

A stepwise regression tree for nonlinear approximation: applications to estimating subpixel land cover

C. HUANG*

Department of Geography, University of Maryland, College Park,
MD 20742, USA

and J. R. G. TOWNSHEND

Department of Geography and Institute for Advanced Computer Studies,
University of Maryland, College Park, MD 20742, USA

(Received 27 September 2000; in final form 11 September 2001)

Abstract. A stepwise regression tree (SRT) algorithm was developed for approximating complex nonlinear relationships. Based on the regression tree of Breiman *et al.* (BRT) and a stepwise linear regression (SLR) method, this algorithm represents an improvement over SLR in that it can approximate nonlinear relationships and over BRT in that it gives more realistic predictions. The applicability of this method to estimating subpixel forest was demonstrated using three test data sets, on all of which it gave more accurate predictions than SLR and BRT. SRT also generated more compact trees and performed better than or at least as well as BRT at all 10 equal forest proportion interval ranging from 0 to 100%. This method is appealing to estimating subpixel land cover over large areas.

1. Introduction

Land cover is one of the major data components in earth science studies and applications (Henderson-Sellers and Pitman 1992, Townshend *et al.* 1994). Traditionally it has been characterized as discrete cover types, i.e. each unit land parcel, or a pixel, is labelled with a single cover type. However, the transition of land cover between unit land parcels is spatially continuous (Wang 1990, Maselli *et al.* 1996). Depending on spatial resolution, pixels often have more than one land cover component in their footprints. Therefore it would be more appropriate and more accurate to characterize such pixels with the proportions of land cover components than to label them with discrete cover types. Many land surface properties required for modelling studies would also be derived with higher accuracies from the continuous distributions of land cover components than from discrete land cover categories. Consequently, characterizing the land surface as the continuous distribution of land cover components in stead of discrete cover types is increasingly required by the user community (DeFries *et al.* 1995).

*Current address: Raytheon ITSS, USGS/EROS Data Center, Sioux Falls, SD 57198, USA; e-mail: huang@usgs.gov.

A general approach to estimating subpixel land cover proportions from remotely sensed data is to model pixel signature from land cover components. Once a model is developed, land cover proportions can be estimated from pixel values through model inversion. By modelling pixel signature in different ways, different methods have been developed. These methods can be grouped into three categories of models: physically based models, spectral mixture models and empirical models. Physically based models are based on physical processes of radiative transfer (e.g. Li and Strahler 1992, Baret *et al.* 1995). Such models are often very complex and generally premature for use in regional land cover applications (Kimes *et al.* 1998). Spectral mixture models are based on the assumption that the spectral signature of a pixel is the mixture of those land cover components, or end-members, within that pixel (Smith *et al.* 1985). The usefulness of this type of models is hindered by two problems. First, many models assume linear mixing among land cover components (e.g. Shimabukuro and Smith 1991, Mathieu *et al.* 1994), which was found invalid in many studies (e.g. Borel and Gerstl 1994, Ray and Murray 1996). Second, spectral end-members identified in mixture analysis, such as green vegetation, non-photosynthetic vegetation and shadow (Roberts *et al.* 1993), do not correspond to such relevant land cover components as forest. Additional efforts are required to estimate the proportion of relevant land cover components (e.g. Adams *et al.* 1995).

While both physically-based models and spectral mixture models depend more or less on knowledge in physical aspects of radiative transfer, empirical models do not depend on such knowledge and have the advantage of giving results that are directly interpretable as the proportion of relevant land cover components. Empirical models are usually developed using linear or nonlinear regression techniques. Linear regression has been employed to estimate subpixel land cover in many studies (e.g. Iverson *et al.* 1989, Zhu and Evans 1994, DeFries *et al.* 1997). A major problem with this approach is that relationships between subpixel land cover proportion and pixel signature may not be linear (e.g. Borel and Gerstl 1994, Ray and Murray 1996). While this problem could be resolved using nonlinear regression techniques, many nonlinear regression techniques require prior knowledge on the nonlinear form of a relationship (Gallant 1987), which is usually unknown in land cover characterization. The purpose of this article is to develop a stepwise regression tree (SRT) method for nonlinear estimation. This method is based on a stepwise linear regression (SLR) technique and the regression tree of Breiman *et al.* (1984) (BRT). Without prior knowledge on the nonlinear form, this method is capable of approximating any form of nonlinear relationship using a set of linear models. The applicability of this method to estimating subpixel land cover is demonstrated by using it to estimate subpixel forest proportions from AVHRR and simulated MODIS images.

2. Development of the stepwise regression tree method

2.1. The least square approach

Most regression techniques implement a least square approach. Suppose there exists a relationship between a dependent variable y and a set of m independent variables $\mathbf{X} = \{x_1, x_2, \dots, x_m\}$:

$$y = f(\mathbf{B}, \mathbf{X}) \quad (1)$$

where \mathbf{B} is a set of necessary parameters. Given sufficient number of simultaneous observations (N) of the dependent and independent variables, \mathbf{B} can be derived by

minimizing the residual sum square (RSS) of the estimation:

$$RSS = \sum_{i=1}^N (y_i - f(\mathbf{B}, \mathbf{X}_i))^2 \quad (2)$$

Linear regression assumes that f is a linear function. To further reduce the impact of multi-collinearity, i.e. correlations among predictor variables (Gunst and Mason 1980), stepwise linear regression techniques that establish prediction models using least correlated variables have been developed. The stepwise linear regression (SLR) method evaluated in this study exhausts all possible combinations of predictor variables in searching for the one that minimizes the impact of multi-collinearity without losing a significant portion of the explanatory power of a data set (Miller 1990).

