

Feature selection and land cover classification of a MODIS-like data set for a semiarid environment

J. S. BORAK and A. H. STRAHLER

Department of Geography and Center for Remote Sensing, 725
Commonwealth Avenue, Boston, MA 02215, USA; e-mail: borak@crsa.bu.edu

(Received 7 July 1997; in final form 13 March 1998)

Abstract. The advent of the Earth Observing System (EOS), and the Moderate-resolution Imaging Spectroradiometer (MODIS) in particular, will usher in a new era of global remote sensing by providing very large data volumes for interpretation and processing. Since many data streams will contain correlated data, feature selection is an important practical problem for such activities as classification of global land cover based on spectral, temporal, spatial and directional data. Tree-based classification methods offer a suite of promising approaches to extraction of meaningful features from large measurement spaces. This research develops a tree-based model that performs feature selection on a satellite database containing information on land covers in a semiarid region in Cochise County, Arizona. In addition, we test the abilities of several classifiers to correctly label land cover using this reduced set of inputs under various sampling schemes. Results from this analysis indicate that decision trees can reduce a high-dimension dataset to a manageable set of inputs that retain most of the information of the original database, while remaining largely insensitive to choice of sampling strategy, and that Fuzzy ARTMAP, a type of artificial neural network classifier, achieves highest accuracy in comparison to maximum-likelihood or decision-tree classifiers.

1. Introduction

The Moderate-resolution Imaging Spectroradiometer (MODIS) is a primary instrument for terrestrial observation on the Earth Observing System (EOS) scheduled for launch in mid-1998 (Salomonson *et al.* 1989). MODIS will provide a combination of precise radiometry, frequent coverage and moderate spatial resolution that has previously been unavailable with any single instrument. Pre-launch MODIS data products are being developed from existing satellite data sources such as the Advanced Very High Resolution Radiometer (AVHRR) and Landsat Thematic Mapper (TM) instruments (Running *et al.* 1994a).

The MODIS land cover product is designed to utilize spectral, temporal, spatial and directional information in the assignment of land cover categories in a 1-km global land cover database. To do so, hundreds of inputs per resolution cell are available for collection and processing. Given that there are approximately 150M km² of global land surface, the computational costs of analyzing several hundred dimensions of input information on a per-pixel basis are prohibitive. In addition, many of the data streams are correlated and contain redundant information.

This research examines a decision tree-based method for reducing the input measurement space of the MODIS land cover product. To evaluate the capabilities of different algorithms to correctly classify the reduced set of inputs, the selected features were used as input to three types of classifiers: the decision tree itself, a maximum-likelihood classifier, and a neural network classifier.

The remote sensing inputs to be used as a MODIS-like dataset were derived from multi-temporal TM imagery, and covered the spectral, temporal and spatial domains expected for MODIS data (directional data were unavailable). A digital elevation layer was also present in the complete set of inputs. Reference land cover data were based on a map generated for the Gap Analysis Program (GAP) (Scott *et al.* 1993).

2. Background

2.1. Global-scale land cover classification

2.1.1. AVHRR-based approaches

Much of the recent research into land cover classification from remotely sensed data, particularly at the global scale, has focused on imagery collected by NOAA's Advanced Very High Resolution Radiometer (AVHRR) instrument (Justice *et al.* 1985, 1991, Townshend *et al.* 1987, Loveland *et al.* 1991, Running *et al.* 1994b, Ehrlich and Lambin 1996). The high temporal resolution associated with AVHRR imagery presents an attractive scenario for global land cover mapping (Townshend and Tucker 1984). Since cloud cover often obscures surface features of interest, frequent satellite overpasses are required in order to obtain clear views of desired locations (Goward 1989).

These frequent overpasses can produce unwieldy data volumes that often contain redundant information. One solution to this problem is data reduction. Probably the most common way to reduce satellite database size in a statistically valid manner is through a linear transformation of the full, correlated feature space to a set of uncorrelated factors or components (Kauth and Thomas 1976, Jackson 1983, Ingebritsen and Lyon 1985, Richards 1994). A subset of these factors retains most of the variance in the original data set in fewer dimensions. From a practical standpoint, however, the MODIS dataset is potentially so large that the processing required in a production environment for a linear transformation approach is prohibitive.

Another approach to data reduction that has been adopted for many global-scale land cover classification algorithms is to use a time sequence of maximum-value composites of the Normalized Difference Vegetation Index (NDVI) as input data. The NDVI is the normalized difference of near-infrared and red spectral reflectances, and an NDVI composite is an image in which each pixel contains the maximum NDVI observed for that location over a selected period of time (Justice *et al.* 1985). Maximum-value compositing aids in eliminating cloud cover and exploits the seasonal characteristics of vegetation using time series data, while limiting input databases to manageable sizes by reducing the spectral dimension to a single variable.

2.1.2. MODIS-based approaches

In spite of its high temporal resolution, several complications exist with AVHRR imagery as it relates to land cover classification. First, AVHRR channel 2 suffers from variable water vapour absorption, contributing to atmospheric noise. Second, calibration is problematic in the absence of onboard standards (Goward *et al.* 1991).

Third, the AVHRR instrument possesses only five broad spectral bands that are designed for cloud, rather than land observation. This limited dimensionality is insufficient to distinguish subtle differences in vegetation types that are detectable by land-sensing instruments such as the Thematic Mapper. Last, even NDVI-composited data are susceptible to problems of misregistration, persistent cloudiness and angular effects (Moody and Strahler 1994).

The advantages of the MODIS design over that of the AVHRR for land sensing are well documented (Running *et al.* 1994a). MODIS possesses temporal resolution comparable to that of the AVHRR, but with superior spectral characteristics. For example, the signal-to-noise ratios for MODIS bands are higher than those of the AVHRR, arising from the fact that the spectral bands of MODIS are less subject to atmospheric attenuation (see table 1 for a comparison of AVHRR and MODIS spectral band positions). Also, MODIS has three on-board calibration systems to detect and account for very small changes in instrument performance in a timely manner. In addition, MODIS simultaneously acquires atmospheric sounding data designed to facilitate atmospheric correction of the surface-imaging bands.

In the spatial domain, MODIS is also superior to the AVHRR instrument. While the AVHRR senses at a nominal nadir resolution of 1.1 km in all five channels, MODIS will acquire data at nominal nadir resolutions of 500 m in five of seven land bands, and at 250 m in the remaining two. Thus, MODIS data are better suited for applications that require fine-scale information.

2.2. Advanced technology classifiers

2.2.1. Artificial neural networks

Coincident with the development of enhanced tools for terrestrial observation has been the development of advanced classification algorithms that partition measurement space in innovative ways. Recently, researchers have begun to employ artificial neural network (ANN) classifiers to discriminate land cover types in remotely-sensed data (Hepner *et al.* 1990, Benediktsson *et al.* 1990, Heerman and Khazenie 1992, Bischof *et al.* 1992). The generic ANN consists of layers of processing elements, or neurons, that are linked via weighted connections. Often, an input layer is linked to an output layer by one or more intermediate layers.

The primary advantages of using artificial neural networks for classification of land cover concern processing structure, fault tolerance and statistical flexibility (Fischer and Gopal 1993). In terms of processing, ANNs are inherently parallel, and

Table 1. AVHRR, Landsat TM and MODIS reflective spectral band positions (μm).

