it may be used to increase classification accuracy (Strahler, 1980). Incorporating prior probabilities, defined in direct proportion to the number of fields in each class in the training set, into the discriminant analysis resulted in an increase in classification accuracy (Figure 6). It is evident, however, that the classifications derived from the artificial neural network were slightly, but not significantly, more accurate than those from the discriminant analysis even with the inclusion of prior information for a training classification. With the classification of the testing data, however, the discriminant analysis with prior information was able to classify the data to the same accuracy as the artificial neural network. It was also apparent that the distribution of class allocation evident in the confusion matrices derived from the discriminant analysis with *a priori* information was similar to that from the artificial neural network. These results, therefore, appear to indicate that the artificial neural network is superior to the discriminant analysis in terms of characterizing class appearance and so indicating a high potential for accurate inter-class discrimination. This may not, however, always translate into a more accurate classification of an independent testing set if prior information on class occurrence is available to the discriminant analysis. The use of prior information within an artificial neural network is more

DA (PRIORS = EQUAL)

TRAINING

Predicted

Actual	SpB	SuB	WW	SW	P	C	G
SpB	28		1			1	
SuB	1	24			4		1
WW	1		8	11		6	4
SW	1			4			
P		1			11		
C	1		1	1	1	1	2
G	1		1			2	4

Accuracy = 65.57%

TESTING

Predicted

	SpB	SuB	WW	SW	P	C	G
SpB	8		1	1	1	1	
SuB		9			3		
WW			5	5		2	

Accuracy = 61.11%

ANN

Actual	SpB	SuB	WW	SW	P	C	G
SpB	29		1				
SuB	1	28			1		
WW	1		29				
SW			5				
P	2	2			8		
C	2		2		1	1	1
G	1		2				5

Accuracy = 81.97%

	SpB	SuB	WW	SW	P	C	G
SpB	10		2				
SuB		10			2		
WW	2		10				

Accuracy = 83.33%

Figure 5. Confusion matrices for the classifications of the SAR data by a discriminant analysis (DA) with equal prior probabilities of class occurrence (priors=equal) assumed and by the artificial neural network (ANN). Note that the training set comprised seven classes and the testing set three classes. The class codes are: SpB - Spring barley; SuB - Sugar beet; WW - Winter wheat; SW - Spring wheat; P - Potatoes; C - Carrots; G - Grass.

DA (PRIORS = SIZE)

TRAINING

Predicted

Actual	SpB	SuB	WW	SW	P	C	G
SpB	28		2				
SuB	1	27			1		1
WW	1		29				
SW	1		4				
P	2	1			9		
C	1		2		1	1	2
G	1		3				4

Accuracy = 80.33%

TESTING

Predicted

	SpB	SuB	WW	SW	P	C	G
SpB	8		3		1		
SuB		11			1		
WW	1		11				

Accuracy = 83.33%

Figure 6. Confusion matrices for the classifications of the SAR data by a discriminant analysis (DA) with *a priori* information of class occurrence. Prior probabilities were defined in direct proportion to the size of each class in the training set (priors=size).

problematic and is the focus of on-going research with present emphasis on the effect of training set size on the opportunity to learn class appearance (Foody *et al.*, 1992b; 1993).

Another important issue is the reaction of the classification to cases that are members of an untrained class. Frequently, in a supervised image classification, the image will contain classes for which a training set has not been acquired. For instance, an image used for a crop classification may contain classes of no interest to the user such as forest, water, or urban land for which training sets were not defined. Cases of these untrained classes must, however, be allocated to one of the trained classes by the classification. The effect of untrained classes on the performance of the two classifiers was assessed for a data set in which the training set comprized 90 cases of the three most predominant classes but was used to classify a testing test containing 68 cases of all seven classes (Figure 7). As with the previous classifications, the artificial neural network classified the training data to a higher accuracy than did the discriminant analysis. It was evident, however, that the discriminant analysis clas-

