Relating the Land-Cover Composition of Mixed Pixels to Artificial Neural Network Classification Output

Giles M. Foody

Abstract

Artificial neural networks are attractive for use in the classification of land cover from remotely sensed data. In common with other classification approaches, artificial neural networks are used typically to derive a "hard" classification, with each case (e.g., pixel) allocated to a single class. However, this may not always be appropriate, especially if mixed pixels are abundant in the data set. This paper investigates the potential to derive information on the land-cover composition of mixed pixels from an artificial neural network classification. The approach was based on relating the activation level of artificial neural network output units, which indicate the strength of class membership, to land-cover composition. Two case studies are discussed which illustrate that the activation level of the artificial neural network outputs themselves were not strongly related to pixel composition. However, re-scaling the activation levels, to remove the bias towards very high and low strengths of class membership imposed by the unit activation function, produced measures that were strongly related to the land-cover composition of mixed pixels. In both case studies, significant correlations (all r > 0.8) between the re-scaled activation level of an output unit and the percentage cover of the class associated with the unit were obtained.

Introduction

Remotely sensed data have been used to map land cover at a range of spatial and temporal scales. The accuracy and value of the derived land-cover maps are dependent on a range of factors related to the data sets and methods used. Thus, for example, the accuracy of maps derived from conventional supervised image classification techniques is a function of factors related to the training, allocation, and testing stages of the classification (e.g., Swain, 1978; Thomas et al., 1987).

Conventional image classification techniques assume that all the pixels within the image are pure, that is, that they represent an area of homogeneous cover of a single land-cover class. This assumption is often untenable with pixels of mixed land-cover composition abundant in an image. These land-cover class mixes may arise from the gradual intergradation of continuous land-cover classes (Csaplovics, 1992; Foody et al., 1992) or, perhaps more commonly, as a consequence of the relationship between the sensor's spatial resolution and the fabric of the landscape (Irons et al., 1985; Campbell, 1987). Irrespective of their origin, mixed pixels are a major problem in land-cover mapping applications. For example, while a mixed pixel must contain at least two clas-

ses, the classification procedures generally used to produce a land-cover map are "hard" techniques which force allocation to one class. Moreover, depending on the nature of the mixture and its composite spectral response, the allocated class need not even be one of the pixel's component classes (Campbell, 1987).

The proportion of mixed pixels generally increases with a coarsening of the spatial resolution of the sensing system (Townshend and Justice, 1981; Crapper, 1984). Consequently, the effects of the mixed pixel problem may be felt most strongly when mapping land-cover from coarse spatial resolution data sets. Unfortunately, many regional to global scales studies are often constrained to the use of relatively coarse spatial resolution sensor data. At regional to global scales, however, existing land-cover data sets are known to be of poor quality, and remote sensing is the only feasible approach for land-cover mapping (Townshend et al., 1991; DeFries and Townshend, 1994). For instance, maps of tropical vegetation are required to assess the role of land-cover change, particularly deforestation, on the global carbon cycle (Wisniewski and Sampson, 1993). The available land-cover data sets, however, vary considerably and, consequently, estimates of phenomena such as deforestation vary markedly (Grainger, 1993; Curran and Foody, 1994), limiting our understanding of the carbon cycle. Although remote sensing has considerable potential for mapping tropical land cover, the only practical sensing system to use is the NOAA AVHRR which has a relatively coarse spatial resolution, 1.1 km at best. The large proportion of mixed pixels in AVHRR data can lead to significant errors in the estimation of forest extent and its change over time (Cross et al., 1991; Skole and Tucker, 1993; Curran and Foody, 1994).

Conventional "hard" image classification techniques may therefore provide a poor representation of the distribution of land cover and be a poor base for the estimation of the areal extent of land-cover classes. In some applications it is therefore desirable to unmix pixels into their component parts. A range of spectral mixture models have been developed for this task (e.g., Clark and Canas, 1993; Holben and Shimabukuro, 1993; Settle and Drake, 1993). Of these, linear mixture models are the most widely used. These, however, may not always be appropriate as the assumptions made, often including normally distributed data as well as linear mixing, are often untenable. One simple alternative approach which is widely available, and may be appropriate when the aim is to map land cover, is to "soften" the classification

Department of Geography, University of Wales Swansea, Singleton Park, Swansea, SA2 8PP, United Kingdom.

Presently with the Telford Institute of Environmental Systems, Department of Geography, University of Salford, Salford, M5 4WT, United Kingdom.

output. A softened classification output would indicate the strength of class membership a pixel had to each class, not just the code of the class with which it had the highest strength of membership. For each pixel, the strengths of class membership derived in the classification may be related to its land-cover composition (Fisher and Pathirana, 1990; Foody et al., 1992). Fuzzy classification techniques are attractive here as the concept of a pixel having a degree of membership to all classes is fundamental to fuzzy-sets-based techniques (Bosserman and Ragade, 1982; Klir and Folger, 1988). However, it is possible to soften other classifiers, including, for instance, the widely available maximum-likelihood classification (Wang, 1990; Foody et al., 1992; Maselli et al., 1994). Whatever the technique used, the basic aim is to relate the land-cover composition of pixels to measures of the strength of class membership derived from the classification. Ideally, the measures of the strength of class membership would reflect the land-cover composition of a pixel. Thus, the output for a pure pixel, representing an area of homogeneous cover of one land-cover class, should be a very high strength of membership to the actual class of membership and negligible strength of membership to other classes. Alternatively, in the output for a mixed pixel the strengths of class membership derived should reflect the relative coverage of the land-cover classes in the area represented by the pixel.

A major problem with the fuzzy-sets and probabilistic methods is that they are slow and computationally demanding. For analyzing large data sets and rapid processing, which will be vital in the EOS era when a data rate in the order of 1Tb per day is expected from the EOS sensors alone (Gershon and Miller, 1993), alternative techniques are required. One particularly attractive approach is the use of artificial neural networks. These are non-parametric techniques which have been shown to generally be capable of classifying data as or more accurately than conventional classifiers (Benediktsson et al., 1990; Foody et al., 1995). Furthermore, although there may be problems associated with training an artificial neural network, particularly in relation to overtraining and training time (Benediktsson et al., 1993; Hammerstrom, 1993), an artificial neural network, once trained, may classify data extremely rapidly as the classification process may be reduced to the solution of a large number of extremely simple calculations which may be performed in parallel (Aleksander and Morton, 1990; Schalkoff, 1992). Artificial neural networks have, however, traditionally been used to provide a hard classification (e.g., Kanellopoulos et al., 1992). The aim of this paper is to assess the potential to derive information on the land-cover composition of mixed pixels from measures of the strength of class membership that may be derived from an artificial neural network classification.