Band	Sensor		
	AVHRR	TM	MODIS
Blue	NA	0.450–0.520	0.459–0.479
Green	NA	0.520–0.600	0.545–0.565
Red	0.580–0.680	0.630–0.690	0.620–0.670
NIR	0.720–1.10	0.760–0.900	0.841–0.876
SWIR	NA	NA	1.23–1.25
SWIR	NA	1.55–1.75	1.63–1.65
SWIR	NA	2.08–2.35	2.11–2.16

parallel processing is faster than the serial structures used in most classifiers. Secondly, neural networks tend to be more robust with respect to missing or noisy information. Finally, neural networks are not bound by statistical assumptions; they are completely nonparametric.

One ANN architecture of note is Fuzzy ARTMAP (Carpenter *et al.* 1992). This algorithm incorporates fuzzy logic for class membership and an adaptive resonance theory (ART) approach to partitioning of measurement space. In general, a Fuzzy ARTMAP network is comprised of two ART modules, ART_a and ART_b that are connected via a mapping field. Each ART module is itself composed of two layers of processing elements known as F1 (input signals) and F2 (output categories) that are linked by weighted connections.

In the training stage, ART_a organizes input patterns, while ART_b organizes output class patterns. The mapping field monitors agreement between the modules, and weights are adjusted accordingly whenever there is disagreement (Carpenter *et al.* 1992). For categorical land cover classification, the ART_b module is unnecessary as output categories are provided directly by the mapping field. Once the network has been trained, new inputs are fed to the classifier, and outputs are generated according to the structure of the network established during the learning phase.

In one of the few remote sensing-based tests of this architecture, Gopal *et al.* (1993) compared performances of Fuzzy ARTMAP and a common back-propagation neural network for classification of Sahelian land cover based on multitemporal NDVI data derived from AVHRR imagery. They found that Fuzzy ARTMAP did better in terms of both classification accuracy and computer processing considerations.

2.2.2. *Tree-based classifiers*

Another advanced classification algorithm that has appeared relatively recently in the remote sensing literature is the tree-based classifier (TBC) (Breiman *et al.* 1984). The TBC is a tool that combines certain advantages of ANN algorithms while avoiding some of their disadvantages. The approach employs tree-structured rules that recursively partition the feature spaces of data sets into increasingly homogeneous subsets based on a splitting rule. These subsets are represented as nodes in the tree structure. Nodes at the bottom of the tree, known as terminal nodes, contain the output from the classifier.

The TBC that has received the most attention thus far is the Classification and Regression Tree (CART) algorithm originally developed by Breiman *et al.* (1984). The CART approach employs binary decision rules, and is a supervised algorithm that requires training data, known as a learning sample. As with ANN approaches, class discrimination is completely nonparametric (McLachlan 1992).

CART, and TBCs in general, offer several advantages over ANN classifiers. Classification trees do not employ hidden layers. Thus, one source of ambiguity inherent in some ANN approaches is avoided completely. In addition, the number of splits in a tree can be optimized in a systematic manner (Breiman *et al.* 1984). Training data are used to grow an initial tree of maximal size and then prune it back to an optimal structure that balances parsimony with accuracy of classification. Thus, a sequence of candidate trees is tested for optimal performance. This may require considerable training time, but TBCs are often computationally less expensive than ANNs (Atlas *et al.* 1990, Brown *et al.* 1993), and the degree of ambiguity associated with optimization of structure is considerably reduced.

In addition to solving classification problems, TBCs may also be used to uncover structure in data (Breiman *et al.* 1984, Safavian and Landgrebe 1990, Sethi 1990, Guo and Gelfand 1992, Brown *et al.* 1993). The hierarchical relationships revealed by partitioning feature space in TBCs illustrate the explanatory capabilities of the predictor variables. This knowledge can be used to eliminate redundant or noisy features in input data.

For example, Brown *et al.* (1993) used decision trees to extract important features for an electronic signal identification problem, and then used this reduced feature space to train a back-propagation neural network. They classified test data both with the trained network, and with the tree that was used to reduce the measurement space. Their results indicated no significant differences between these two classifiers. In addition, they found that the neural network actually produced slightly higher classification accuracies when trained with the selected dataset than when the full set of features was included.

Until recently, tree-based methods had been limited mainly to the literature of the machine-learning and artificial intelligence fields. Currently, however, TBCs are receiving increased attention in the remote sensing literature. For example, Michaelsen *et al.* (1994) employed a TBC to optimally stratify terrain categories for integrating ground data using multitemporal remotely sensed data. The characteristics of the original cover classes that accounted for the greatest variability in the data were used to produce more meaningful stratifications for sampling purposes. Hansen *et al.* (1996) classified a global $1^\circ \times 1^\circ$ satellite dataset using both tree-based and maximum-likelihood algorithms. Results were similar for the two classifiers, but the TBC tended to outperform the maximum-likelihood approach in separating spectrally similar cover classes.

3. Methodology

3.1. Data set

The MODIS land cover classification algorithm employs a test site approach whereby a distributed network of well-characterized validation and training sites will be used to supervise the classification and to validate the output product (Strahler *et al.* 1994). For the current research, a test site database has been compiled for the desert ecosystem of Cochise County, Arizona. The vegetation distribution throughout the study region is strongly influenced by the underlying topography such that land cover types vary from evergreen needleleaf forest at the highest elevations to barren playa on the desert floor. The most common cover type at the site is grassland, but a significant riparian zone exists along the San Pedro River Basin which contains cottonwood and willow tree species. Total areal coverage of the site is approximately 13 000 km².

Prior to the launch of MODIS, the main source of remote sensing data at each test site is Landsat Thematic Mapper imagery owing to its similarity in spectral coverage to that of MODIS data (see table 1). In addition, since the MODIS land cover classification algorithm is designed to ingest surface reflectances from 12-monthly composite input databases (Strahler *et al.* 1994), an intra-annual sequence of TM scenes that simulates this coverage is ideal for test scenarios. Each site database also includes elevation information and reference land cover data.

For the Cochise County test site, cloud cover and availability restrictions reduced the number of viable TM scenes to seven collected over a span of seven months (see table 2), but the timing still allowed good coverage of the phenologies of the various

Table 2. Landsat TM scene acquisitions.

Scene no.	Date of acquisition
1	23 April 1992
2	10 June 1992
3	26 June 1992
4	12 July 1992
5	30 September 1992
6	1 November 1992
7	11 November 1992

vegetation covers. Therefore, the set of satellite data employed in this research consisted of the six reflective channels from seven TM scenes collected in 1992 over the study site (TM path/row = 35/38). Corresponding elevation data were subsetted from a 30 arc-second digital elevation model (DEM) derived from the Digital Chart of the World by Earth Resources Observation Systems (EROS) Data Center.

Reference land cover information was derived from a vegetation map which was produced as part of GAP's state-by-state effort to identify gaps in natural communities (Scott *et al.* 1993). A rasterized map of the area based on the associated vegetation layer for Arizona at 30 m spatial resolution (Graham 1995) was acquired in order to both train and test the classification algorithms.

The satellite imagery, DEM and land cover data layers were first co-registered. Then, the remote sensing data were atmospherically corrected to surface reflectance using the MODTRAN atmospheric characterization algorithm (Berk *et al.* 1989). Surface and upper-air information used in the correction process was compiled from observations collected at several local meteorological stations.