DA (PRIORS = EQUAL)

TRAINING (Predicted)

Actual	SpB	SuB	WW
SpB	28		2
SuB	2	28	
WW	1		29

Accuracy = 94.44%

TESTING (Predicted)

	SpB	SuB	WW
SpB	9		3
SuB		12	
WW	1		11
SW	1		4
P	1	11	
C	2		5
G	1		7

Accuracy = 47.06%

ANN

Actual	SpB	SuB	WW
SpB	30		
SuB		30	
WW	1		29

Accuracy = 98.89%

	SpB	SuB	WW
SpB	8	1	3
SuB		12	
WW	2		10
SW		1	4
P		12	
C	1	1	5
G	1		7

Accuracy = 44.12%

Figure 7. Confusion matrices for the classifications of SAR data by discriminant analysis (DA) and artificial neural network (ANN). Note that here the training set comprised three classes but the testing set contained seven classes, four of which were untrained classes.

sified the testing set to a marginally, but not statistically significant, higher accuracy than the artificial neural network, and inspection of Figure 7 shows that the difference in classification accuracy was not caused by cases of the untrained classes. The value of the classifications could be improved if untrained classes were identifiable. By outputting the relative strength of class membership in addition to the most likely class of membership, it may be possible to identify cases that may be members of untrained classes. Measures of the strength of class membership can be derived from a range of classifiers and used to improve the value of the classification (e.g., Foody *et al.*, 1992b). The outputs of an artificial neural network classifier, computing over the real numbers, may be interpreted as such a measure of class membership, with the convention that the output with the largest value is taken as the predicted class of membership. The potential of this measure for identifying members of untrained classes was assessed with reference to the strengths of class membership derived from a previous classification, the training classification of 122 cases drawn from all seven classes. Analysis of the relative strength of membership each case displayed to each class revealed that a case typically showed a high strength of membership to its allocated class and a low strength to others (Figure 8). This indicated that the artificial neural network provided a fairly hard allocation, like the discriminant analysis (Table 1), and that this measure of the strength of class membership may not be useful for the identification of cases of untrained classes.

The ability of the classifiers to accurately determine class membership for an independent testing set will be dependent on the quality of the description of class appearance generated in the training stage. Two key issues related to this are the representativeness of the training data and their statistical distribution. While the latter should only affect the performance of the discriminant analysis because the artificial neural network is a distribution free technique (Medina and Vasquez, 1991), it is essential for all classifiers that the training data are representative of the class from which they are drawn if they are to be of value in determining class membership for cases of previously unknown class. The effects of data distribution and representativeness were assessed in a series of classifications using a simulated data set. The latter comprised four classes, A, B, C, and D, each with 100 cases. Each class was generated with a normal distribution, the characteristics of which are given in Table 2; normality was assessed by the Kolmogorov-Smirnov test. The data for each class were then split into training and testing sets comprising 33 and 67 cases, respectively. The architecture of the artificial neural network was modified from that shown in Figure 4 to one with one input, three hidden, and four output units, a 1:3:4 structure.

As with the analyses of the SAR data, the accuracy with which the training and testing data could be classified was assessed for all classifications. In the first set of analyses, the training data for each class were sampled to ensure that they were representative of the class from which they were drawn (Table 3); representativeness was assessed by the Mann-Whitney *U* test. Classifications of the training and testing data by the discriminant analysis and artificial neural network were found to be of comparable accuracy, although the artificial neural network classifications were marginally more accurate (Figure 9). Keeping the size of the training and testing sets the same but taking a non-representative sample (Table 4) to form the training set, the classifications were repeated; here an unrepresentative sample is considered to be one drawn from the population but having a mean that is dissimilar to that of the population at the 95 percent level of confidence. Again, the performance of both classifiers was fairly similar, but there was a marked difference in the accuracy with which the training and testing sets were classified (Figure 10). This result is to be expected because the training set for each class would not be representative of that class in the testing set. It was apparent that both classifiers were affected similarly by unrepresentative training data, and users may therefore need to exercise care in selecting training sites and ensure, for instance, that atypical cases are excluded from training sets through the application of a training set purification procedure (Mather, 1987b; Arai, 1992). It may be possible to increase the accuracy of the artificial neural network classification by including a "noise" term which lessens the sensitivity of the classification to the exact characteristics of the training data set (Holmstrm and Koistinen, 1992).