2.2. The regression tree of Breiman et al. (1984)

The above linear regression assumes linear relationships between subpixel land cover and satellite signal, which are often invalid over large areas. As mentioned earlier, use of nonlinear regression techniques is not feasible because the nonlinear function f is often unknown (Gunst and Mason 1980). To overcome this problem, Breiman *et al.* (1984) proposed a decision tree approach, which will be referred to as Breiman's regression tree (BRT) hereafter, to nonlinear regression. This method recursively partitions the observations into subsets such that the total RSS of all subsets is minimized. The RSS of the k th subset is calculated as:

$$RSS_k = \sum_{i=1}^{N_k} (\hat{y}_i - y_i)^2 \quad (3)$$

where N_k is the number of samples in the k th subset, \hat{y}_i and y_i the predicted and actual values of the response variable. The total RSS is the sum of those of all subsets (leaf nodes of a regression tree). In each subset the predicted value \hat{y}_i is not estimated from independent variables using linear regression techniques. Rather it simply takes the mean value of the dependent variable of all training samples within that subset. Figure 1(a) shows how this method can approximate a nonlinear

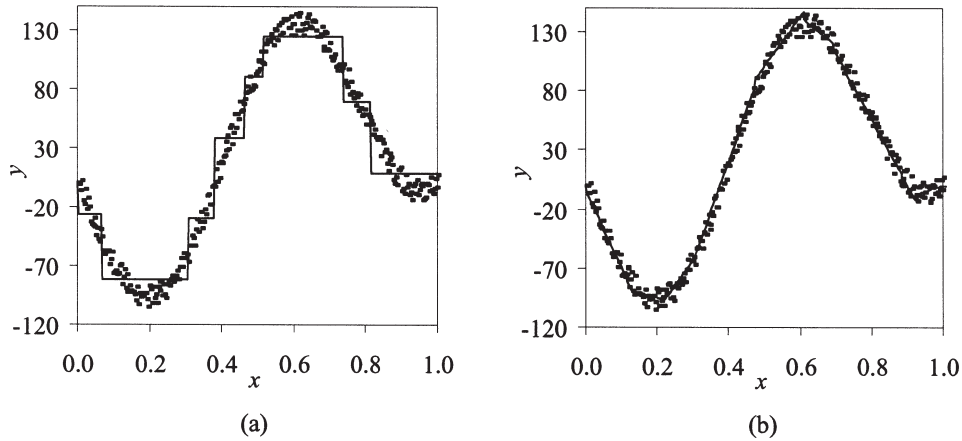


Figure 1. A comparison of (a) BRT and (b) SRT predictions (line) on data samples (\cdot) simulated using an arbitrary function. x and y are arbitrary variables with no unit, representing the independent and dependent variables respectively.

relationship using a set of discrete predictions without knowing the mathematical form of that relationship. However, a major problem with this method is that the prediction is not continuous. As illustrated in figure 1(a), BRT predictions may consist of substantial local errors even for a data set that can be fitted using a curve.

2.3. Stepwise regression tree

The above analysis reveals that the advantages and limitations of the least square approach and BRT are somewhat complementary. The least square approach may give more accurate predictions but there is a need to know the mathematical form of the target relationship in advance, while BRT does not need to know the mathematical form of the target relationship but tends to give less accurate predictions. Such limitations might be mitigated by combining the two approaches. A main reason for the BRT giving less satisfactory predictions is that in each subset the dependent variable \hat{y}_i is predicted as a single value—the mean of the response variable of all observations in that subset. The variance of independent variables in each subset is not accounted for in making a prediction. This problem can be alleviated by using a least square regression method to estimate the dependent variable \hat{y}_i from independent variables in each subset.

The stepwise regression tree employs a stepwise linear regression technique to estimate the dependent variable in each subset. Compared to nonlinear regression techniques, linear regression has the advantage of being simple and fast. While a relationship may not be linear globally, a linearity assumption may be valid for a small subset of the data. Stepwise linear regression has a further advantage of allowing one to use a small subset of least correlated variables without losing a significant portion of the explanatory power of the data, thus minimizing the impacts of multi-collinearity on the regression and on result analysis (Gunst and Mason 1980). Stepwise linear regression techniques have been detailed in many textbooks (e.g. Miller 1990).

Figure 2 gives the pseudo code of the stepwise regression tree (SRT). Initially all observations are in a set, the root node of the regression tree. The programme starts from the root node and recursively partitions the samples into subsets until no leaf node can be further split. A node is split into two subsets when a split leads to substantial reduction in the residual sum square (RSS) of estimation. For a given node RSS is calculated using equation (3), and is referred to as RSS_{node} . Similarly the RSS s of the two subsets after a split, referred to as RSS_{left} and RSS_{right} , are also calculated using this equation. The total RSS after a split, referred to as $RSS_{\text{splitting}}$, is the sum of RSS_{left} and RSS_{right} . The minimum $RSS_{\text{splitting}}$ of all possible splits of a node is referred to as $MinRSS_{\text{splitting}}$. A node can not be further split when the number of samples in that node is less than a predefined minimum node size or when the improvement from splitting that node is less than a predefined threshold value. The improvement from splitting a node is calculated as:

$$\text{Improvement} = \frac{RSS_{\text{node}} - MinRSS_{\text{splitting}}}{RSS_{\text{node}}} \times 100\% \quad (4)$$

Figure 1(b) shows that SRT likely will give more accurate and more realistic predictions than BRT on a data set simulated using an arbitrary nonlinear function.

3. Application to estimating subpixel forest cover

The applicability of the stepwise regression tree to deriving subpixel land cover was tested in comparison with Breiman's regression tree (BRT) and a stepwise linear

1. Initially all observations are in the root node;
2. Start from the root node, partition the samples using the following recursive procedure;
3. Giving a tree node;
 4. Perform stepwise linear regression on samples in current node;
 5. Calculate the residual sum square (RSS_{node});
 6. If (number of samples > predefined minimum node size)
 7. For each predictor variable x_i ;
 8. Sort x_i in ascending order;
 9. For each value d in the above sorted list
 10. Using d as the threshold value, partition the samples within current node into two subsets;
 11. Perform stepwise regression on each subset and calculate the corresponding RSS ;
 12. Calculate the subtotal RSS as the sum of those of the two subsets;
 13. End of for loop on threshold value d ;
 14. Find the minimum subtotal RSS achievable from splitting current node on x_i ;
 15. End of for loop on predictor variable x_i ;
 16. Find the minimum RSS of all possible splits of current node ($MinRSS_{splitting}$);
 17. Calculate the improvement from splitting the current node using equation (4)
 18. If (Improvement > predefined minimum improvement)
 19. Split current node into two new nodes using the variable and threshold value that give the $MinRSS_{splitting}$;
 20. For each of the two new nodes, go to step 3;
 21. End of if in step 18;
 22. End of if in step 6;
 23. End of recursion.