Finally, the categories in the land cover layer were mapped from the ones used for GAP to those used in the International Geosphere Biosphere Programme's Data and Information System (IGBP-DIS) 1-km Land-Cover Project (Belward and Loveland 1995), since these are the categories recognized by the MODIS land cover product. The IGBP-DIS 1-km Land-Cover Project has mapped global land cover from AVHRR data in order to provide an accurate land cover database for global change studies of the IGBP core projects.

In order to relabel the original GAP classes as IGBP-DIS categories, several botanical and ecological references were consulted (Kearney and Peebles 1951, Benson and Darrow 1981, Dick-Peddie 1993). The relabelling produced a relatively small number of rather large land cover polygons (approximately 200 for the entire site), illustrating a general spatial smoothing characteristic of many datasets derived from vector coverages. A list of cover types and their relative areal proportions is compiled in table 3.

Next, the remote sensing inputs were spatially degraded to simulate the characteristics of MODIS data by employing a fast Fourier transform approach developed by the MODIS Calibration Science Team. The associated computer program convolves a TM-to-MODIS Gaussian blur filter with Landsat TM digital counts to degrade the fine-scale data to 240, 480 or 960 m MODIS-like resolution. Since the MODIS land cover classification algorithm will employ a 1000 m nadir-equivalent surface reflectance database (Borak 1996), all six reflective TM bands were degraded to 960 m spatial resolution for each scene. In addition, bands 3 and 4 were degraded

Table 3. Site cover types and areal proportions (IGBP-DIS classification legend).

Label	Class description	Areal proportion
ENL	Evergreen needleleaf forest	0.016
EBL	Evergreen broadleaf forest	0.024
DBL	Deciduous broadleaf forest	0.001
MXF	Mixed forest	0.086
WSV	Woody savanna	0.016
GRS	Grassland	0.645
CSL	Closed shrubland	0.012
OSL	Open shrubland	0.108
CRP	Cropland	0.072
UBU	Urban and built-up	0.007
BAR	Bare soil and rocks	0.013

to 240 m resolution in order to derive a spatial texture metric and a soil adjusted vegetation index (SAVI) from the satellite data. The image texture measure was defined as the ratio of the standard deviation to the mean of all 240 m band 4 pixel reflectances within a given 960 m ground element. The SAVI was derived following the technique of Kaufman and Tanré (1992) using the 240 m data generated from bands 3 and 4.

3.2. Classification

3.2.1. Sampling of training data

Two stratified sampling strategies were used in this research. The first was a random approach which drew an equal proportion (0.20) of the total pixels from each cover type. The other was a random scheme which drew an equal number of samples (300) from each cover class. These two strategies were used for both feature selection and classification.

3.2.2. Feature selection and classification

Feature selection was performed following the tree-based approach of Breiman *et al.* (1984), as implemented in the *Splus* statistical software package. Initially, decision trees of maximum size were generated using the full suite of input features. Next, the trees were pruned in order to reduce the fitting of noise in the inputs. Lastly, those features that contributed most to the variance in the training data were retained for the classification phase.

Three approaches were examined to test classification performance on the reduced set of inputs. First, the predictive capability of the decision tree itself was explored. Although primarily used for feature selection in this analysis, the ability of the tree to label previously unseen data was of interest because this indicated whether acceptable classification results might be produced without additional processing. A second approach followed the methodology of Brown *et al.* (1993) by using the features selected by the decision tree to build a neural network classifier. The ANN architecture of choice was Fuzzy ARTMAP, as implemented in the *NeuralWorks Professional II/PLUS* software package. The third algorithm was a maximum-likelihood classifier (MLC). The intention was to use the performance of the MLC as a benchmark for evaluating the results generated by the advanced techniques.

4. Results and discussion

4.1. Accuracy and error computation

In the following discussion, we will refer to percentage errors of omission and commission. These errors are calculated assuming that the GAP reference data are true. The percentage error of omission is the percentage of GAP pixels in a particular class that are incorrectly classified by the algorithm. The percentage error of commission is the percentage of incorrectly labelled pixels for a given category in the classified data. Overall accuracy refers to the percentage of all GAP pixels of all classes that are correctly classified. Accuracy and error statistics include pixels used for training the classifiers.

4.2. Proportional sampling

4.2.1. Tree-based classifier

The pruned decision tree for the proportional sampling, depicted in figure 1, selected 14 features from the original set of 57 (see table 4 for a complete listing of selected features), which corresponded to a 75.4% reduction in dimensionality. Each spectral band was selected at least once in the time series, the SAVI was selected three times (scenes 2, 6 and 7), and image texture was selected in scene 2. These results indicated that useful information was present in all of the remote sensing data streams, including the derived data layers. On any given date, however, specific features were more important than others.

Overall, the three most important features were SAVI from scene 6, band 7 surface reflectance from scene 7, and elevation. In general, the SAVI split discriminated grassland and other sparse cover types from the denser cover types. The band 7 reflectance criterion separated cropland from forest and shrub classes, possibly based on differences in water content, and the elevation information appeared to account for differences in temperature and/or moisture regimes among the sparse covers.

In terms of classification accuracy, the 14-feature tree correctly labelled the classes in the reference map at a rate of 73.5%. This level of accuracy mainly reflected the low omission error rate of grassland which accounted for 65% of the observations (see table 5). Aside from that, the only other categories that had omission errors below 50% were evergreen needleleaf forest, mixed forest and bare soil and rocks. Since grassland was the most common class in the training data, the tree tended to split on factors that selected that class, sometimes at the expense of the smaller categories. This is illustrated by the moderate commission error rate for grassland in table 5, and also by the fact that half of the terminal nodes in figure 1 are grassland categories.

4.2.2. Maximum-likelihood classifier

The MLC produced a classification of lower overall accuracy (56.3%) relative to the TBC because it had problems with the grassland class (see table 5). The high errors of omission and low errors of commission implied a narrow category definition in measurement space. Thus, the classifier tended to exclude many examples of the grassland class. In some of the other classes, however, maximum-likelihood outperformed the decision tree. For example, the omission and commission error rates for cropland represented improvements in accuracy over the tree classifier.