Thus far, the simulated data for each class had displayed a normal distribution. To assess the effect of data non-normality on the relative performance of the classifiers, the data for each class were deliberately skewed through the application of power functions (Table 5). These non-normal data were then used in a set of classifications using representative and non-representative training samples as before.

With a representative sample of the non-normally distributed data used for training both classifiers, the accuracy with which the artificial neural network and discriminant analysis could classify the training and testing data were assessed. The classifications were less accurate than the com-

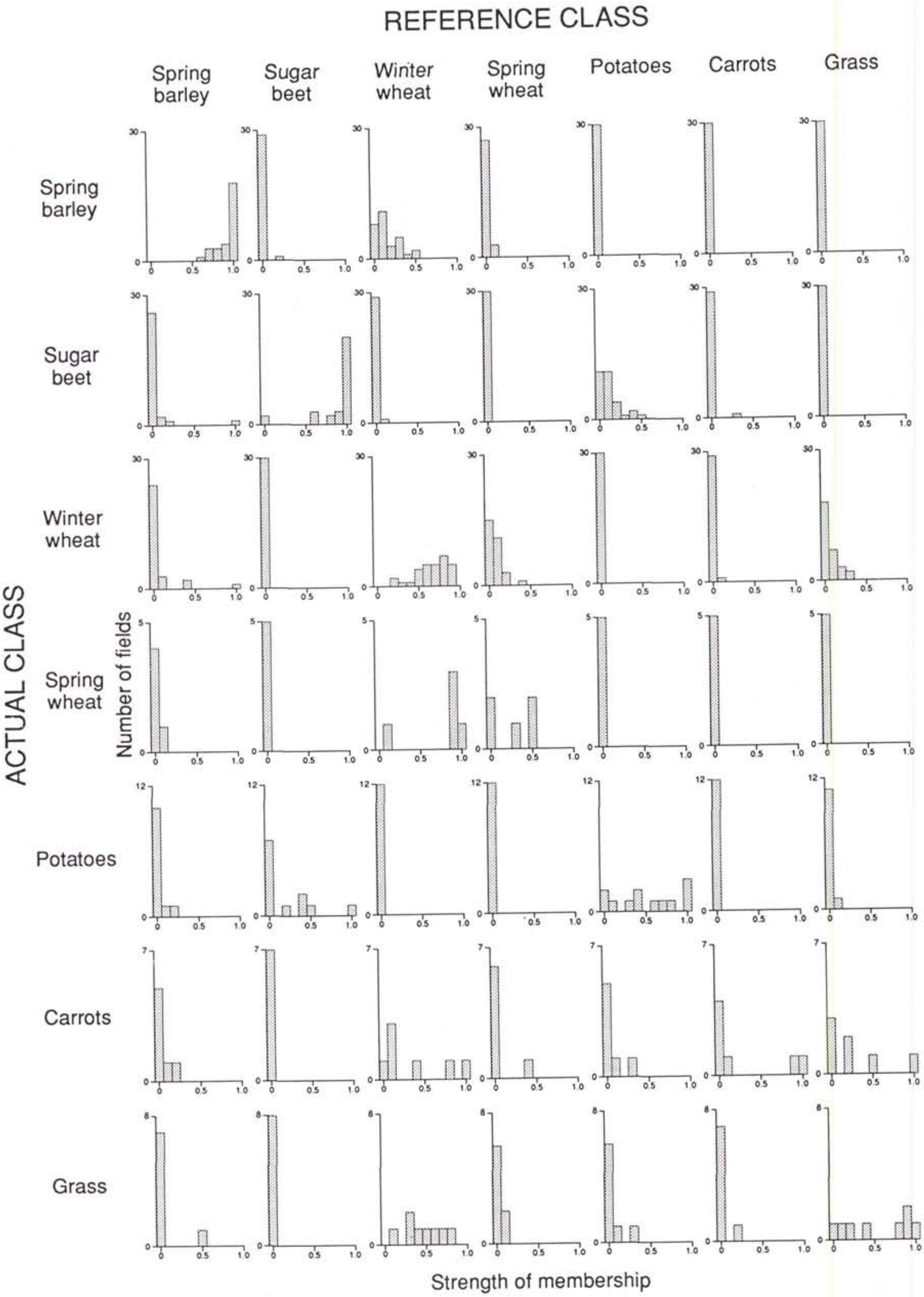


Figure 8. Variations in the strength of class membership derived from the artificial neural network for all cases to each class. Derived from the training classification of 122 cases of seven classes.

TABLE 1. *A POSTERIORI* PROBABILITIES OF CLASS MEMBERSHIP DERIVED FROM THE TRAINING CLASSIFICATION OF 122 CASES OF SEVEN CLASSES BY THE DISCRIMINANT ANALYSIS. NOTE THAT, AS EXPECTED, THE CLASSIFICATION IS RELATIVELY HARD.