Artificial Neural Network Classification

Although artificial neural networks have a wide range of potential applications in remote sensing, their main use has been for image classification. Of the range of network types and architectures (Davalo and Naïm, 1991; Schalkoff, 1992), classification has generally been achieved with a basic layered, feedforward network architecture (Schalkoff, 1992). Such networks may be envisaged as comprising a set of simple processing units arranged in layers, with each unit in a layer connected by a weighted channel to every unit in the next layer (Figure 1). The number of units and layers in the artificial neural network are determined by factors relating, in part, to the nature of the remotely sensed data and desired classification, with an input unit for every discriminating variable and an output unit associated with each class in the classification. Typically, the output from the network is a hard classification, with only the code (i.e., nominal value)

of the predicted class of membership indicated for each pixel.

To derive a measure of the strength of class membership, and thus soften the output of the classification, it is important not to treat the artificial neural network as a black box that simply transforms input data into a class allocation, but, rather, focus on the way data are processed in the network. Each unit in the network consists of a number of input channels, an activation function and an output channel which may be connected to other units in the network. Signals impinging on a unit's inputs are multiplied by the weight of the inter-connecting channel and are summed to derive the net input (net_j) to the unit

$$net_j = \sum_{i=1}^{n} w_{ji} o_i \qquad (1)$$

where w_{ji} is the weight of the interconnection channel to unit j from unit (or input) i, o_i is the output of unit i (or external input i), and n is the number of connections to unit j (Figure 1). This net input is then transformed by the activation function to produce an output (o_j) for the unit (Rumelhart et al., 1986; Aleksander and Morton, 1990; Schalkoff, 1992). There are a range of activation functions that may be used but typically a sigmoid activation function such as

$$o_j = 1/1 + e^{-\lambda net_j} \qquad (2)$$

where λ is a gain parameter, which is often set to 1, and a bias weight are often used (Schalkoff, 1992).

The values for the weighted channels between units are not set by the analyst for the task at hand but rather determined by the network itself during training. The latter involves the network attempting to learn the correct output for the training data. Conventionally, a backpropagation learning algorithm (Rumelhart et al., 1986; Aleksander and Morton, 1990) is used which iteratively minimizes an error function over the network outputs and a set of target outputs, taken from a training data set. Training begins with the entry of the training data to the network, in which the weights connecting network units were set randomly. These data flow forward through the network to the output units. Here the network error, the difference between the desired and actual network output, is computed. This error is then fed backward through the network towards the input layer, with the weights connecting units changed in proportion to the error. The whole process is then repeated many times until the error rate is minimized or reaches an acceptable level, which may be a very time consuming process. Conventionally, the overall output error is defined as half the overall sum-of-the-squares of the output errors, which, for the pth training pattern, is

$$E_p = 0.5 \sum_{j=1}^{m} (t_{pj} - o_{pj})^2 \qquad (3)$$

where t_{pj} is the desired output, o_{pj} is the actual network output of unit j, and m is the number of units in the output layer of the network. The accumulated error after one complete pass of the training set, the total epoch error, is

$$E = \sum_{p=1}^{r} E_p \qquad (4)$$

where r is the total number of training patterns.

On each iteration, backpropagation recursively computes the gradient or change in error with respect to each weight, $\partial E/\partial w$, in the network, and these values are used to modify the weights between network units. The weights are changed by
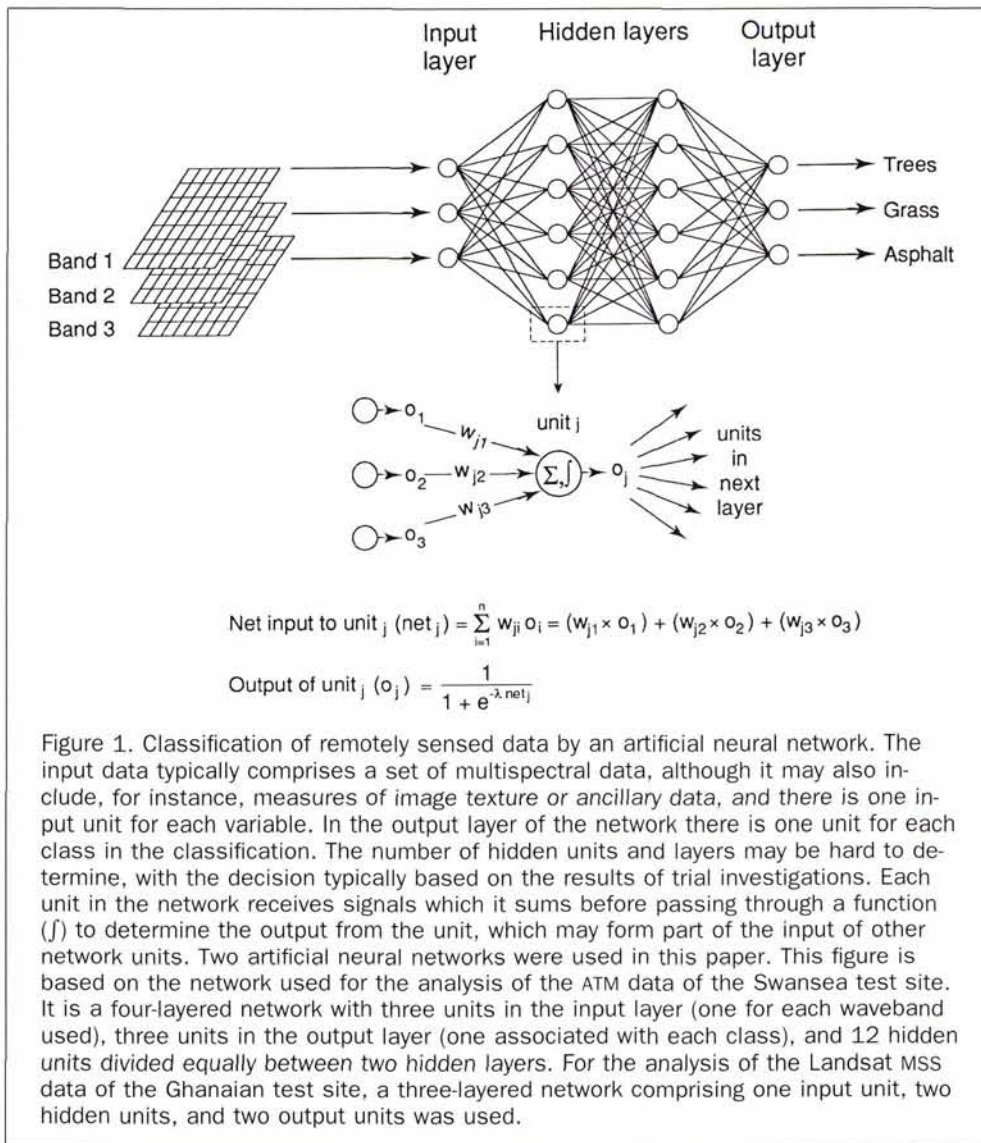
Figure 1. Classification of remotely sensed data by an artificial neural network. The input data typically comprises a set of multispectral data, although it may also include, for instance, measures of image texture or ancillary data, and there is one input unit for each variable. In the output layer of the network there is one unit for each class in the classification. The number of hidden units and layers may be hard to determine, with the decision typically based on the results of trial investigations. Each unit in the network receives signals which it sums before passing through a function ($\int$) to determine the output from the unit, which may form part of the input of other network units. Two artificial neural networks were used in this paper. This figure is based on the network used for the analysis of the ATM data of the Swansea test site. It is a four-layered network with three units in the input layer (one for each waveband used), three units in the output layer (one associated with each class), and 12 hidden units divided equally between two hidden layers. For the analysis of the Landsat MSS data of the Ghanaian test site, a three-layered network comprising one input unit, two hidden units, and two output units was used.