When the prior probabilities for the classes were adjusted to the known proportions of the cover types in the study area, the performance of the MLC improved to

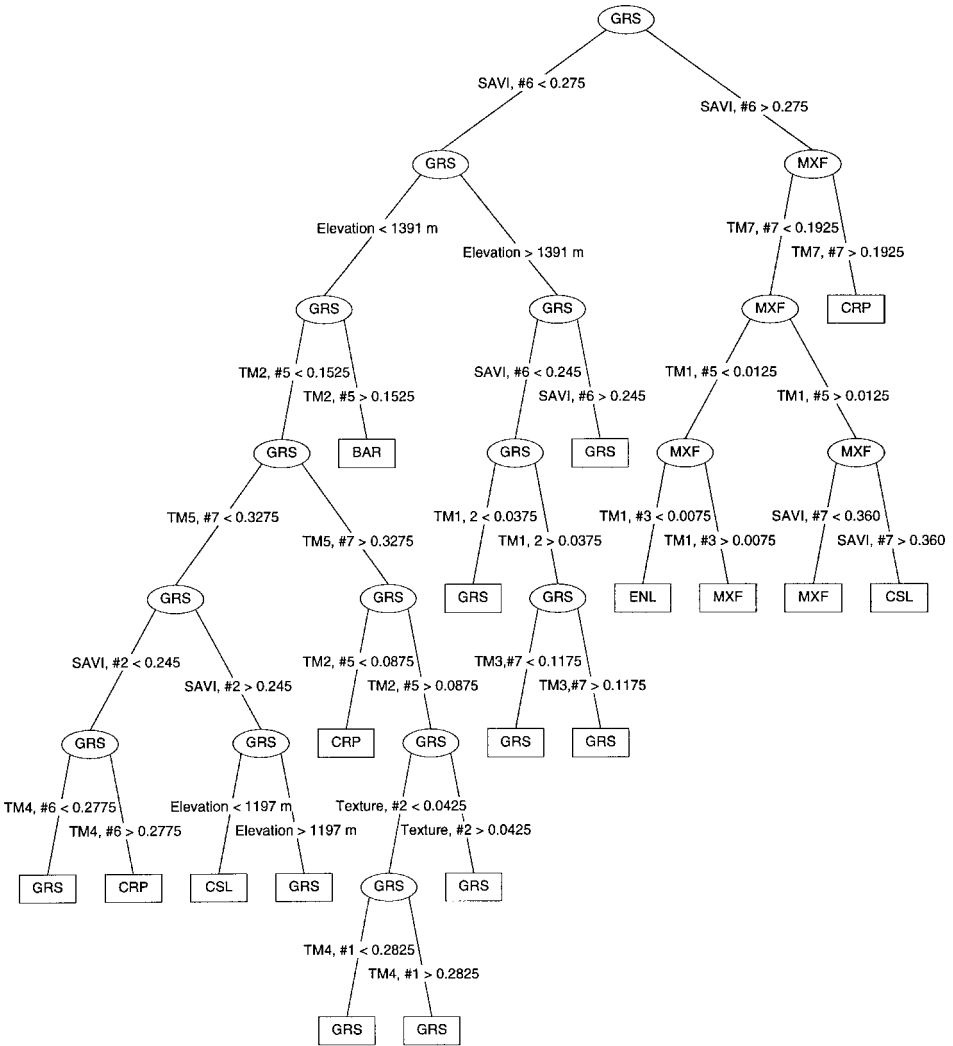


Figure 1. Proportional sampling decision tree.

Table 4. Features selected by proportional (P) and equal (E) sampling approaches; elevation (not shown) was selected by both approaches.

	TM1	TM2	TM3	TM4	TM5	TM7	Texture	SAVI
Scene 1				P				E
Scene 2	P						P, E	P, E
Scene 3	P, E							
Scene 4								
Scene 5	P	P	E					
Scene 6				P, E				P, E
Scene 7		E	P		P, E	P		P

Table 5. Omission and commission errors with overall accuracies (%), proportional sampling.

Cover category	Classifier											
	Decision tree		Maximum-likelihood with equal priors		Maximum-likelihood with area-weighted priors		Fuzzy ARTMAP reduced features		Fuzzy ARTMAP all features		Commission	
	Omission	Commission	Omission	Commission	Omission	Commission	Omission	Commission	Omission	Commission	Omission	Commission
Evergreen needleleaf forest	39.2	22.1	34.9	39.0	42.1	28.4	38.8	36.9	36.4	22.2	36.4	22.2
Evergreen broadleaf forest	100.0	0.0	28.8	80.6	39.5	71.6	46.7	55.9	65.4	53.7	65.4	53.7
Deciduous broadleaf forest	100.0	0.0	100.0	0.0	100.0	0.0	72.2	0.0	77.8	0.0	77.8	0.0
Mixed forest	17.7	38.3	50.4	26.7	38.7	24.7	28.3	24.3	25.1	27.8	25.1	27.8
Woody savanna	100.0	0.0	41.8	72.9	60.7	50.9	42.8	51.7	59.7	42.6	59.7	42.6
Grassland	6.2	23.1	51.8	9.3	24.2	14.6	12.9	14.4	13.6	16.3	13.6	16.3
Closed shrubland	54.4	59.6	38.6	77.6	50.0	64.1	48.7	35.7	59.5	44.8	59.5	44.8
Open shrubland	100.0	0.0	14.6	70.7	34.4	60.8	38.7	42.0	45.2	47.7	45.2	47.7
Cropland	56.7	42.7	19.9	36.9	33.2	24.9	39.8	26.4	34.3	27.4	34.3	27.4
Urban and built-up	100.0	0.0	59.8	46.4	67.4	21.1	60.9	2.7	66.3	13.9	66.3	13.9
Bare soil and rocks	4.1	29.0	1.2	24.2	2.3	22.0	5.8	5.8	7.6	7.6	7.6	7.6
Overall accuracies		73.5		56.3		71.1		78.6		77.3		77.3

71.1% overall accuracy, a figure nearly as high as that of the decision tree. This difference was due to the MLC's weighting of the various class probability distribution functions (PDFs) to reflect these relative proportions. Classification results for the three largest classes, grassland, cropland and mixed forest illustrated the overall improvement.

4.2.3. *Fuzzy ARTMAP*

The Fuzzy ARTMAP classification, with 78.6% overall accuracy, was the most accurate of the classifiers trained with proportional sampling. This was due in part to its high accuracy in the large grassland class (see table 5). Low errors of commission and omission indicated that this category was defined broadly in measurement space, but not so broadly as to become overly inclusive.

Among the smaller classes, Fuzzy ARTMAP produced results of varying accuracies. For example, the mixed forest class was well characterized, as was evident in the omission and commission error rates. In contrast, the classifier did not perform as well for the evergreen broadleaf forest class. The associated errors of omission and commission were not as problematic as they might appear, since 46.4% of the commission errors were misclassified mixed forest pixels. Considering that the pixels were nearly 1 km in spatial resolution, this did not seem to be an entirely negative result.

For comparison purposes, table 5 also shows the results of running Fuzzy ARTMAP using the full-dimensional training set. The overall accuracy of this classification (77.3%) was slightly lower than that obtained using the reduced measurement space. In terms of per-class performance, 6 of the 11 classes exhibited equal or greater rates of both omission and commission errors. Only one category, evergreen needleleaf forest, exhibited lower rates of both error types. The other four classes appeared to exchange omission for commission errors, or vice versa, with no obvious trends.

4.3. *Equal sampling*

4.3.1. *Tree-based classifier*

For the equal sampling approach, certain classes were too small for the analysis, and were eliminated from the dataset. The classes that remained were evergreen broadleaf forest, mixed forest, grassland, open shrubland and cropland. The pruned decision tree for this scenario, depicted in figure 2, selected 10 features from the original set of 57 (see table 4 for a complete listing of selected features), which corresponded to an 82.5% reduction in dimensionality. Five of the six reflectance channels (1–5) were selected at least once in the time series, the SAVI was selected three times (in scenes 1, 2 and 6), and image texture was selected in scene 2.

The selected features were more evenly distributed temporally, in comparison to those produced with proportional sampling. This was probably due to a lack (in the training data) of a dominant class with a strong seasonal signal. The selection of elevation (a static input) as the most important feature further indicated a suppressed temporal signal.