Actual class	*A posteriori* probabilities of class membership									
	Cases allocated correctly			Casses allocated incorrectly						
	To actual class of membership			To actual class of membership			To predicted class of membership			
	mean	min	max	mean	min	max	mean	min	max	Number of cases allocated incorrectly
Spring barley	0.8058	0.4259	0.9990	0.0906	0.0408	0.1403	0.3886	0.3433	0.4339	2
Sugar beet	0.9067	0.5905	0.9952	0.2369	0.0000	0.4899	0.6490	0.4943	0.8684	6
Winter wheat	0.4114	0.3706	0.4593	0.2559	0.0331	0.3540	0.4952	0.3338	0.9242	22
Spring wheat	0.6342	0.3914	0.7918	0.0458	0.0458	0.0458	0.2500	0.2500	0.2500	1
Potatoes	0.7867	0.4452	0.9723	0.4471	0.4471	0.4471	0.5514	0.5514	0.5514	1
Carrots	0.4248	0.4248	0.4248	0.2301	0.1053	0.4171	0.4901	0.2806	0.7623	6
Grass	0.6449	0.5495	0.7762	0.2497	0.0466	0.3987	0.3651	0.3383	0.4443	4

TABLE 2. BASIC DESCRIPTION OF THE STATISTICAL DISTRIBUTION OF THE SIMULATED DATA SETS.

	Mean	Standard Deviation
Class A	60.74	4.49
Class B	74.17	14.17
Class C	90.71	10.05
Class D	120.13	4.98

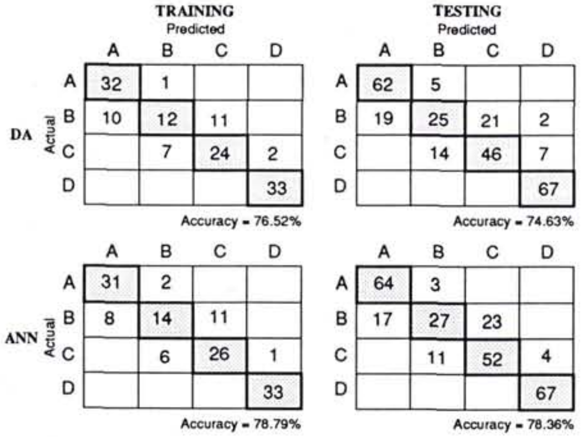
TABLE 3. MEAN AND STANDARD DEVIATION (STD DEV) OF THE TRAINING (REPRESENTATIVE SAMPLE) AND TESTING DATA USED IN THE CLASSIFICATIONS ON WHICH FIGURE 9 IS BASED.