Figure 2. Pseudo code of the stepwise regression tree method.

regression (SLR) method using simulated and actual images. Spectral mixture models, though having been employed to estimate subpixel land cover in many studies (e.g. Quarmby *et al.* 1992, Settle and Drake 1993), were not considered in this comparison for reasons mentioned in section 1.

3.1. Data and preprocessing

A moderate resolution imaging spectroradiometer (MODIS) image simulated from Landsat thematic mapper (TM) data and two advanced very high resolution radiometer (AVHRR) images were used in this study (table 1). These two sensors were selected because they have been among the major sensor systems for collecting data sets for global land cover study (Running *et al.* 1994, Hansen *et al.* 2000), and because at their spatial resolutions, pixels are more likely to be mixed than at finer resolutions.

The MODIS image was simulated from a TM subscene acquired on 14 August 1985. The subscene covered an area of $22.8 \times 22.8 \text{ km}^2$ around Annapolis, Maryland, USA. The TM digital number was converted to at-satellite reflectance according to

Table 1. Characteristics of data sets used in this study.

Data set	Source image and pixel size	Independent variables
Maryland data set	MODIS image simulated from TM data, 256.5 m	Landsat TM bands 1 to 5, 7 and NDVI
Central African Republic (CAR) data set (1) and (2)	AVHRR images, 1 km	Annual minimum, maximum and mean of AVHRR bands 1 to 4 and NDVI, bands 1 to 4 of the greenest month, and bands 1 to 3 and NDVI of the warmest month

Markham and Barker (1986). Atmospheric correction was not performed because the image was quite clear within the study area. A simulation program embedded with the point spread functions of both TM and MODIS was used to simulate 250 m MODIS pixels from the TM subscene for all 6 spectral bands (Barker and Burelhach 1992). The output pixel size was adjusted to 256.5 m so that each MODIS pixel lies on exactly 9×9 28.5 m TM pixels. The size of the output MODIS image is 88 by 88 pixels. The normalized difference vegetation index (NDVI) was calculated from 256.5 m red and near infrared (NIR) bands using the following equation:

$$NDVI = \frac{NIR - RED}{NIR + RED} \quad (5)$$

A simple land cover map was developed at the 28.5 m resolution by visually interpreting the TM subscene and aerial photographs covering the same area. Field work was conducted to resolve confusions and to validate this map, which was found highly accurate at the 28.5 m resolution. By overlaying 256.5 m MODIS grids on this land cover map, an image of subpixel forest proportions was derived at 256.5 m resolution. This is the reference forest proportion image for the Maryland data set.

The two Central African Republic (CAR) data sets cover two areas of the Central African Republic, each 160 km \times 160 km. The images were taken from a global 1 km AVHRR data set processed at the EROS data center of the US Geological Survey (USGS) under the guidance of the international geosphere biosphere programme (Eidenshink and Faundeen 1994, Townshend *et al.* 1994). Each data set consists of 23 annual matrices derived from 12 monthly composites for the phenological year between April 1992 and March 1993 following previous studies (DeFries *et al.* 1997, DeFries *et al.* 1998, Hansen *et al.* 2000), including the maximum, minimum and mean of AVHRR bands 1 to 4 and NDVI, bands 1 to 4 of the greenest month (the month having maximum NDVI), and bands 1 to 3 and NDVI of the warmest month (the month having maximum band 4 temperature). AVHRR band 5 was not included because it was found highly correlated with band 4. Estimated from this data set, the coefficient of determinations (R^2) between the two bands were 1.0 for 6 of the 12 monthly composites, and were higher than 0.9 for all 12 monthly composites.

Reference forest proportion images for the two CAR data sets were derived by overlaying 1 km grids on a 30 m forest map derived from Landsat TM images acquired in early 1990s. Within each 1 km grid forest proportion was calculated as the proportion of TM forest pixels within that grid. Developed with intensive human

interventions by the University of Maryland through the Landsat Pathfinder Humid Tropical Deforestation Project (LPHTDP),¹ this map is deemed highly reliable, especially for the purposes of this study because only forest is concerned.

3.2. Subpixel forest cover estimation

Each of the three data sets was split into a training subset and a test subset. For the Maryland data set, 20% (or 1548) pixels were randomly selected from the subscene as training samples. For the two CAR data sets, 10% pixels (2560) were selected as training samples. The sizes of the training samples were deemed sufficient for training the three algorithms in the three study areas.

One of the advantages of stepwise linear regression over regular regression is that the former can substantially reduce the number of predictor variables to be used without losing a substantial portion of explanatory power of the data. Figure 3 shows that for each data set, the standard error of estimation is a function of the number of predictor variables. For the Maryland data set, a large portion of the explanatory power of seven variables can be represented by two variables (TM band 5 and NDVI), while for the two CAR data sets, three variables (mean AVHRR band 4, mean NDVI, and NDVI of the warmest month) carry most of the explanatory power of 23 variables. Therefore the maximum number of variables to be selected by SLR was limited to two for the Maryland data set and three for the two CAR data sets. Little improvement can be achieved by using more variables.

For both BRT and SRT, the recursive partitioning of data samples into subsets needs to be controlled in order to yield trees of appropriate size. While underfitted trees may not be able fully to explore the explanatory power of data, overfitted ones may generalize poorly to unseen samples (Breiman *et al.* 1984, Quinlan 1993). Table 2 gives the conditions under which the splitting of a node should be stopped for the three data sets. These conditions were empirically based. They yielded trees of reasonable sizes for the three data sets. Based on figure 3, the number of predictor variables to be selected by the stepwise regression of SRT in each node was limited to two for the Maryland data set and three for the two CAR data sets.