Overall, the three most important features were elevation, band 5 surface reflectance from scene 7, and band 1 surface reflectance from scene 3. For the most part, the elevation split appeared to separate forest types from non-forest types, most likely based on temperature and/or moisture distributions correlated with elevation. The channel 5 reflectance criterion probably discriminated lowland vegetation stands

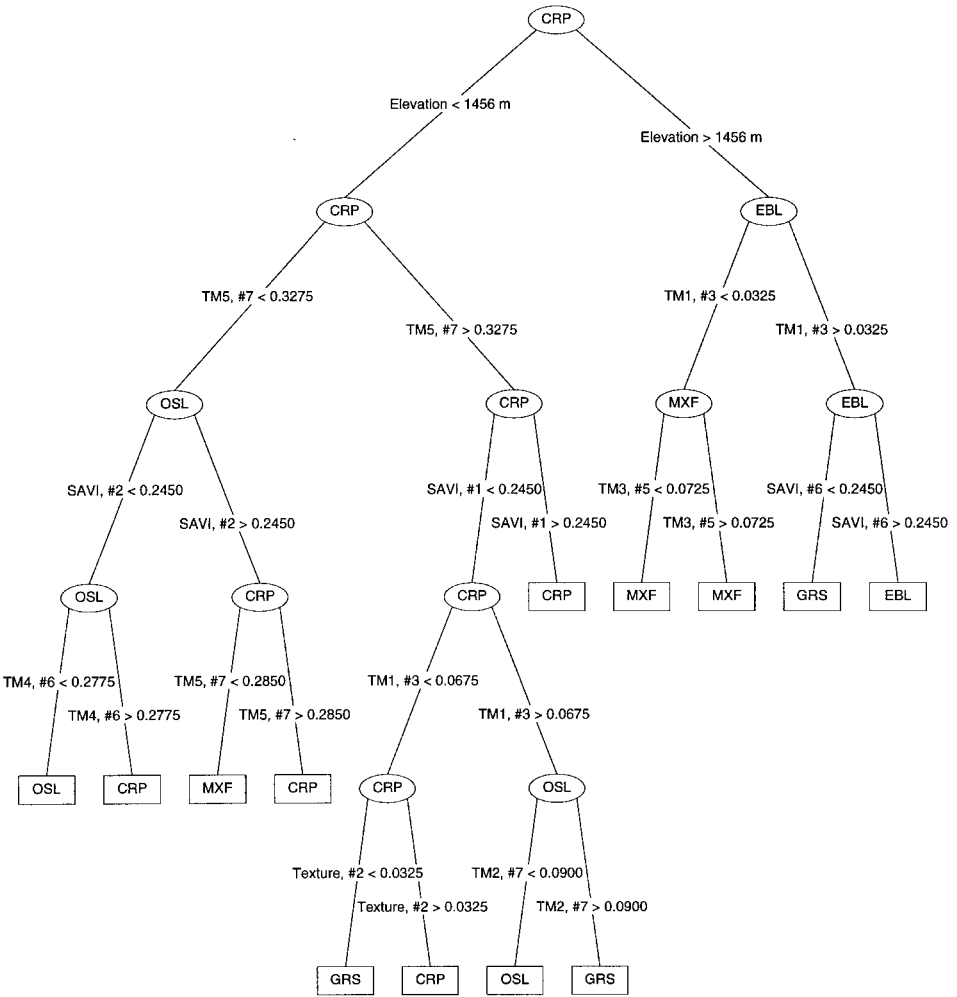


Figure 2. Equal sampling decision tree.

based on moisture content, and the band 1 reflectance data seemed to separate upland cover types based on sparseness of cover.

Interpreting the meaning of this tree's structure was somewhat more difficult than for the proportional sampling case. A striking difference between the trees was the location of the split for the SAVI layer from scene 6. For the equal sampling approach, this layer appeared to be useful only for discriminating between grassland and evergreen broadleaf forest at higher elevations. For the proportional sampling scheme, however, this layer was the most important feature, as it was useful for separating grassland and other sparse covers from denser ones, irrespective of elevation (see Section 4.2.1).

The tree's performance was degraded in terms of overall accuracy (47.2%) because it characterized grassland poorly (see table 6). Omission error rate was quite high, but commission error rate was low. This implied a narrow definition of this category

Table 6. Omission and commission errors with overall accuracies (%), equal sampling.

Cover category	Classifier														
	Decision tree			Maximum-likelihood with equal priors			Maximum-likelihood with area-weighted priors			Fuzzy ARTMAP reduced features			Fuzzy ARTMAP all features		
	Omission	Commission	Overall	Omission	Commission	Overall	Omission	Commission	Overall	Omission	Commission	Overall	Omission	Commission	Overall
Evergreen broadleaf forest	16.7	81.2	12.7	82.4	35.6	67.3	0.7	74.3	1.0	69.0	63.6	11.6	54.7	64.1	
Mixed forest	35.1	36.1	42.5	18.7	28.8	19.0	22.8	22.5	21.6	26.5	63.6	43.7	7.2	65.5	
Grassland	66.6	7.4	54.8	7.2	18.3	14.4	44.0	7.0	43.7	7.2	63.6	24.7	65.5	54.7	
Open shrubland	19.3	73.7	14.9	73.0	50.3	57.7	23.9	65.3	24.7	65.5	63.6	11.6	54.7	64.1	
Cropland	13.0	64.9	23.1	45.7	36.2	34.8	15.3	54.0	11.6	54.7	63.6	11.6	54.7	64.1	
Overall accuracies	47.2	54.4	75.2	54.4	63.6	75.2	15.3	63.6	11.6	54.7	63.6	11.6	54.7	64.1	

in feature space, much like the MLC example with equal prior probabilities in Section 4.2.2.

4.3.2. *Maximum-likelihood classifier*

This approach was theoretically equivalent to an MLC with equal prior probabilities and proportional sampling, as in Section 4.2.2, since the underlying mean variance–covariance matrix for each class is the same regardless of sample size. Thus, overall accuracies for the two cases were quite close, although low overall. At 54.4%, the outcome here was actually somewhat better than that of the decision tree. The maximum-likelihood omission error rates were generally lower, and commission rates higher than those generated for the tree. For example, omission error rate for grassland was 54.8%, but commission error rate was only 7.2% (see table 6). Thus, the classifier tended to exclude many examples of the grassland class, and overall accuracy was degraded.

Once again, when the prior probabilities were adjusted to the cover type proportions in the study area, the performance of the MLC improved. Overall accuracy was 75.2%, which was the most accurate for the classifiers trained with equal sampling. This superiority was due to the expansion of the PDF associated with the grassland class, as illustrated by the low omission error rate (see table 6). Conversely, the commission error rate was approximately twice that of the other classifiers.

4.3.3. *Fuzzy ARTMAP*

The Fuzzy ARTMAP classification, with 63.6% overall accuracy, was the next most accurate of the classifiers trained with equal sampling. Omission error rate for the grassland class was moderate for the neural network when compared with the other classifiers (see table 6). On small classes, Fuzzy ARTMAP produced results of medium to high accuracy. For example, the evergreen broadleaf forest class was overly inclusive, as illustrated by the low omission error rate, and high commission error rate. The mixed forest category was well characterized, as was evident in the omission and commission error rates.