$$\Delta_p w_{ji} = \beta \varepsilon_{pj} o_{pi} \qquad (5)$$

where $\Delta_p w_{ji}$ is the change for the weight which connects the jth unit with its ith incoming connection, β is a constant that defines the learning rate, ε_{pj} is the computed error, and o_{pi} is the value of the ith incoming connection; often, the previous weight change is also used to add momentum to learning. For training by epoch, an overall correction to a weight is made after each presentation of all the training data and is

$$\sum_{p=1}^{r} \Delta_p w_{ji}. \qquad (6)$$

The calculation of the error, ε_{pj}, varies for output and hidden units. Because the desired output is known for the training data, the error for the output units may, assuming the use of a sigmoid activation function with $\lambda = 1$, be calculated from

$$\varepsilon_{pj} = (t_{pj} - o_{pj}) o_{pj} (1 - o_{pj}) \qquad (7)$$

whereas for a hidden unit, whose outputs are connected to k other units, the error is defined in proportion to the sum of the errors of all k units as modified by the weights connecting these units by

$$\varepsilon_{pj} = (\sum_{h=1}^{k} \varepsilon_{ph} w_{hj}) o_{pj} (1 - o_{pj}). \qquad (8)$$

Once the error has declined to an acceptable level, which is often determined subjectively, training ceases and the network is ready for the classification of cases of unknown class membership. In the classification each case is allocated to the class associated with the output unit with the highest activation level (i.e., the class associated with the output unit with the largest o_j). Further details on artificial neural network learning may be found in Rumelhart *et al.* (1986), Aleksander and Morton (1990), and Schalkoff (1992).

By outputting solely the code of the class associated with the unit in the output layer with the highest activation level, information on the magnitude of the activation level of the output units is wasted in the same way that maximum-likelihood classification is wasteful of information by discarding the probability of class membership (Wang, 1990; Foody *et al.*, 1992). The activation level of an output unit, however, indicates the strength of membership of a pixel to the class associated with the output unit. Typically, the activation level of a unit lies on a scale from 0 to 1 which reflects the variation from extremely low to extremely high strength of membership to the class associated with the out-

Figure 2. ATM image acquired in the 605 to 625-nm waveband for the test site adjacent to the University of Wales Swansea. The approximate location of the test site is highlighted by the box.

put unit. The aim of this paper is to determine if the magnitude of the output unit activation levels may be related to the land-cover composition of mixed pixels.

Data and Methods

Two test sites and data sets which provided different mixture problems were used. First, airborne thematic mapper (ATM) data acquired in 11 wavebands by a Daedalus 1268 sensor with a spatial resolution of approximately 1.5 m along a north-south flight line over the University of Wales Swansea were used. To simplify the analysis, only a subset of the wavebands recorded were used, and attention was focused on a 0.05-km² area adjacent to the University campus (Figure 2). Because ATM data have generally been found to be three-dimensional in character for a range of land covers, with the main dimensions relating to reflectance in the visible, near infrared, and middle infrared wavelengths (Townshend, 1984; Weaver, 1987), only the data acquired in three such wavebands were used. These were the data acquired in the 605 to 625-nm, 695 to 750-nm, and 1550 to 1750-nm wavebands.

The test site was comprised mainly of three land-cover classes: trees, grass, and asphalt (car park). All other classes (e.g., buildings, cars) were excluded from the analysis. For the purpose of this investigation, all the pixels in this image were assumed to be pure, that is, to comprise a homogeneous cover of one class. On the basis of this assumption, the image was classified into the three classes, and the class allocation was found to be correct for a sample of 30 pixels. This classification was used as the reference data in the evaluation of the relationship between the artificial neural network output unit activation levels and the land-cover composition of pixels in a coarser spatial resolution image. For the purpose of this paper, the coarser spatial resolution image was simulated by degrading spatially the original ATM image with an 11 by 11 low-pass (mean) filter. This spatially de-

graded image was taken to represent an image of the same region as the original image but with a coarser spatial resolution; in reality, the degraded image is only an approximation of a coarse spatial resolution image. For each pixel in the spatially degraded image, it was possible to determine the proportion of each of the three land-cover classes it contained from the classification of the spatially undegraded data. Although not ideal, the use of a classification of fine spatial resolution data to evaluate the performance of classifications of coarse spatial resolution data has been used in other studies (e.g., Iverson et al., 1989; Spanner et al., 1989). From the spatially degraded image, the DN of 50 pixels in each of the three wavebands and their corresponding land-cover composition data were extracted. At least five of these pixels were located in regions that were of homogeneous cover of each land-cover class. These 15 pure pixels were used to train the artificial neural network.

The second data set was an approximately 2200-km² extract from a cloud-free Landsat MSS image of southern Ghana, West Africa (Figure 3). Moist evergreen and moist semi-deciduous forest predominate in this region (Hall and Swaine, 1981; Whitmore, 1990) and are bordered typically by savanna, much of which is used for agro-forestry and agriculture. The boundary between the forest and savanna was abrupt and may have been sharpened by fire within the savanna (Hall and Swaine, 1981). To simplify the analysis, only the data in the 800- to 1100-nm waveband (near infrared) were used. The data in this waveband only were selected because, relative to the data acquired in the other wavebands, the effects of atmospheric attenuation were minimized and the contrast between forest and non-forest was high (Malingreau et al., 1989).