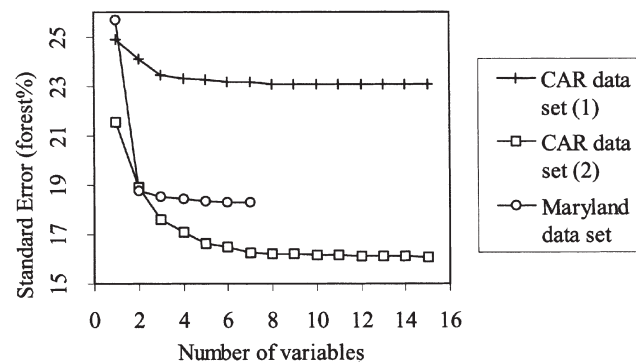


Figure 3. Standard error of estimation as a function of the number of input variables in stepwise linear regression analysis.

¹Details of this project are reported at: <http://www.geog.umd.edu/tropical/main.html>.

Table 2. Conditions under which the splitting of a node should be stopped in regression tree analysis.

BRT	SRT
Node RSS <1% of root node or Node size <80	Improvement <0.1 or Node size <80, 150*

Note: *80 for the Maryland data set, 150 for the two CAR data sets

3.3. Performance assessment

The accuracy of predicted subpixel proportions was measured using the root mean square of error (*RMSE*) of the prediction:

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (\hat{y}_i - y_i)^2} \quad (6)$$

where \hat{y} and y are the predicted and actual subpixel forest proportions respectively, and N the number of testing samples. An overall *RMSE* as well as the *RMSE* values at 10 equal intervals of sub-pixel forest proportion ranging from 0 to 100% were calculated for each method on each data set (table 3 and figure 4).

3.3.1. SLR versus regression tree methods

Table 3 shows that in overall SRT outperformed SLR on all three data sets. This is expected because by minimizing the residual sum square (*RSS*) of estimation by partitioning data samples into subsets, theoretically SRT should always perform better than or at least as well as SLR or any other linear regression techniques. If the relationship between a dependent variable and predictor variables is linear, little improvement in *RSS* can be achieved from partitioning data samples into subsets (CAR data set (1), table 3). However, if the relationship is nonlinear, partitioning data samples into subsets such that the relationship within each subset is linear or quasi-linear can remove a substantial portion of *RSS* in the prediction (Maryland data set and CAR data set (2), table 3). The structure of a developed tree reveals how *RSS* is reduced as the tree grows (figure 5). In this tree developed for the Maryland data set, each splitting of a node removed a portion of model *RSS*. In overall splitting the training data set into nine subsets reduced the *RSS* by 61%. In other words, the *RSS* of SRT (total *RSS* of all leaf nodes) was only 39% of that of SLR (*RSS* of root node). As the root square of averaged *RSS* (equation (5)), the *RMSE* estimated from training samples using the SRT should be about 62% (root

Table 3. *RMSE* of predicted subpixel forest proportion and percent *RMSE* reduced by using SRT instead of SLR and BRT.

Method	Maryland data set	CAR data set (1)	CAR data set (2)
RMSE of predicted subpixel forest proportion (%)			
SLR	17.00	23.02	16.72
BRT	12.25	23.60	14.68
SRT	11.50	22.21	13.50
RMSE reduced by using SRT instead of SLR and BRT (%)			
SLR	32.35	3.52	19.26
BRT	6.12	5.89	8.04