For comparison, table 6 also shows the results of running Fuzzy ARTMAP using the full-dimensional training set. The overall accuracy of this scenario (64.1%) was only slightly higher than that obtained using the reduced measurement space. Omission and commission error rates were nearly identical for the two runs in all cover categories.

4.4. *Use of κ to compare classifications*

A quantitative comparison of classifier performances was accomplished using the κ test statistic (Congalton *et al.* 1983, Hudson and Ramm 1987). Table 7 presents κ values for all of the classification runs. All of the classifiers performed as well as or

Table 7. κ statistics.

Classifier	Proportional sampling	Equal sampling
Decision tree	0.446	0.036
Maximum-likelihood no priors	0.401	0.357
Maximum-likelihood with priors	0.518	0.519
Fuzzy ARTMAP reduced features	0.613	0.450
Fuzzy ARTMAP all features	0.583	0.455

better with proportionally sampled training data when compared with the equal sampling cases. Although the $\hat{\kappa}$ figure generated for the MLC with adjusted prior probabilities was slightly higher for the equal sampling run, this difference was not significant ($p < 0.5$).

The classifications generated by the MLC with unadjusted prior probabilities and the TBC were less accurate than those produced by the other approaches. When prior probabilities were taken into account, the MLC outperformed the TBC regardless of sampling approach, and actually was the most accurate of any classifier with equal sampling. This superiority illustrates the importance of prior probabilities and prior knowledge of cover type proportions in maximum-likelihood classification in general.

Fuzzy ARTMAP produced the most accurate classification overall. In the proportional sampling case with reduced features, $\hat{\kappa}$ was 0.613. With the full set of features, $\hat{\kappa}$ decreased to 0.583. These figures were statistically different ($p < 0.0007$). Also, the best Fuzzy ARTMAP run was significantly different from the best MLC run. Thus, the neural network classifier performed significantly better than any other classifier, and better still with the reduced feature space than with the full feature space.

In the equal sampling case with reduced features, $\hat{\kappa}$ for Fuzzy ARTMAP was 0.450. For the full feature space, $\hat{\kappa}$ was 0.455. These figures were not found to be statistically different ($p < 0.288$). Thus, the use of a reduced feature space did not statistically degrade neural network classifier performance under either of the sampling schemes.

4.5. Spatial patterns and classifier performance

The classification results of the proportional sampling runs are presented graphically in figure 3. Figure 3(a) is an example of a spatially degraded TM image of the study area for comparison purposes, while figure 3(b) is the land cover reference dataset. Parts (c)–(e) display the results of the three types of classifiers for the proportional sampling case.

For the TBC classification, depicted in figure 3(c), it was clear that the classifier overestimated the grassland class. This overestimation came mainly at the expense of open shrubland. Broadleaf forest and woody savanna were also underestimated and were mostly mislabelled as mixed forest pixels. The MLC classification with prior probabilities, shown in figure 3(d), mapped the smaller cover classes such as broadleaf forest and woody savanna much better than the TBC. Distribution of the cropland class was also improved. The main drawback seemed to be an overestimation of open shrubland, which was generally identified as grassland in the reference data.

The Fuzzy ARTMAP classification with reduced features is shown in figure 3(e). Clearly, this map was best at recreating the patterns in the reference data. The spatial patterns of the grassland, open shrubland, woody savanna and cropland classes were all reproduced reasonably well. Difficulties still existed with the urban and built-up class, but this was more of a land use than a land cover category, and may have confused the classifier for that reason.

4.6. Classification system mismatches and degraded accuracies

The reader may well note that the range of overall classification accuracies reported here falls below the 85–90% range which is considered an achievable target for remote classification of land cover (Belward 1996). The authors believe that the

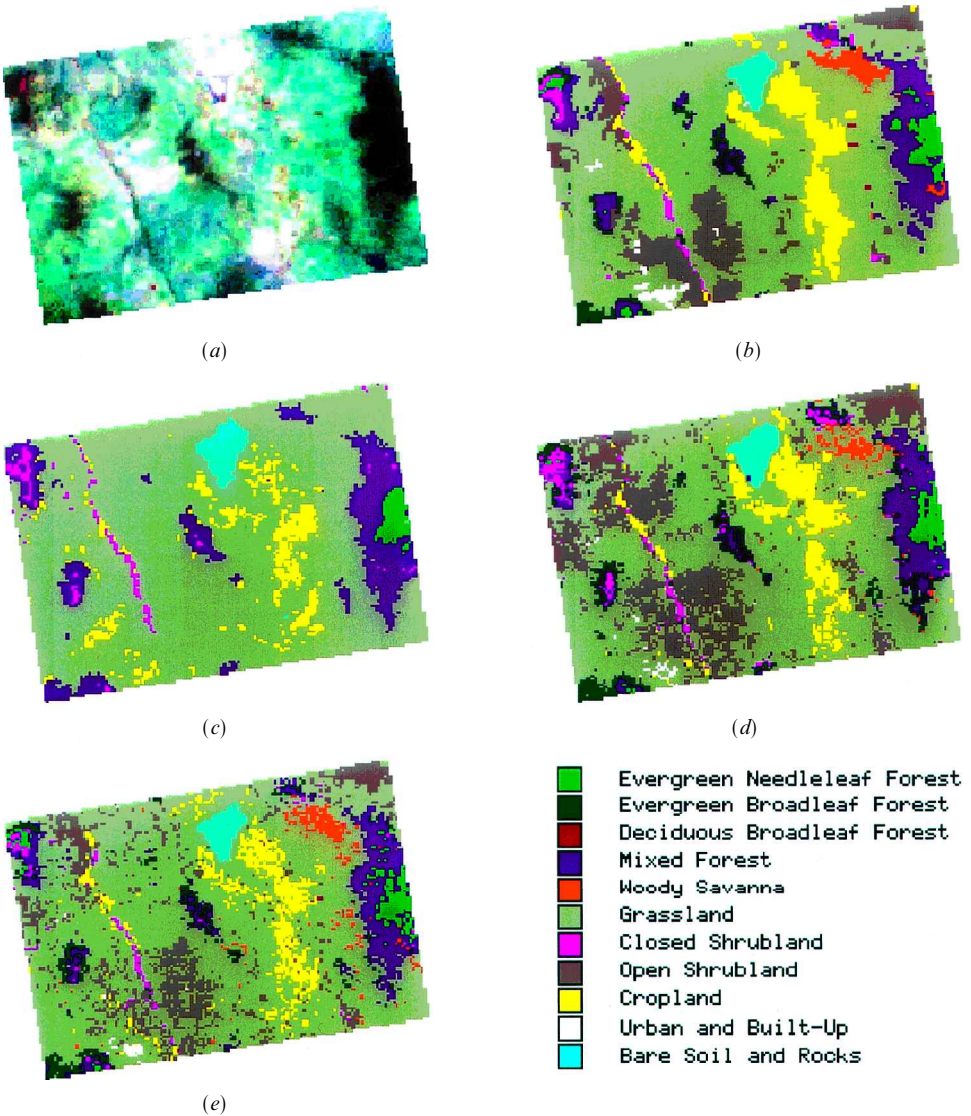


Figure 3. Comparison between 4, 5, 3 color composite of spatially degraded Landsat TM data from scene 5 (a), reference land cover data (b), TBC with proportional sampling (c), MLC with proportional sampling and prior probabilities (d), and Fuzzy ARTMAP with proportional sampling, reduced features (e).

low accuracies are partly attributable to the nature of the GAP database, which does not map directly to the IGBP-DIS classification system. The best evidence of this mismatch is the fact that approximately 45% of the TM-resolution GAP database is labelled 'mixed grass/mixed scrub', which becomes grassland when cross-walked to the IGBP-DIS legend. This relabelling is somewhat arbitrary in that no proportional information on cover type mixtures is available. Thus, the grassland category exhibits a high degree of heterogeneity and overlaps with other types in measurement space. Since these inconsistencies produce mislabelled pixels in the

reference database, they reduce classification accuracy both by training the classifiers on spurious data and by increasing omission and commission error rates.