As with the analyses of the data for the Swansea test site, the original image was degraded spatially to simulate a coarse spatial resolution data set. The land-cover composition of the pixels in this simulated coarse spatial resolution
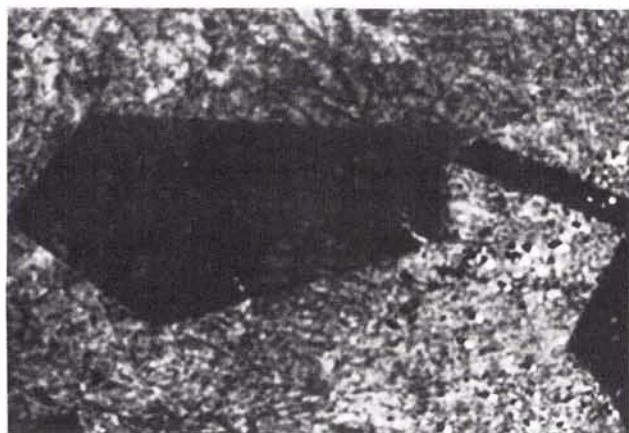
Figure 3. Extract of the Landsat MSS image acquired in the 800 to 1100-nm waveband showing tropical forest, in darker tone, in southern Ghana.

data set could then be derived from a classification of the original, undegraded spatially, image. Because the characteristics of the Landsat MSS sensor and imagery are relatively well-known, it was possible to derive a more realistic simulated coarse spatial resolution image than with the imagery of the ATM data of the Swansea test site. The Landsat MSS image had been resampled to 57-m pixel size with a cubic convolution resampling algorithm. It was then degraded to simulate imagery with a spatial resolution of 1.2 km, approximately that of NOAA AVHRR data, using a filtering approach similar to that described by Justice *et al.* (1989) which provides a set of co-registered imagery that may be considered identical in all aspects except spatial resolution. For each pixel in the simulated coarse spatial resolution image, the percentage which was forest covered was derived from a forest/non-forest classification of the original, undegraded spatially, Landsat MSS image. In the absence of accurate ground data, the forest cover estimates derived in this way were used as the reference data, although, as with the analyses of the ATM data set, they are themselves only an approximation. In total, 89 pixels were sampled from the simulated coarse spatial resolution image and their DN were extracted for the analysis. The DN of 12 pixels located at each end of the forest cover continuum were used to train the artificial neural network. The remaining 65 pixels had variable forest cover and were used to assess the relationship between the activation level of the artificial neural network output unit associated with forest and the percentage of the pixel forest covered.

The analyses were performed with a fully connected feed-forward artificial neural network. The Quickprop learning algorithm, a variant of the backpropagation learning technique outlined above, was used as it may speed-up network learning and has been shown to be superior to backpropagation in some situations (Fahlman, 1988). Relatively rapid training is achieved with Quickprop as information computed during the previous training epoch, particularly the previous error derivative ($\partial E/\partial w_{(t-1)}$), is used. With Quickprop, the weight update is achieved by

$$\Delta w_{(t)} = \frac{(\partial E/\partial w_{(t)}) \cdot \Delta w_{(t-1)}}{(\partial E/\partial w_{(t-1)}) - (\partial E/\partial w_{(t)})} \qquad (9)$$

where t indicates the iteration number (Fahlman, 1988).

The artificial neural network architecture for each analy-

sis was derived from a series of trial investigations. For the analyses of the ATM data, a four-layered network was used. This had three input units, one for each of the wavebands used, and three output units, one for each class (Figure 1). The two hidden layers each comprised six units. For the analyses of the Landsat MSS data, a simpler architecture was used, comprising just one input unit, two output units, and one hidden layer made up of two units. The parameters for the learning rate, decay, and maximum growth factor were set at 0.5, 4.0, and 1.75, respectively, which are within the range of values often used (Fahlman, 1988; NCS, 1992: Paugam-Moisy, 1993). Both networks were left to run for 700 iterations, although the error rate in each analysis had stabilized much earlier.

Results and Discussion

With the ATM data, the artificial neural network was trained using the five selected pure pixels of each class. The quality of the trained network was evaluated by determining the accuracy with which it could classify the training data; although this will give an inflated estimate of inter-class separability (Swain, 1978), it will provide a guide to the ability of the network to discriminate between the three classes. All 15 training pixels were found to be correctly classified by the artificial neural network. Furthermore, for each pixel the activation level of the artificial neural network output unit associated with the actual class of membership was very high, while those of the other output units were negligible (Table 1). For these pure pixels, the artificial neural network was therefore able to provide a very accurate and hard classification, with high activation levels associated only with the actual class of membership. For a mixed pixel, however, it may be anticipated that the magnitude of the activation level of an output unit would be related to the coverage of the associated land-cover class in the pixel.

The DN for the 35 pixels of variable land-cover composition were then input to the artificial neural network, and the activation levels of the output units were derived. Because the land-cover composition of each of these pixels was known, the relationship between the magnitude of the activation level of the output unit associated with a class and the percentage cover of that class could be derived. These relationships are shown in Figure 4, which show that the output unit activation levels were typically high or low with little variation between the extremes. Thus, for each class the activation level of the unit associated with it in the output layer was typically close to 1 if the pixel comprised 50 percent or more of that class or was otherwise close to 0. These results

TABLE 1. ACTIVATION LEVELS OF THE ARTIFICIAL NEURAL NETWORK OUTPUT UNITS FOR THE CLASSIFICATION OF THE 15 PURE PIXELS USED TO TRAIN THE ANALYSIS.