	Training		Testing	
	Mean	Std Dev	Mean	Std Dev
Class A	60.32	4.28	60.95	4.61
Class B	73.13	14.03	74.69	14.31
Class C	89.95	9.73	91.09	10.26
Class D	119.19	4.90	120.30	5.05

parable classifications derived using normally distributed data (Figures 9 and 11). This result appears to have arisen as a function of increased overlap between classes W and Z as a result of the deliberate skewing of the distributions rather than as a result of the distributions being non-normal. As with the classifications using normally distributed data with a representative training set, the accuracy of the training and testing classifications were similar for both the artificial neural network and discriminant analysis classifications, although the accuracy of the artificial neural network classifications were again marginally higher.

The accuracy of classifications using a non-representative training set of non-normally distributed data were found to be higher from the artificial neural network than from the discriminant analysis. For both classifiers, there was also a difference in the accuracy of the training and testing classifications, although the magnitude was similar for both techniques (Figure 12).

The results of the classifications of the simulated data set therefore appear to show that the artificial neural network classifies the data as or more accurately than the discriminant analysis. It seems that the ability of the two classifiers to characterize classes, as indicated by the training classifications, is similar, except where the data have a non-normal distribution. In the latter situation, the artificial neural network classified the data to a higher accuracy. Also, there was only an appreciable difference between the accuracy of training and testing classifications when an unrepresentative training set was used. These results, especially given the difficulties in defining a truly representative training set and in assessing and correcting for non-normality in the data, indicate that artificial neural networks have considerable potential for the classification of remotely sensed data, especially as the statistical distribution of the data for each class is often unknown and the analysis is often based on a small training set.



DA — TRAINING (Predicted)

Actual	A	B	C	D
A	32	1		
B	10	12	11	
C		7	24	2
D				33

Accuracy - 76.52%

DA — TESTING (Predicted)

Actual	A	B	C	D
A	62	5		
B	19	25	21	2
C		14	46	7
D				67

Accuracy - 74.63%

ANN — TRAINING (Predicted)

Actual	A	B	C	D
A	31	2		
B	8	14	11	
C		6	26	1
D				33

Accuracy - 78.79%

ANN — TESTING (Predicted)

Actual	A	B	C	D
A	64	3		
B	17	27	23	
C		11	52	4
D				67

Accuracy - 78.36%

Figure 9. Confusion matrices for the classifications of the simulated data by the discriminant analysis (DA) and artificial neural network (ANN). The data sets were distributed normally and representative training sets were used.

The undoubted advantages of the artificial neural network based classifications over the discriminant analysis in certain circumstances must, however, be seen in the light of their limitations. Among the latter is included the largely subjective manner in which the network architecture is de-

Table 4. Mean and Standard Deviation (Std Dev) of the Training (Non-Representative Sample) and Testing Data Used in the Classifications on which Figure 10 is Based.

	Training		Testing	
	Mean	Std Dev	Mean	Std Dev
Class A	63.32	4.93	59.20	3.35
Class B	83.47	13.49	69.60	12.18
Class C	97.86	10.21	87.19	7.93
Class D	123.57	4.92	118.90	4.09

Table 5. Characteristics of the Skewed Simulated Data Set. Classes W, X, Y, and Z All Have Skewed Distributions and Are Non-Normally Distributed at the 90 Percent Level of Confidence.