A second factor reducing classification accuracies is the spatial smoothing of the reference data. As shown in figure 3, the spatial pattern of all three classified images is more detailed than that of the reference data, which are generalized in the cartographic process. As a result, mismatches are unavoidably generated within smoothed reference polygons. In this case, the classifiers are probably doing a better job of showing the true land cover pattern at the 1 km scale.

Although these problems degrade classifier performance by adding noise, they do not invalidate our analysis nor the conclusions we draw regarding the reduction of feature space and the intercomparison of classifiers. In fact, the noise provides a more challenging test of each algorithm in partitioning an overly complex measurement space.

5. Conclusions

Feature selection by the decision trees was effective in reducing the dimensionality of the input measurement space. Although the trees reduced the number of input features by approximately 75–80%, classification accuracies were not significantly degraded by this reduction. The neural network classifiers performed as well, or better, with the feature subsets than with the full complement of features.

Based on the features selected by the decision trees, all spectral bands and the SAVI are useful at the test site. This indicates that the entire spectral range contained relevant information. The SAVI was particularly useful in that it was selected more often than any other data source in either sampling scheme. The spatial measurement, image texture, was somewhat useful in that it was selected as a secondary feature for each of the sampling schemes.

The temporal domain was also relevant, but not in all cases. For example, the features selected for the proportional sampling approach were predominantly located at the end of the temporal sequence. For the equal sampling approach, there was little, if any, temporal signal. This was probably due to the selection of elevation, a static variable, as the primary feature in the subset.

Use of prior probabilities is important, especially in environments where one or two large classes are dominant spatially. The MLC required explicit provision of prior probabilities, while the decision tree and neural network classifiers learned them from the distributions of the training data. Thus, it is important not to compare maximum-likelihood classification results without prior probability information to classifiers that implicitly use this information. An implication for global land cover classification is that the proportion of training samples within classes should not be too different from the proportions within the region to be classified.

For proportional sampling, the neural network with selected features performed best. With equal sampling, the MLC with adjusted prior probabilities was most accurate. The neural network classification was the more accurate of the two, and this difference was statistically significant. Considering that accurate estimates of prior probabilities are not always available, it is of note that Fuzzy ARTMAP produced the next most accurate classification with equal sampling of training data, with no statistical difference between reduced and full measurement spaces. This provides additional evidence to support the use of feature selection as a pre-processing filter for neural network classifiers.

In summary, feature selection using a tree-based classifier greatly reduced the

dimensionality of the input dataset without compromising classification accuracies. Fuzzy ARTMAP was the superior classifier and should perform well if applied to situations with or without dominant cover types. Additional research at other test sites will be necessary to validate these results on a global distribution of land cover types.

Acknowledgments

This work was funded by NASA under the MODIS project, NAS-5-31369, Earth Observing System. Meteorological data were provided by R. d'Entremont, Atmospheric and Environmental Research Inc. and J. Washburne, University of Arizona Department of Hydrology and Water Resources. The authors also wish to thank M. Friedl for comments on an earlier draft of this manuscript.

References

- ATLAS, L., COLE, R., MUTHUSAMY, Y., LIPPMAN, A., CONNOR, J., PARK, D., EL-SHARKAWI, M., and MARKS, R. J., II, 1990, A performance comparison of trained multilayer perceptrons and trained classification trees. *Proceedings of the IEEE*, **78**, 1614–1619.
- BELWARD, A. S. (editor), 1996, The IGBP-DIS global 1 km land cover data set proposal and implementation plans. *IGBP-DIS Working Paper no. 13*, International Geosphere–Biosphere Programme, Toulouse.
- BELWARD, A., and LOVELAND, T. R., 1995, The IGBP-DIS 1-km land cover project. *Proceedings, RSS'95 Remote Sensing in Action, 21st Annual Conference of the Remote Sensing Society, Southampton, UK, 11–14 September 1995* (Nottingham: Remote Sensing Society), pp. 1099–1106.
- BENEDIKTSSON, J. A., SWAIN, P. H., and ERSOY, O. K., 1990, Neural network approaches versus statistical methods in classification of multisource remote sensing data. *IEEE Transactions on Geoscience and Remote Sensing*, **28**, 540–552.
- BENSON, L. D., and DARROW, R. A., 1981, *Trees and Shrubs of the Southwestern Deserts* (Tucson, AZ: University of Arizona Press).
- BERK, A., BERNSTEIN, L. S., and ROBERTSON, D. C., 1989, *MODTRAN: a moderate resolution model for LOWTRAN 7*. Report AFGL-TR-89-0122, Air Force Geophysics Laboratory, Bedford, MA.
- BISCHOF, H., SCHNEIDER, W., and PINZ, A. J., 1992, Multispectral classification of Landsat-images using neural networks. *IEEE Transactions on Geoscience and Remote Sensing*, **30**, 482–490.
- BORAK, J. S., 1996, *MODIS Land Cover Algorithm Version 1 Release* (Seabrook, MD: MODIS Science Data Support Team).
- BREIMAN, L., FRIEDMAN, J. H., OLSEN, R. A., and STONE, C. J., 1984, *Classification and Regression Trees* (Belmont, CA: Wadsworth).
- BROWN, D. E., CORRUBLE, V., and PITTARD, C. L., 1993, A comparison of decision tree classifiers with backpropagation neural networks for multimodal classification problems. *Pattern Recognition*, **26**, 953–961.
- CARPENTER, G. A., GROSSBERG, S., MARKUZON, N., REYNOLDS, J. H., and ROSEN, D. B., 1992, Fuzzy ARTMAP: a neural network architecture for incremental supervised learning of analog multidimensional maps. *IEEE Transactions on Neural Networks*, **3**, 698–713.
- CONGALTON, R. G., ODERWALD, R. G., and MEAD, R. A., 1983, Assessing Landsat classification accuracy using discrete multivariate analysis statistical techniques. *Photogrammetric Engineering and Remote Sensing*, **49**, 1671–1678.
- DICK-PEDDIE, W. A., 1993, *New Mexico Vegetation: Past, Present, and Future* (Albuquerque, NM: University of New Mexico Press).
- EHRlich, D., and LAMBIN, E. F., 1996, Broad scale land-cover classification and interannual climatic variability. *International Journal of Remote Sensing*, **17**, 845–862.
- FISCHER, M. M., and GOPAL, S., 1993, Neurocomputing—a new paradigm for geographic information processing. *Environment and Planning A*, **25**, 757–760.
- GOPAL, S., SKLAREW, D. M., and LAMBIN, E., 1993, Fuzzy-neural networks in multitemporal classification of landcover change in the Sahel. In *New Tools for Spatial Analysis*,

Proceedings of the Workshop (Luxembourg: Office for Official Publications of the European Communities), pp. 69–81.