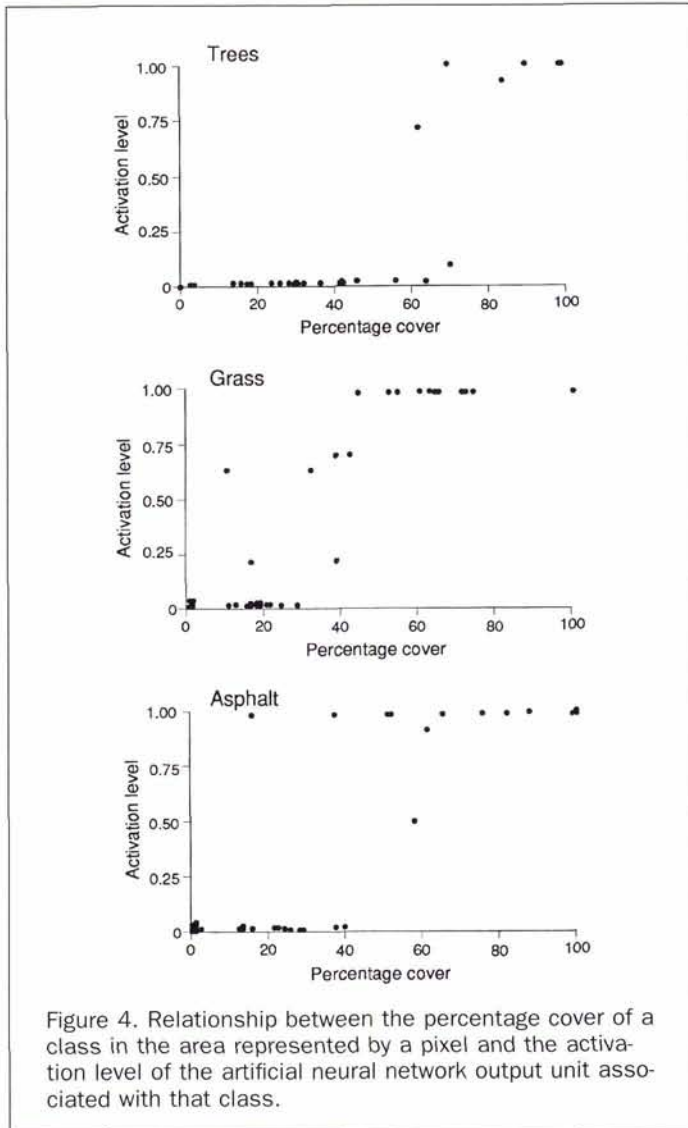| Actual Class	Activation level of output unit associated with:		
	Trees	Grass	Asphalt
Trees	0.999218	0.000552	0.000205
	0.999555	0.000192	0.000299
	0.999509	0.000241	0.000268
	0.999491	0.000244	0.000276
	0.999558	0.000186	0.000307
Grass	0.000403	0.999477	0.000264
	0.000397	0.999438	0.000293
	0.000396	0.999434	0.000296
	0.000398	0.999450	0.000286
	0.000397	0.999443	0.000291
Asphalt	0.000383	0.000307	0.999444
	0.000386	0.000293	0.999459
	0.000395	0.000265	0.999485
	0.000393	0.000287	0.999451
	0.000392	0.000293	0.999441

Figure 4. Relationship between the percentage cover of a class in the area represented by a pixel and the activation level of the artificial neural network output unit associated with that class.

indicate a fairly hard analysis and only a limited potential to derive information on the land-cover composition of mixed pixels from the activation level of the artificial neural network output units.

The activation level of an output unit is a function of the net input to it (net_j) and the unit's activation function. Here the latter was a conventional sigmoid activation function (Equation 2). This activation function aims to encourage the output from a unit towards 0 or 1 (Aleksander and Morton, 1990; Schalkoff, 1992). Although this was useful in training the network, because the training set was composed of pure pixels and the network was attempting to unambiguously learn the training data, it provides a distinctly non-linear measure of the strength of class membership. The output unit activation levels could, however, be re-scaled to remove the bias towards very low and high values imposed by the unit activation function. This was achieved here by deriving the net input (net_j) to the output units, by re-writing Equation 2, and relating this measure to the land-cover composition of the pixels. These re-scaled activation levels could be considered equivalent to the output that would have been derived if the output unit activation function had been switched, after training, to a linear function, with the only difference being that the activation function would act to re-scale the values to lie on a 0 to 1 scale.

The relationship between the re-scaled output unit activation levels and the land-cover composition of the pixels was assessed. The re-scaled activation levels were found to be more strongly related to the land-cover composition than the original network outputs (Figure 5); with Spearman rank correlations of 0.923, 0.836, and 0.808 observed for the relationships between the percentage cover of a class and re-scaled activation level of the unit associated with that class for the trees, grass, and asphalt classes, respectively (all the correlations were significant at the 99 percent level of confidence). Moreover, these results compare favorably to those derived using other approaches to unmixing the land-cover composition of pixels (Foody and Cox, 1994).

The results of the analysis of the ATM data indicated some potential for the estimation of the land-cover composition from an artificial neural network classification if attention focused on the re-scaled activation levels. This was investigated further for the simpler, but more practical, two-class situation with the Landsat MSS data. Furthermore, the approach used to spatially degrade the Landsat MSS imagery provided a better simulation of a coarse spatial resolution image than that used on the ATM data.
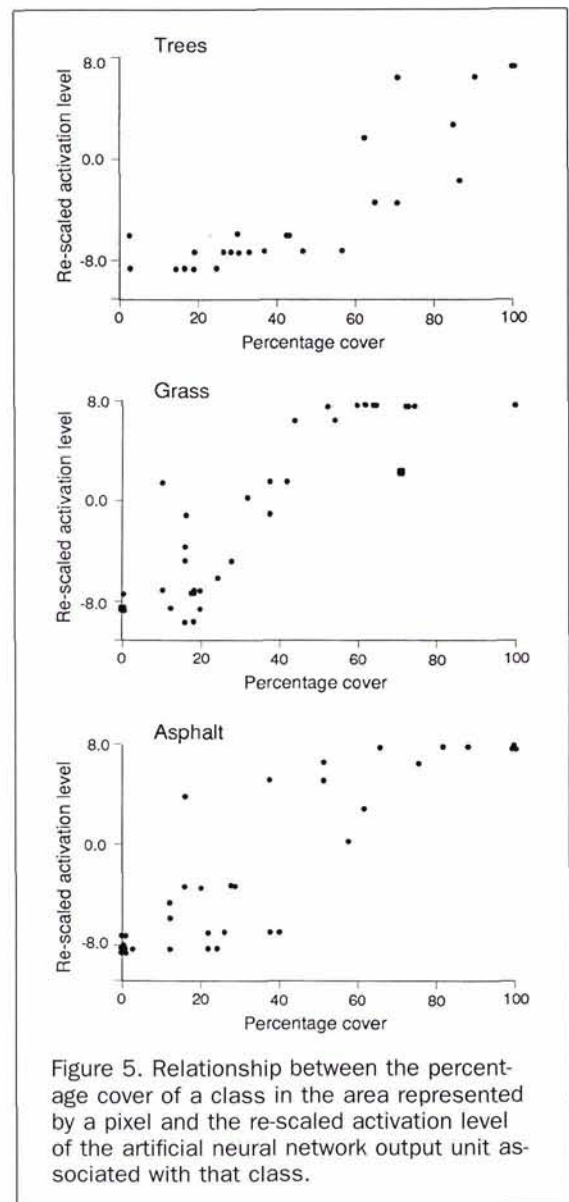

Figure 5. Relationship between the percentage cover of a class in the area represented by a pixel and the re-scaled activation level of the artificial neural network output unit associated with that class.