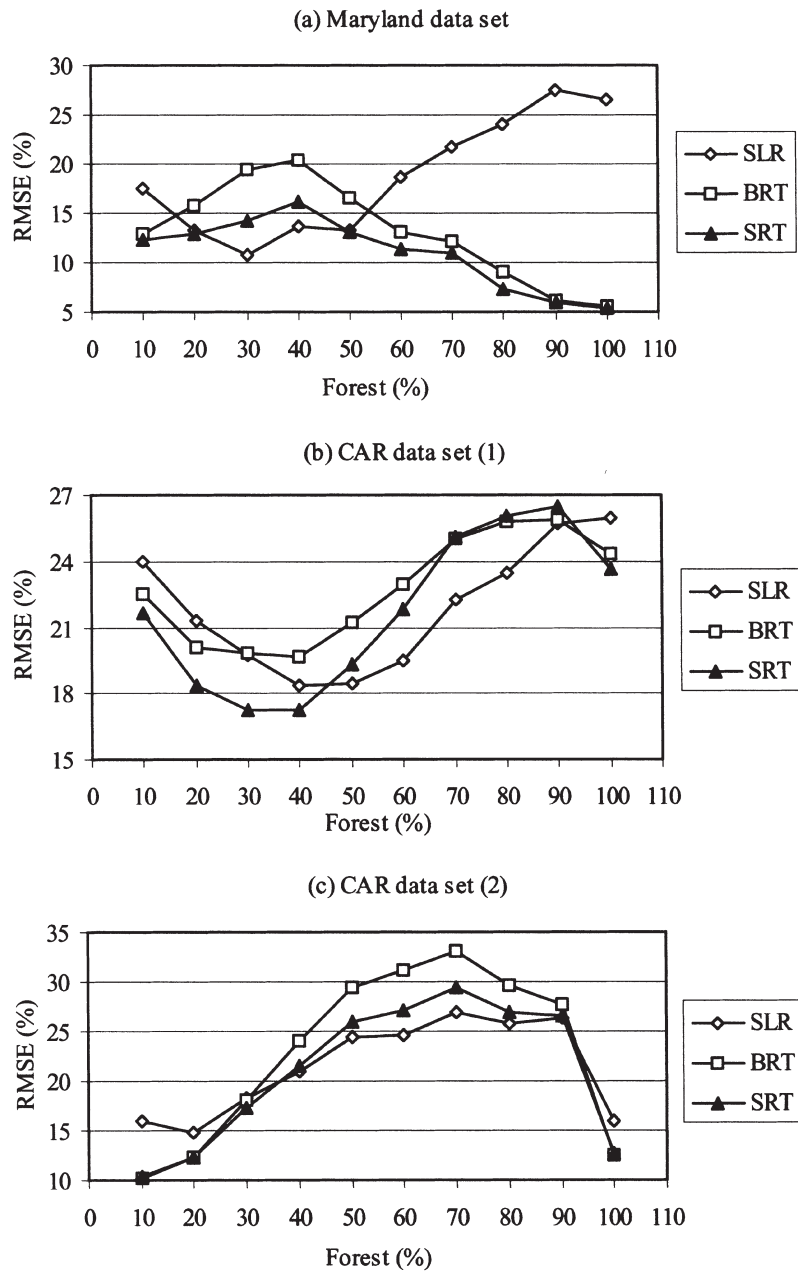


Figure 4. RMSE of predicted subpixel forest proportion at 10 equal forest proportion intervals ranging from 0 to 100%. The value on the x-axis refers to the upper limit of an interval, e.g. the value 10 refers to the interval between 0 and 10%, the value 20 the interval between 10 and 20%, and so on.

square of 39%) of that estimated using SLR. According to table 3, the *RMSE* of forest proportion estimated using SRT from test samples was 67.65% of that estimated using SLR (table 3).

BRT had lower *RMSE* values than SLR on the Maryland data set and CAR

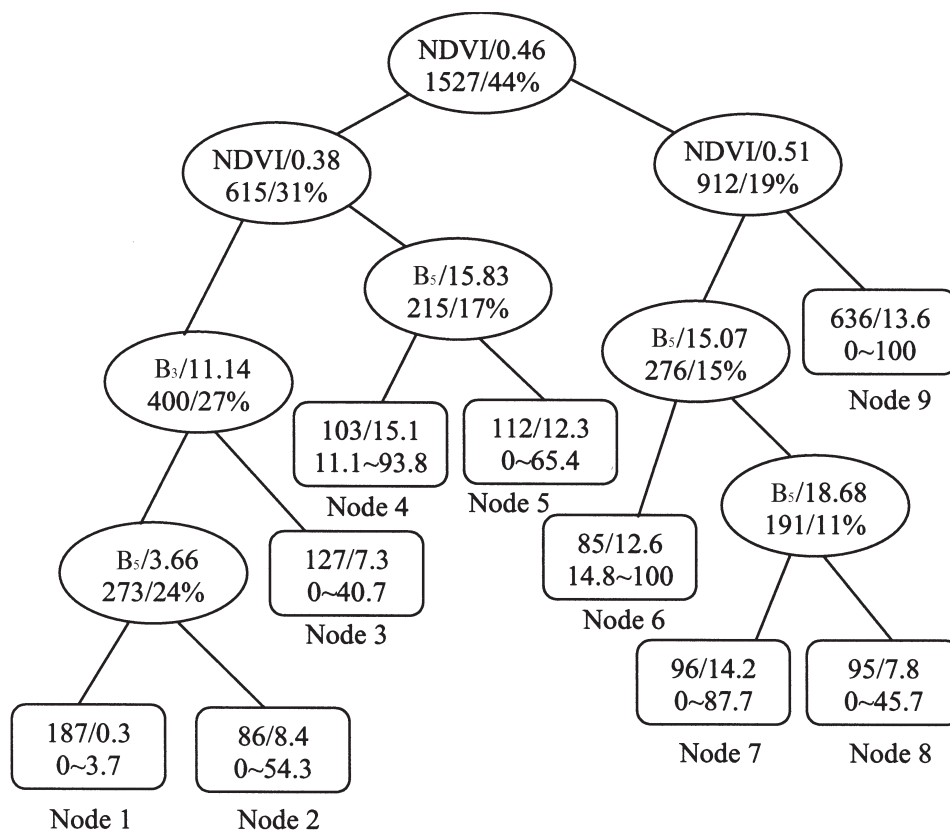


Figure 5. The stepwise regression tree for estimating subpixel forest proportion from the Maryland data set. Ellipsoids represent non-leaf nodes and rounded rectangles represent leaf nodes. The left node of a splitting has splitting variable \leq threshold and the right node has splitting variable $>$ threshold. B_1 , B_2 , B_3 , B_4 , B_5 and B_7 in non-leaf nodes represent TM bands 1 to 5 and 7. The four values in non-leaf nodes: upper—splitting variable/threshold value, lower—number of samples in current node/improvement in RSS from splitting current node. The four values in leaf nodes: upper—number of samples in current node/RMSE of prediction model for current node, lower—lower ~ upper limits of the predicted value from current prediction model (%).

data set (2), but had a slightly higher RMSE value than SLR on the CAR data set (1). This is probably because most pixels in the Maryland data set and CAR data set (2) had either very low or very high subpixel forest proportions, while pixels in the CAR data set (1) had subpixel forest proportion distributed more evenly over the entire range of forest proportion (0–100%). Figure 4 shows that both BRT and SRT had substantially lower RMSE values than SLR in predicting extremely low or high subpixel forest cover. In areas where forest and non-forest were highly mixed, e.g. forest proportion intervals roughly between 20% and 50% for the Maryland data set, and between 50% and 80% for the CAR data sets (1) and (2), the two tree-based methods had higher RMSE values than SLR. Similar results on the comparative performances between BRT and SLR were reported in DeFries *et al.* (1997).