- GOWARD, S. N., 1989, Satellite bioclimatology. *Journal of Climate*, **2**, 710–720.
- GOWARD, S. N., MARKHAM, B., DYE, D. G., DULANEY, W., and YANG, J., 1991, Normalized difference vegetation index from the Advanced Very High Resolution Radiometer. *Remote Sensing of Environment*, **35**, 257–277.
- GRAHAM, L. A., 1995, *Arizona natural vegetation, as mapped for the Arizona GAP Analysis Program*, Digital GIS file, School of Renewable Natural Resources, University of Arizona, Tucson, AZ.
- GUO, H., and GELFAND, S. B., 1992, Classification trees with neural network feature extraction. *IEEE Transactions on Neural Networks*, **3**, 923–933.
- HANSEN, M., DUBAYAH, R., and DEFRIES, R., 1996, Classification trees: an alternative to traditional land cover classifiers. *International Journal of Remote Sensing*, **17**, 1075–1081.
- HEERMAN, P. D., and KHAZENIE, N., 1992, Classification of multispectral remote sensing data using a back-propagation neural network. *IEEE Transactions on Geoscience and Remote Sensing*, **30**, 81–88.
- HEPNER, G. F., LOGAN, T., RITTER, N., and BRYANT, N., 1990, Artificial neural network classification using a minimal training set: comparison to conventional supervised classification. *Photogrammetric Engineering and Remote Sensing*, **56**, 469–473.
- HUDSON, W. D., and RAMM, C. W., 1987, Correct formulation of the kappa coefficient of agreement. *Photogrammetric Engineering and Remote Sensing*, **53**, 421–422.
- INGEBRITSEN, S. E., and LYON, R. J. P., 1985, Principal components analysis of multitemporal image pairs. *International Journal of Remote Sensing*, **6**, 687–696.
- JACKSON, R. D., 1983, Spectral indices in n-space. *Remote Sensing of Environment*, **13**, 409–421.
- JAMES, M., 1985, *Classification Algorithms* (New York: J Wiley).
- JUSTICE, C. O., TOWNSHEND, J. R. G., HOLBEN, B. N., and TUCKER, C. J., 1985, Analysis of the phenology of global vegetation using meteorological satellite data. *International Journal of Remote Sensing*, **6**, 1271–1318.
- JUSTICE, C. O., TOWNSHEND, J. R. G., and KALB, V. L., 1991, Representation of vegetation by continental data sets derived from NOAA-AVHRR data. *International Journal of Remote Sensing*, **12**, 999–1021.
- KAUFMAN, Y. J., and TANRÉ, D., 1992, Atmospherically resistant vegetation index (ARVI) for EOS-MODIS. *IEEE Transactions on Geoscience and Remote Sensing*, **30**, 261–270.
- KAUTH, R. J., and THOMAS, G. S., 1976, The tasseled cap—a graphic description of the spectral-temporal development of agricultural crops as seen by Landsat. *Symposium Proceedings on Machine Processing of Remotely Sensed Data, Purdue University, West Lafayette, Indiana, 29 June–1 July 1976* (IEEE Catalog Number 76 CH 1103-1 MPRSD), pp. 41–51.
- KEARNEY, T. H., and PEEBLES, R. H., 1951, *Arizona Flora* (Berkeley, CA: University of California Press).
- KUSTAS, W. P., and GOODRICH, D. C., 1994, Preface. *Water Resources Research*, **30**, 1211–1225.
- LOVELAND, T. R., MERCHANT, J. W., OHLEN, D. L., and BROWN, J. F., 1991, Development of a land-cover characteristics database for the conterminous U.S. *Photogrammetric Engineering and Remote Sensing*, **57**, 1453–1463.
- MCLACHLAN, G. J., 1992, *Discriminant Analysis and Statistical Pattern Recognition* (New York: Wiley).
- MICHAELSEN, J., SCHIMMEL, D. S., FRIEDL, M. A., DAVIS, F. W., and DUBAYAH, R. C., 1994, Regression tree analysis of satellite and terrain data to guide vegetation sampling and surveys. *Journal of Vegetation Science*, **5**, 673–686.
- MOODY, A., and STRAHLER, A. H., 1994, Characteristics of composited AVHRR data and problems in their classification. *International Journal of Remote Sensing*, **15**, 3473–3491.
- RICHARDS, J. A., 1994, *Remote Sensing Digital Image Analysis. An Introduction* (New York: Springer).
- RUNNING, S. W., JUSTICE, C. O., SALOMONSON, V., HALL, D., BARKER, J., KAUFMAN, Y. J., STRAHLER, A. H., HUETE, A. R., MULLER, J.-P., VANDERBILT, V., WAN, Z. M., TEILLET, P., and CARNEGIE, D., 1994a, Terrestrial remote sensing science and algorithms planned for EOS/MODIS. *International Journal of Remote Sensing*, **15**, 3587–3620.

- RUNNING, S. W., LOVELAND, T. R., and PIERCE, L. L., 1994b, A vegetation classification logic based on remote sensing for use in global biogeochemical models. *Ambio*, **23**, 77–81.
- SAFAVIAN, S. R., and LANDGREBE, D., 1990, *A survey of decision tree classifier methodology*. Technical Report TR-EE 90-54, Purdue University School of Electrical Engineering, West Lafayette, IN.
- SALOMONSON, V. V., BARNES, W. L., MAYMON, P. W., MONTGOMERY, H. E., and OSTROW, H., 1989, MODIS: advanced facility instrument for studies of the earth as a system. *IEEE Transactions on Geoscience and Remote Sensing*, **27**, 145–153.
- SCOTT, J. M., DAVIS, F., CSUTI, B., NOSS, R., BUTTERFIELD, B., GROVES, C., ANDERSON, H., CAICCO, S., D'EHRCHIA, F., EDWARDS, T. C., JR, ULLIMAN, J., and WRIGHT, R. G., 1993, Gap analysis: a geographic approach to protection of biological diversity. *Wildlife Monographs*, **123**, 1–41.
- SETHI, I. K., 1990, Entropy nets: from decision trees to neural networks. *Proceedings of the IEEE*, **78**, 1605–1613.
- STRAHLER, A. H., MOODY, A., LAMBIN, E., HUETE, A., JUSTICE, C., MULLER, J.-P., RUNNING, S., SALOMONSON, V., VANDERBILT, V., and WAN, Z., 1994, *MODIS Land Cover Product: Algorithm Technical Basis Document, Version 3.1* (Washington, DC: National Aeronautics and Space Administration).
- TOWNSHEND, J. R. G., and TUCKER, C. J., 1984, Objective assessment of Advanced Very High Resolution Radiometer data for land cover mapping. *International Journal of Remote Sensing*, **5**, 497–504.
- TOWNSHEND, J. R. G., JUSTICE, C. O., and KALB, V., 1987, Characterization and classification of South American land cover types using satellite data. *International Journal of Remote Sensing*, **8**, 1189–1207.