The artificial neural network for the analysis of the spatially degraded Landsat MSS data was trained using the 12 pure forest pixels and 12 pure non-forest pixels. Once trained, the remaining 65 pixels sampled from the data were input to the artificial neural network, and the activation level of the output units for each pixel was derived. The activation level of the output unit associated with the forest class was then related to the percentage forest cover for each pixel (Figure 6). As with the results from the analyses of the ATM data set, although more apparent in this simpler situation, the output was relatively hard, with the activation level of output units close to 1 if the pixel contained more than 50 percent forest cover, or else close to 0.

To compensate for the bias towards high and low activation levels imposed by the unit activation function, the activation levels derived from the unit associated with the forest class were re-scaled. As with the ATM data set, the re-scaled activation levels equated to the net input into the output unit associated with the forest class, derived from Equation 2. The re-scaled activation levels derived showed a strong positive relationship with the percentage of the pixel area that was forested (Figure 7). Because the relationship was linear and based on a relatively large sample, its strength was evaluated with a linear correlation. The correlation coefficient obtained, $r = 0.94$, was significant at the 99 percent level of confidence. Furthermore, given the simplicity of this mixture problem, two classes and one waveband of data, it should be noted that the relationship between the re-scaled activation level of the output unit associated with the forest class and the percentage forest was stronger than that between image tone (DN) and percentage forest cover ($r = -0.91$).

These results indicate the potential to derive accurate estimates of the land-cover composition of mixed pixels from an artificial neural network classification and indicate future avenues of research. The output unit activation levels could, for instance, be used to indicate the quality of classification allocation on a per-case basis, and so indicate uncertainty in the classification which may be beneficial to some later users of the classification. Different results may be obtained from the use of other unit activation functions, and this is currently under investigation. However, activation functions such as the sigmoid are beneficial in training the artificial neural network. It may therefore generally be most appropriate to use an activation function such as the sigmoid but to re-scale the output unit activation levels to remove the bias towards high or low values. The use of such an approach in the two case studies discussed above provided outputs which were strongly correlated with the land-cover composition of mixed pixels.

Summary and Conclusions.
Mixed pixels may be a major problem in some analyses of remotely sensed data. A land-cover classification, for example, assumes pixels to be pure and allocates each pixel to a single class. Clearly, a mixed pixel cannot be usefully represented by such a classification, resulting in a poor model of the spatial distribution of the land-cover classes and erroneous estimates of land-cover class extent.

A number of approaches have been developed to resolve the mixed pixel problem. The image classification could, for example, be softened. This would involve outputting measures of the strength of class membership each pixel has to each class instead of only the code of the most likely class of membership. These measures of the strength of class membership may then be related to the land-cover composition of the pixel. Alternatively, the analyst may elect to unmix the composition of pixels with a spectral mixture model. With both of these approaches, however, the techniques widely used make a number of often untenable assumptions about
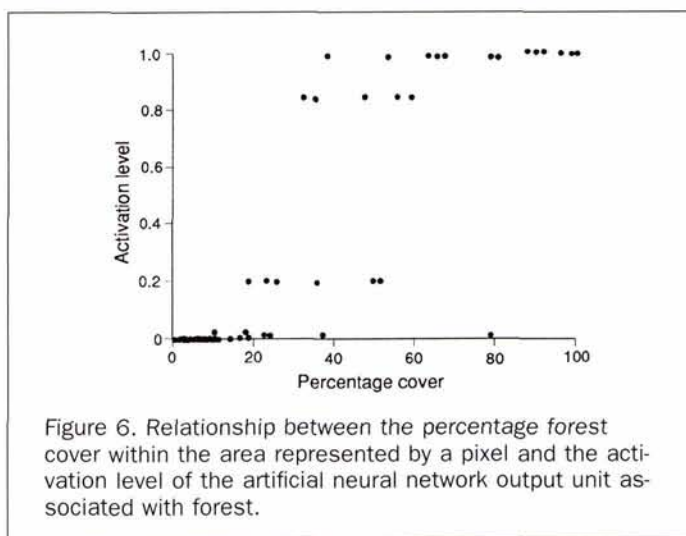

Figure 6. Relationship between the percentage forest cover within the area represented by a pixel and the activation level of the artificial neural network output unit associated with forest.

the data (e.g., some assume normally distributed data, linear mixing, etc.), may require large training sets, and may be computationally slow and demanding. Of the range of alternative approaches that could be used, artificial neural networks have been shown to have considerable potential.

Although, in classification applications, artificial neural networks have typically been used to derive a conventional hard allocation, it is, as with other classifiers, possible to soften the classification output. Thus, instead of simply allocating each pixel to the class associated with the most highly activated unit in the network's output layer, the magnitude of the activation level for all output units could be derived. This information would indicate the strength of class membership a pixel has to all the classes and may be related to the land-cover composition of the pixel. The potential of using the activation level of artificial neural network output units to unmix the land-cover composition of mixed pixels was investigated with reference to two case studies.

In both case studies, an artificial neural network was trained on a set of pure pixels as in a conventional statistical supervised classification; training on mixed pixels may also be beneficial (Foody, 1995) but this was not investigated here. For each pixel in an independent testing set, the activation level of all the network output units was derived and re-
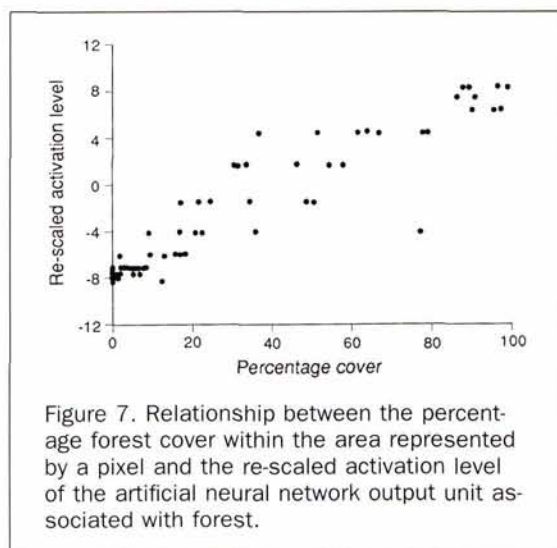