3.3.2. BRT versus SRT

Table 3 shows that SRT gave more accurate predictions than BRT on all three data sets. While their differences in overall *RMSE* were small on these data sets, figure 4 shows that SRT predictions were almost always more accurate than or at least as accurate as BRT predictions at all 10 forest proportion intervals. In addition, SRT has several desirable properties. First, It gave more realistic predictions than BRT (figure 6). While subpixel forest cover can be anywhere between 0 and 100% in a pixel, for reasons discussed in section 2.2., BRT only gave predictions at a limited set of discrete values. Second, SRT produced smaller trees than BRT. Table 4 shows that trees generated by SRT on the three data sets had 50% to 90% fewer leaf nodes than those generated by BRT. With substantially fewer leaf nodes, a smaller tree has fewer decision rules to interpret, and because each leaf node is associated with more data samples, the decision rules are less likely to be trivial, and therefore should be better generalized to unseen samples (Quinlan 1993).

Another appealing property of SRT is that the developed decision rules may be useful for stratification purpose. Figure 7 shows the spatial distribution of the leaf nodes of the stepwise regression tree developed from the Maryland data set (figure 5). Some leaf nodes delineated regions that are physically distinctive. For example, nodes 1, 2 and 3 in figure 5 delineated open water, coastal zone and urban area, respectively (figure 7). Node 9 delineated a region characterized by the mixing of forest and grass/crop. The other five leaf nodes did not show physically distinctive spatial patterns. The leaf nodes were modeled differently and had different errors (table 5). The SRT trees for the two CAR data sets also exhibited certain spatial patterns. However, it was not clear if their leaf nodes delineated physically distinctive regions, because the coarse resolution of these data sets prevented visual identification of physically distinctive regions.

4. Discussion

The above analysis demonstrated the better performance of SRT over SLR and BRT in estimating subpixel forest proportion. It would be interesting to compare this method to other commonly used subpixel estimation techniques like spectral mixture analysis (SMA). However, SMA can not be directly used to estimate subpixel forest proportion, because forest is not a spectral end-member. The pure forest pixels, i.e. those with 100% forest cover, could not be used to represent a spectral end-member, because they scattered in wide ranges in the spectral spaces. SRT does not have this problem. It can be used to estimate any land cover component, as long as adequate training samples containing the subpixel proportion of that component are available.

Remotely sensed data sets can have from several to several hundred variables. In theory, SRT should be able to handle a large number of input variables. However, its training speed may decrease rapidly as the number of input variables increases. One solution to this problem is to use techniques like principal component analysis to condense the data into fewer variables without losing a significant portion of explanatory power.

As discussed earlier, the extrapolation capability of a tree-based model may be compromised if the model is underfitted or overfitted. SRT is controlled by two parameters: improvement as calculated in equation (4) and minimum node size. A sensitivity analysis using more data sets should reveal whether there exist universal threshold values that can be used to prevent developing an ill-fitted tree. Our

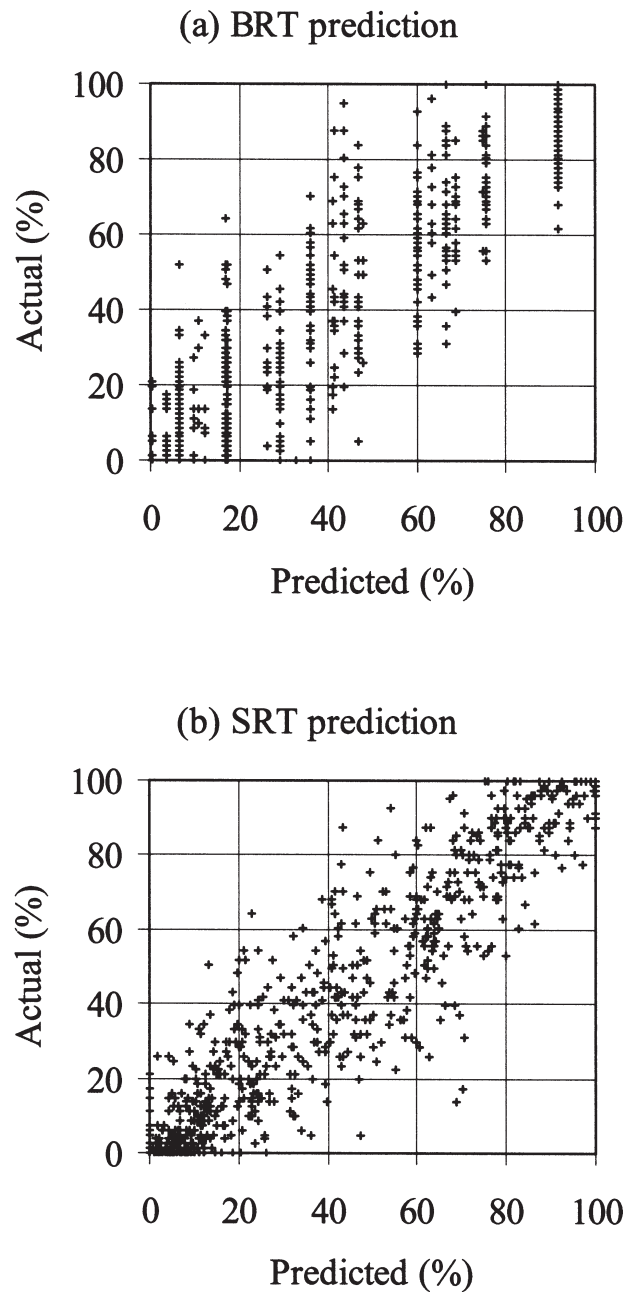


Figure 6. Distribution of subpixel forest proportions estimated from the Maryland data set using (a) BRT and (b) SRT.

experiences with some other tree-based algorithms including BRT and the classification tree of Quinlan (1993) suggest that some trials are always needed in order to develop an appropriately pruned tree for a specific data set.

Deriving land cover information from satellite images over a large area is often far more complex than over a small study area. In addition to the many issues

Table 4. Number of leaf nodes of SRT and BRT on the three data sets.