Input class	Function applied	New class	New class mean
A	$A^2 \times 10^{-3}$	W	37.1
B	$B^5 \times 10^{-9}$	X	30.6
C	$C^3 \times 10^{-6}$	Y	7.7
D	$D^6 \times 10^{-12}$	Z	30.8

		TRAINING Predicted					TESTING Predicted			
		A	B	C	D		A	B	C	D
DA	A	32	1			A	67			
Actual	B	6	18	9		B	38	28	1	
	C	1	7	24	1	C	1	45	21	
	D			1	32	D			2	65
		Accuracy - 80.30%					Accuracy - 67.54%			
		A	B	C	D		A	B	C	D
ANN	A	32	1			A	67			
Actual	B	6	20	7		B	38	29		
	C	1	7	24	1	C	1	46	20	
	D			1	32	D			3	64
		Accuracy - 81.82%					Accuracy - 67.16%			

Figure 10. Confusion matrices for the classifications of the simulated data by the discriminant analysis (DA) and artificial neural network (ANN). The data sets were distributed normally and non-representative training sets were used.

		TRAINING Predicted					TESTING Predicted			
		W	X	Y	Z		W	X	Y	Z
DA	W	24	2		7	W	49	4		14
Actual	X	12	6	12	3	X	25	14	22	6
	Y			33		Y			67	
	Z	9	13	1	10	Z	19	28		20
		Accuracy - 55.30%					Accuracy - 55.97%			
		W	X	Y	Z		W	X	Y	Z
ANN	W	31			2	W	62	1		4
Actual	X	11	4	11	7	X	20	11	22	14
	Y			33		Y			66	1
	Z	19		2	12	Z			39	28
		Accuracy - 60.61%					Accuracy - 62.31%			

Figure 11. Confusion matrices for the classifications of the simulated data by the discriminant analysis (DA) and artificial neural network (ANN). The data sets were distributed non-normally and representative training sets were used.

		TRAINING Predicted					TESTING Predicted			
		W	X	Y	Z		W	X	Y	Z
DA	W	14	6		13	W	11			56
Actual	X	4	16	6	7	X	4	6	40	17
	Y			33		Y			67	
	Z	9	4	2	18	Z	4		15	48
		Accuracy - 61.36%					Accuracy - 49.25%			
		W	X	Y	Z		W	X	Y	Z
ANN	W	23	2		8	W	19			48
Actual	X	7	14	4	8	X	7	4	33	23
	Y			33		Y			67	
	Z	14	2	1	16	Z	5		2	60
		Accuracy - 65.15%					Accuracy - 55.97%			

Figure 12. Confusion matrices for the classifications of the simulated data by the discriminant analysis (DA) and artificial neural network (ANN). The data sets were distributed non-normally and non-representative training sets were used.

fined. Furthermore, the artificial neural network is computationally more demanding than the discriminant analysis in the training stage of the classification; in the analyses performed here, the training stage of the artificial neural network classifications involved some 500 iterations.

Conclusions

From classifications of SAR data and simulated data sets by an artificial neural network and discriminant analysis, four main conclusions may be drawn:

- The artificial neural network consistently provided a higher training classification accuracy than did the discriminant analysis, indicating that it is more able to accurately characterize class appearance. The differences, however, were only significant if the data were non-normally distributed.
- The better training classification performance of the artificial neural network does not always imply a better ability to classify an independent testing set. The results of a series of classifications showed that higher classification accuracies were derived from the artificial neural network but the differences were at times insignificant, particularly when *a priori* information was available to the discriminant analysis. The use of *a priori* information in an artificial neural network is currently under further investigation.
- Non-representative training data lead, as expected, to significant differences between training and testing classification accuracies, and the effect was fairly similar for both the artificial neural network and discriminant analysis classifications.

- The artificial neural network classifications were as or more accurate than those derived from the discriminant analysis; except for one classification when the accuracy was slightly and insignificantly lower. Furthermore, unlike conventional statistical classifiers, the artificial neural network is not based on often untenable assumptions about the data. Consequently artificial neural network techniques are likely to have considerable potential for the classification of remotely sensed data.

Acknowledgments

We are grateful to the Commission of the European Communities for the provision of the SAR and ground data and the anonymous referees for their comments. MBM is supported by NERC research studentship GT4/90/TLS/60 and WBY by SERC research studentship 898 00 829. An earlier version of this paper was presented at the Remote Sensing Society's annual conference in 1992 at Dundee University.