Figure 7. Relationship between the percentage forest cover within the area represented by a pixel and the re-scaled activation level of the artificial neural network output unit associated with forest.

lated to the land-cover composition of the pixels. In both investigations the original activation levels derived were only poorly related to the land-cover composition of the pixels, with a fairly hard class allocation apparent; generally, if a class covered more than half the pixel's area, the activation level of the output unit associated with that class was high and that for other output units was low. This type of output from the artificial neural network was favored through the use of the sigmoid unit activation function. However, by re-scaling the activation levels of the output units, it was possible to derive data that were strongly correlated with the land-cover composition of mixed pixels. Using the unit activation function to re-scale the activation levels, to a value equating to the net input to each output unit, a measure of the strength of class membership that was strongly correlated to the land-cover composition of the pixels was produced. In both case studies, significant correlations (all $r>0.8$) between the percentage cover of a class in a mixed pixel and the re-scaled activation level of the output unit associated with the class were obtained.

References

Aleksander, I., and H. Morton, 1990. *An Introduction to Neural Computing*, Chapman and Hall, London, 240 p.

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(4):540–551.

————, 1993. Conjugate-gradient neural networks in classification of multisource and very-high-dimensional remote sensing data, *International Journal of Remote Sensing*, 14(15):2883–2903.

Bosserman, R.W., and R.K. Ragade, 1982. Ecosystem analysis using fuzzy set theory, *Ecological Modelling*, 16:191–208.

Campbell, J.B., 1987. *Introduction to Remote Sensing*, Guilford, New York, 551 p.

Clark, C., and T. Canas, 1993. Spectral identification by artificial neural network, *Towards Operational Applications* (K. Hilton, editor), Remote Sensing Society, Nottingham, pp. 137–144.

Crapper, P.F., 1984. An estimate of the number of boundary cells in a mapped landscape coded to grid cells, *Photogrammetric Engineering & Remote Sensing*, 50(10):1497–1503.

Cross, A.M., J.J. Settle, N.A. Drake, and R.T.M. Pävinen, 1991. Sub-pixel measurement of tropical forest cover using AVHRR data, *International Journal of Remote Sensing*, 12(5):1119–1129.

Csaplovics, E., 1992. Analysis of colour infrared aerial photography and SPOT satellite data for monitoring land cover change of a heathland region of the Causse du Larzac (Massif Central, France), *International Journal of Remote Sensing*, 13(3):441–460.

Curran, P.J., and G.M. Foody, 1994. The use of remote sensing to characterise the regenerative states of tropical forests, *Environmental Remote Sensing from Regional to Global Scales* (G.M. Foody and P.J. Curran, editors), Wiley, Chichester, pp. 44–83.

Davalo, E., and P. Naïm, 1991. *Neural Networks*, Macmillan, Basingstoke, 145 p.

DeFries, R.S., and J.R.G. Townshend, 1994. Global land cover: comparison of ground based data sets to classifications with AVHRR data, *Environmental Remote Sensing from Regional to Global Scales* (G.M. Foody and P.J. Curran, editors), Wiley, Chichester, pp. 84–110.

Fahlman, S.E., 1988. Faster-learning variations of backpropagation: An empirical study, *Proceedings of the 1988 Connectionist Models Summer School* (D. Touretzky, G. Hinton, and T. Sejnowski, editors), Morgan Kaufmann, San Mateo, California, pp. 38–51.

Fisher, P.F., and S. Pathirana, 1990. The evaluation of fuzzy membership of land cover classes in the suburban zone, *Remote Sensing of Environment*, 34:121–132.

Foody, G.M., 1995. Fully fuzzy supervised image classification, *Remote Sensing in Action*, Remote Sensing Society, Nottingham, 1187–1194.

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., 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., 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.

Grainger, A., 1993. Rates of deforestation in the humid tropics: estimates and measurements, *The Geographical Journal*, 159(1):33–44.

Gershon, N.D., and C.G. Miller, 1993. Dealing with the data deluge, *IEEE Spectrum*, 30(7):28–32.

Hall, J.B., and M.D. Swaine, 1981. *Distribution and Ecology of Vascular Plants in a Tropical Rain Forest. Forest Vegetation in Ghana*, Dr Junk, The Hague, 383 p.

Hammerstrom, D., 1993. Working with neural networks, *IEEE Spectrum*, 30(7):46–53.

Holben, B.N., and Y.E. Shimabukuro, 1993. Linear mixing models applied to coarse spatial resolution data from multispectral satellite sensors, *International Journal of Remote Sensing*, 14(11): 2231–2240.

Irons, J.R., B.L. Markham, R.F. Nelson, D.L. Toll, D.L. Williams, R.S. Latty, and M.L. Stauffer, 1985. The effects of spatial resolution on the classification of Thematic Mapper data, *International Journal of Remote Sensing*, 6(8):1385–1403.

Iverson, L.R., E.A. Cook, and R.L. Graham, 1989. A technique for extrapolating and validating forest cover across large regions. Calibrating AVHRR data with TM data, *International Journal of Remote Sensing*, 10(11):1805–1812.

Justice, C.O., B.L. Markham, J.R.G. Townshend, and R.L. Kennard, 1989. Spatial degradation of satellite data, *International Journal of Remote Sensing*, 10(9):1539–1561.

Klir, G.J., and T.A. Folger, 1988. *Fuzzy Sets, Uncertainty, and Information*, Prentice-Hall, New Jersey, 355 p.

Kanellopoulos, I., A. Varfis, G.G. Wilkinson, and J. Megier, 1992. Land-cover discrimination in SPOT HRV imagery using an artificial neural network - a 20-class experiment, *International Journal of Remote Sensing*, 13(5):917–924.

Malingreau, J.P., C.J. Tucker, and N. Laporte, 1989. AVHRR for monitoring global tropical deforestation, *International Journal of Remote Sensing*, 10(4&5):855–867.

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.

NCS, 1992. *NeuralDesk User's Guide*, Neural Computer Sciences, Southampton.

Paugam-Moisy, H., 1993. Parallel neural computing based on network duplicating, *Parallel Algorithms for Digital Image Processing, Computer Vision and Neural Networks* (I. Pitas, editor), Wiley, Chichester, pp. 305–340.

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, Massachusetts, pp. 318–362.

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

Settle, J.J., and N.A. Drake, 1993. Linear mixing and the estimation of ground cover proportions, *International Journal of Remote Sensing*, 14(6):1159–1177.