	SRT	BRT
Maryland data set	9	26
CAR data set (1)	5	52
CAR data set (2)	8	17

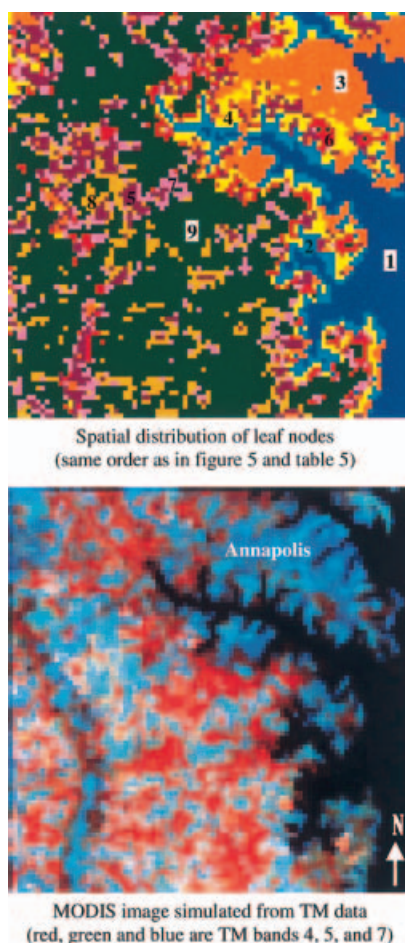


Figure 7. Spatial distribution of leaf nodes of the stepwise regression tree shown in figure 5. The image covers an area of 22.8 km by 22.8 km around Annapolis, Maryland.

arising from handling a large data volume, the relationship between land cover and satellite signal is often more complex in a large area than in a small area. And stratification based on *a priori* knowledge is often required to accommodate such spatial variation of a relationship (e.g. Zhu and Evans 1994, DeFries *et al.* 1997). The above analysis of SRT results from the Maryland data set suggests that with adequate training samples, SRT may provide an alternative approach to simultaneous stratification and model development for large area applications. As a supervised method, its performance relies heavily on the availability and quality of training

Table 5. Regions delineated by the leaf nodes of SRT on the Maryland data set.

Tree node no.	Region characteristics	Prediction model Forest (%)=	Standard error of prediction (forest %)
1	Open water body	0	0.3
2	Coastal area	$19.1B_5 - 28.5B_7 - 25.5^*$	8.4
3	Urban	$78.0NDVI - 1.6B_5 + 11.1$	7.3
4	Not obvious	$6.0B_4 - 8.7B_7 - 47.9$	15.1
5	Not obvious	$4.0B_4 - 5.2B_5 - 2.8$	12.3
6	Not obvious	$57.1B_1 - 70.1B_2 - 18.2$	12.6
7	Not obvious	$10.6B_1 - 9.5B_5 + 14.6$	14.2
8	Not obvious	$-1.4B_4 - 3.0B_5 + 111.8$	7.8
9	Forest-crop/grass mixing	$7.2B_1 - 14.2B_5 + 153.8$	13.6

Note: * B_i is the at-satellite reflectance (%) for TM band i .

data sets. For regional applications, the training pixels do not have to be selected randomly and to be over 10% or more of a study area, as they did in this study. In order to avoid possible biases in model prediction, however, the training samples do need to represent an adequate sampling of both the dynamic range of subpixel forest proportion and the spectral and spatial variability of that area. Fortunately, this is no longer an insurmountable limitation for many regions, as digital air photo and high resolution satellite data are becoming cheaper and are being made available in more and more areas. Deriving training data from high resolution images is practically more feasible than field work, and has been adopted in many similar studies (e.g. DeFries *et al.* 1997).

5. Conclusions

A stepwise regression tree (SRT) algorithm was developed for approximating complex nonlinear relationships. This algorithm is based on the regression tree of Breiman *et al.* (1984) (BRT) and a stepwise linear regression (SLR) method. It represents an improvement over SLR in that it can approximate a nonlinear relationship by decomposing it into a set of linear or quasi-linear ones, and over BRT in that it tries to explain the variance of the dependent variable in each subset using the variance of independent variables. Theoretically, therefore, SRT should always outperform SLR and BRT. Experimental results on three test data sets confirmed this observation. It made predictions with less error than BRT and SLR on all three data sets. While its overall errors were only slightly lower than those of the BRT, it performed better than or at least as well as BRT at all 10 equal forest proportion intervals ranging from 0 to 100%. In addition, SRT generated smaller trees that were easier to interpret and less likely to be overfitted, and produced more realistic predictions than BRT. SRT may also provide an alternative approach to stratification, which is often necessary for large area applications but often requires certain *a priori* knowledge over a study area. These properties of SRT are appealing to deriving subpixel land cover components over large areas. Some land cover components like forest are relevant to many applications, but they may not be directly derived using spectral mixture analysis techniques if they are not spectral end-members.

Acknowledgments

This study is made possible through an NSF grant (BIR9318183) and a contract from the National Aeronautics and Space Administration (NAS596060). The MODIS simulation programme was provided by Kai Yang of Science Systems and Applications, Inc. Comments and suggestions from anonymous reviewers greatly enhanced the quality of this article.