References

Almeida, L. B., and F. M. Silva, 1990. Acceleration techniques for backpropagation algorithm, *Lecture Notes in Computer Science* (L. B. Almeida and C. J. Wellekens, editors), Neural Networks EURASIP Workshop, 412:110–119.

Arai, K., 1992. A supervised Thematic Mapper classification with a purification of training samples, *International Journal of Remote Sensing*, 13:2039–2049.

Baum, E. B., 1990. When are K nearest neighbour and backpropagation accurate for feasible sized sets of examples? *Lecture Notes in Computer Science* (L. B. Almeida and C. J. Wellekens, editors), Neural Networks EURASIP Workshop, 412:2–25.

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

Campbell, J. B., 1981. Spatial correlation effects upon the accuracy of supervised classification of land cover, *Photogrammetric Engineering & Remote Sensing*, 47:355–363.

Clark, W. A. V., and P. L. Hosking, 1986. *Statistical Methods for Geographers*, Wiley, New York.

Congalton, R. G., 1991. A review of assessing the accuracy of classifications of remotely sensed data, *Remote Sensing of Environment*, 37:35–46.

Davalo, E., and P. Naïm, 1991. *Neural Networks*, Macmillan, Basingstoke.

Fahlman, S. E., 1988. Faster-learning variations of backpropagation: an empirical study, *Proceedings of the 1988 Connectionist Models Summer School*, 17-26 June, San Mateo, Morgan Kaufman, Carnegie-Mellon University, pp. 38–51.

Foody, G. M., 1988. The effects of viewing geometry on image classification. *International Journal of Remote Sensing*, 9:1909–1915.

———, G. M., 1992. Classification accuracy assessment: some alternatives to the kappa coefficient for nominal and ordinal level classifications, *Remote Sensing—From Research to Operation*, Remote Sensing Society, Nottingham, pp. 529–538.

Foody, G. M., P. J. Curran, G. B. Groom, and D. C. Munro, 1989. Multi-temporal airborne synthetic aperture radar data for crop classification, *Geocarto International*, 4:19–29.

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

Foody, G. M., M. B. McCulloch, and W. B. Yates, 1992b. An assessment of an artificial neural network for image classification, *Remote Sensing—From Research to Operation*, Remote Sensing Society, Nottingham, pp. –507.

———, 1993. Training set size and the accuracy of artificial neural network and statistical image classification, *Towards Operational Applications*, Remote Sensing Society, Nottingham, pp. 211-218.

Freund, J. E., and F. J. Williams, 1959. *Modern Business Statistics*, Pitman, London.

Hepner, G. F., 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*, 56:469–473.

Holmström, L., and P. Koistinen, 1992. Using additive noise in backpropagation training, *IEEE Transactions on Neural Networks*, 3: 24–38.

Kent, J. T., and K. V. Mardia, 1988. Spatial classification using fuzzy membership models, *IEEE Transactions on Pattern Recognition and Machine Intelligence*, 10:659–671.

Key, J. R., J. A. Maslanik, and R. G. Barry, 1989. Cloud classification from satellite data using a fuzzy sets algorithm: a polar example, *International Journal of Remote Sensing*, 10:1823–1842.

Klecka, W. R., 1980. Discriminant Analysis, Sage Publications, Beverly Hills.

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

Liu, Z. K., and J. Y. Xiao, 1991. Classification of remotely-sensed image data using artificial neural networks, *International Journal of Remote Sensing*, 12:2433–2438.

Mather, P. M., 1987a. *Computer Processing of Remotely Sensed Images*, Wiley, Chichester.

———, 1987b. Preprocessing of training data for multispectral image classification, *Advances in Digital Image Processing*, Remote Sensing Society, Nottingham, pp. 111–120.