Skole, D., and C. Tucker, 1993. Tropical deforestation and habitat fragmentation in the Amazon: Satellite data from 1978–1988, *Science*, 260:1905–1910.

Spanner, M.A., C.A. Hlavka, and L.L. Pierce, 1989. Analysis of forest disturbance using TM and AVHRR data, *IGARSS'89 - Quantitative Remote Sensing: An Economic Tool for the Nineties*, pp. 1387–1390.

Swain, P.H., 1978. Fundamentals of pattern recognition in remote sensing, *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, 268 p.

Townshend, J.R.G., 1984. Agricultural land-cover discrimination using thematic mapper spectral bands, *International Journal of Remote Sensing*, 5 (4):681–698.

———, 1992. Land cover, *International Journal of Remote Sensing*, 13(6&7):1319–1328.

Townshend, J.R.G., and C.O. Justice, 1981. Information extraction from remotely sensed data: a user view, *International Journal of Remote Sensing*, 2(4):313–329.

Townshend, J., C. Justice, W. Li, C. Gurney, and J. McManus, 1991. Global land cover classification by remote sensing: present capabilities and future possibilities, *Remote Sensing of Environment*, 35:243–255.

Wang, F., 1990. Improving remote sensing image analysis through fuzzy information representation, *Photogrammetric Engineering & Remote Sensing*, 56(8):1163–1169.

Weaver, R.E., 1987. Spectral separation of moorland vegetation in airborne thematic mapper data, *International Journal of Remote Sensing*, 8(1):43–55.

Whitmore, T.C., 1990. *An Introduction to Tropical Rain Forests*, Oxford University Press, Oxford, 226 p.

Wisniewski, J., and R.N. Sampson (editors), 1993. *Terrestrial Biospheric Carbon Fluxes: Quantification of sinks and sources of* CO_2, Kluwer, Dordrecht.

Forthcoming Articles

Articles listed in the May Forthcoming Articles are those scheduled through September 1996 only.

M.J. Barnsley and *S.L. Barr*, Inferring Urban Land Use from Satellite Sensor Images Using Kernel-Based Spatial Re-Classification.

Eduardo Brondizio, Emilio Moran, Paul Mausel, and *You Wu,* Land Cover in the Amazon Estuary: Linking of Thematic Mapper with Botanical and Historical Data.

B.H.C. Cheng, R.H. Bourdeau, and *B.C. Pijanowski,* A Regional Information System for Environmental Data Analysis.

O. Dikshit and *D.P. Roy,* An Empirical Investigation of Image Resampling Effects upon the Spectral and Textural Supervised Classification of a High Spatial Resolution Multispectral Image.

Sam Drake, Visual Interpretation of Vegetation Classes from Airborne Videography: An Evaluation of Observer Proficiency with Minimal Training.

John W. Dunham and *Kevin P. Price,* Comparison of Nadir and Off-Nadir Multispectral Response Patterns for Six Tallgrass Prairie Treatments in Eastern Kansas.

Leila M.G. Fonseca and *B.S. Manjunath,* Registration Techniques for Multisensor Remotely Sensed Imagery.

Bruno Garguet-Duport, Jacky Girel, Jean-Marc Chassery, and *Guy Pautou,* The Use of Multiresolution Analysis and Wavelets Transform for Merging SPOT Panchromatic and Multispectral Image Data.

Greg G. Gaston, Peggy M. Bradley, Ted S. Vinson, and *Tatayana P. Kolchugina,* Forest Ecosystem Modeling in the Russian Far East Using Vegetation and Land-Cover Regions Identified by Classification of GVI.

N.G. Kardoulas, A.C. Bird, and *A.I. Lawan,* Geometric Correction of SPOT and Landsat Imagery: A Comparison of Map and GPS Derived Control Points.

Eric F. Lambin, Change Detection at Multiple Temporal Scales: Seasonal and Annual Variations in Landscape Variables.

Sunil Narumalani, John R. Jensen, Shan Burkhalter, John D. Althausen, and *Halkard E. Mackey, Jr.,* Aquatic Macrophyte Modeling Using GIS and Logistic Multiple Regression.

Elijah W. Ramsey III and *John R. Jensen,* Remote Sensing of Mangrove Wetlands: Relating Canopy Spectra to Site-Specific Data.

Joel D. Schlagel and *Carlton M. Newton,* A GIS-Based Statistical Method to Analyze Spatial Change.

C.R. de Souza Filho, S.A. Drury, A.M. Denniss, R.W.T. Carlyon, and *D.A. Rothery,* Restoration of Corrupted Optical Fuyo-1 (JERS-1) Data Using Frequency Domain Techniques.

E. Lynn Usery, A Feature-Based Geographic Information System Model.

A.P. van Deventer, A.D. Ward, P.H. Gowda, and *J.G. Lyon,* Using Thematic Mapper Data to Identify Contrasting Soil Plains and Tillage Practices.

Chris Varekamp, Andrew K. Skidmore, and *Peter A. Burrough,* Using Public Domain Geostatistical and GIS Software for Spatial Interpolation.

Eric A. Williams and *Dennis E. Jelinski,* On Using the NOAA AVHRR "Experimental Calibrated Biweekly Global Vegetation Index."

Zhangshi Yin and *T.H. Lee Williams,* Obtaining Spatial and Temporal Vegetation Data from Landsat MSS and AVHRR/NOAA Satellite Images for a Hydrologic Model.

June 1996 Special Issue on Softcopy Photogrammetry

O. Kölbl and *Bach,* Tone Reproduction of Photographic Scanners.

Eugene E. Derenyi, The Digital Transferscope.

Weiyang Zhou, Robert H. Brock, and *Paul F. Hopkins,* A Digital System for Surface Reconstruction.

T. Ch. Massesware Rao, K. Venugopala Rao, A. Ravi Kumar, D.P. Rao, and *B.L. Deekshatula,* Digital Terrain Model (DTM) from Indian Remote Sensing (IRS) Satellite Data from the Overlap Area of Two Adjacent Paths Using Digital Photogrammetric Techniques.

Charles K. Toth and *Amnon Krupnik,* Concept, Implementation, and Results of an Automatic Aerotriangulation System.

Peggy Agouris and *Toni Schenk,* Automated Aerotriangulation Using Multiple Image Multipoint Matching.

Emmanuel P. Baltsavias, Digital Ortho-Images: A Powerful Tool for the Extraction of Spatial- and Geo-Information.

Kurt Novak and *Fayez S. Shahin,* A Comparison of Two Image Compression Techniques for Softcopy Photogrammetry.