Medina, F. I., and R. Vasquez, 1991. Use of simulated neural networks for aerial image classification, *ACSM-ASPRS Annual Convention*, ACSM and ASPRS, Bethesda, Maryland, pp. 268–274.

Pedley, M. I., and P. J. Curran, 1991. Per-field classification: an example using SPOT HRV imagery, *International Journal of Remote Sensing*, 12:2181–2192.

Quegan, S, C. C. F. Yanasse, H. de Groof, P. N. Churchill, and A. Sieber, 1991. The radiometric quality of AgriSAR data, *International Journal of Remote Sensing*, 12:277–302.

Rumelhart, D. E., G. E. Hinton, and R. J. Williams, 1986. Learning internal representation by error propagation, *Parallel Distributed Processing: Explorations in the Microstructure of Cognition* (D. E. Rumelhart and J. L. McClelland, editors), MIT Press, Cambridge, Massachussetts, pp. 318–362.

Ryan, T. W., P. J. Sementilli, P. Yuen, and B. R. Hunt, 1991. Extraction of shoreline features by neural nets and image processing, *Photogrammetric Engineering & Remote Sensing*, 57:947–955.

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

Schneider, S., 1980. Interpretation of satellite imagery for determination of land use data, *International Journal of Remote Sensing*, 1:85–89.

Short, N., 1991. A real-time expert system and neural network for the classification of remotely sensed data, *ACSM-ASPRS Annual Convention*, ACSM-ASPRS, Bethesda, Maryland, pp. 406–418.

Skidmore, A. K., and B. J. Turner, 1988. Forest mapping accuracies are improved using a supervised non parametric classifier, *Photogrammetric Engineering & Remote Sensing*, 54:1415–1421.

Stornetta, T., and R. Huberman, 1987. An improved three layer backpropagation algorithm, *IEEE First International Conference on Neural Networks*, San Diego, 3:637–644.

Story, M., and R. G. Congalton, 1986. Accuracy assessment: A user's perspective, *Photogrammetric Engineering & Remote Sensing*, 52:397–399.

Strahler, A. H., 1980. The use of prior probabilities in maximum likelihood classification of remotely sensed data, *Remote Sensing of Environment*, 10:135–163.

Swain, P. H., 1978. Fundamentals of pattern recognition, *Remote Sensing: the Quantitative Approach* (P. H. Swain and S. M. Davis, editors), McGraw-Hill, New York, pp. 136–187.

Thomas, I. L., V. M. Benning, and N. P. Ching, 1987. *Classification of Remotely Sensed Images*, Adam Hilger, Bristol.

Tom, C. H., and L. D. Miller, 1984. An automated land-use mapping comparison of the Bayesian maximum likelihood and linear discriminant analysis algorithms, *Photogrammetric Engineering & Remote Sensing*, 50:193–207.

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

Yates, W. B., J. V. Tucker, and B. C. Thompson, 1994. *Artificial Neural Networks as Synchronous Concurrent Algorithms: Algebraic Specification of a Feed Forward Backpropagation Network*, University of Wales, Technical Report (in press).

(Received 6 November 1992; accepted 17 March 1993; revised 16 June 1993)

Giles M. Foody

Giles M. Foody graduated from the University of Sheffield with a B.Sc. degree in geography in 1983 and remained at Sheffield to complete a Ph.D. in remote sensing in 1986. He is currently lecturer in topographic science and from October 1994 Senior Lecturer in geography at the University College of Swansea (University of Wales).

Mary B. McCulloch

Mary B. McCulloch graduated with a B.A. degree in geography from Saint Davids University College (University of Wales) in 1986 and then completed an M.Sc. in Environmental Management at the University of Stirling. She is currently completing a Ph.D. focused on land cover classification from imaging radar data.

William B. Yates

William B. Yates received a B.Sc. degree in computer science from the University College of Swansea (University of Wales) in 1989 and is currently completing a Ph.D. degree in computer science at Swansea on the formal specification and correctness of neurocomputers and their applications.