Reference

- ADAMS, J. B., SABOL, D. E., KAPOS, V., FILHO, R. A., ROBERTS, D. A., SMITH, M. O., and GILLESPIE, A. R., 1995, Classification of multispectral images based on fractions of endmembers: application to land-cover change in the Brazilian Amazon. *Remote Sensing of Environment*, **52**, 137–154.
- BARET, F., CLEVERS, J. G. P. W., and STEVEN, M. D., 1995, The robustness of canopy gap fraction estimates from red and near-infrared reflectances: a comparison of approaches. *Remote Sensing of Environment*, **54**, 141–151.
- BARKER, J. L., and BURELHACH, J. W., 1992, MODIS image simulation from Landsat TM imagery, in *ASPRS/ACSM/RT92, Washington, DC, August 3–8, 1992* (Washington, DC: ASPRS), pp. 156–165.
- BOREL, C. C., and GERSTL, S. A. W., 1994, Nonlinear spectral mixing models for vegetative and soil surfaces. *Remote Sensing of Environment*, **47**, 403–417.
- BREIMAN, L., FRIEDMAN, J. H., OLSHEND, R. A., and STONE, C. J., 1984, *Classification and regression trees* (Belmont, CA: Wadsworth International Group).
- DEFRIES, R., HANSEN, M., STEININGER, M., DUBAYAH, R., SOHLBERG, R., and TOWNSHEND, J., 1997, Subpixel forest cover in central Africa from multisensor, multitemporal data. *Remote Sensing of Environment*, **60**, 228–246.
- DEFRIES, R. S., FIELD, C. B., FUNG, I., JUSTICE, C. O., LOS, S., MATSON, P. A., MATTHEWS, E., MOONEY, H. A., TOWNSHEND, J. R. G., TUCKER, C. J., USTIN, S. L., and VITOUSEK, P. M., 1995, Mapping the land surface for global atmosphere-biosphere models: toward continuous distributions of vegetation's functional properties. *Journal of Geophysical Research*, **100**, 20867–882.
- DEFRIES, R. S., HANSEN, M., TOWNSHEND, J. R. G., and SOHLBERG, R., 1998, Global land cover classifications at 8km spatial resolution: the use of training data derived from Landsat imagery in decision tree classifiers. *International Journal of Remote Sensing*, **19**, 3141–3168.
- EIDENSHINK, J. C., and FAUNDEEN, J. L., 1994, The 1 km AVHRR global land data set: first stages in implementation. *International Journal of Remote Sensing*, **15**, 3443–3462.
- GALLANT, A. R., 1987, *Nonlinear statistical models* (New York: John Wiley & Sons).
- GUNST, R. F., and MASON, R. L., 1980, *Regression analysis and its application: a data-oriented approach* (New York: Marcel Dekker, Inc.).
- HANSEN, M., DEFRIES, R. S., TOWNSHEND, J. R. G., and SOHLBERG, R., 2000, Global land cover classification at 1 km spatial resolution using a classification tree approach. *International Journal of Remote Sensing*, **21**, 1331–1364.
- HENDERSON-SELLERS, A., and PITMAN, A. J., 1992, Land-surface schemes for future climate models: specification, aggregation and heterogeneity. *Journal of Geophysical Research*, **97**, 2678–2696.
- IVERSON, L. R., COOK, E. A., and GRAHAM, R. L., 1989, A technique for extrapolating and validating forest cover across large regions: calibrating AVHRR data with TM data. *International Journal of Remote Sensing*, **10**, 1805–1812.
- KIMES, D. S., NELSON, R. F., MANRY, M. T., and FUNG, A. K., 1998, Attributes of neural networks for extracting continuous vegetation variables from optical and radar measurements. *International Journal of Remote Sensing*, **19**, 2639–2663.
- LI, X., and STRAHLER, A. H., 1992, Geometric optical bidirectional reflectance modeling of the discrete crown vegetation canopy: effects of crown shape and mutual shadowing. *IEEE Transactions on Geoscience and Remote Sensing*, **30**, 276–292.
- MARKHAM, B. L., and BARKER, J. L., 1986, Landsat MSS and TM post-calibration dynamic ranges, exoatmospheric reflectances and at-satellite temperatures. *EOSAT Landsat Technical Notes*, **1**, 3–8.

- MASELLI, F., RODOLFI, A., and CONESE, C., 1996, Fuzzy classification of spatially degraded Thematic Mapper data for the estimation of sub-pixel components. *International Journal of Remote Sensing*, **17**, 537–551.
- MATHIEU, S., BERTHOD, M., and LEYMARIE, P., 1994, Determination of proportions and entropy of land use mixing in pixels, in *IGARSS'94, Pasadena, California* (Pasadena, CA: Institute of Electrical and Electronics Engineers), pp. 798–800.
- MILLER, A. J., 1990, *Subset selection in regression* (New York: Chapman and Hall).
- QUARMBY, N. A., TOWNSHEND, J. R. G., and SETTLE, J. J., 1992, Linear mixture modeling applied to AVHRR data for crop area estimation. *International Journal of Remote Sensing*, **13**, 415–425.
- QUINLAN, J. R., 1993, *C4.5 programs for machine learning* (San Mateo, CA: Morgan Kaufmann Publishers).
- RAY, T. W., and MURRAY, B. C., 1996, Nonlinear spectral mixing in desert vegetation. *Remote Sensing of Environment*, **55**, 59–64.
- ROBERTS, D. A., SMITH, M. O., and ADAMS, J. B., 1993, Green vegetation, nonphotosynthetic vegetation, and soils in AVIRIS data. *Remote Sensing of Environment*, **44**, 255–269.
- RUNNING, S. W., JUSTICE, C. O., SALOMONSON, V., HALL, D., BARKER, J., KAUFMANN, Y. J., STRAHLER, A. H., HUETE, A. R., MULLER, J. P., VANDERBILT, V., WAN, Z. M., TEILLET, P., and CARNEGIE, D., 1994, Terrestrial remote sensing science and algorithms planned for EOS/MODIS. *International Journal of Remote Sensing*, **15**, 3587–3620.
- SETTLE, J. J., and DRAKE, N. A., 1993, Linear mixing and the estimation of ground cover proportions. *International Journal of Remote Sensing*, **14**, 1159–1177.
- SHIMABUKURO, Y. E., and SMITH, J. A., 1991, Least-squares mixing models to generate fraction images derived from remote sensing multispectral data. *IEEE Geoscience and Remote Sensing*, **29**, 16–20.
- SMITH, M. O., JOHNSON, P. E., and ADAMS, J. B., 1985, Quantitative determination of mineral types and abundances from reflectance spectra using principal components analysis. *Journal of Geophysical Research*, **90**, C797–C804.
- TOWNSHEND, J. R. G., JUSTICE, C. O., SKOLE, D., MALINGREAU, J.-P., TEILLET, J. C. P., SADOWSKI, F., and RUTTENBERG, S., 1994, The 1 km resolution global data set: needs of the International Geosphere Biosphere Programme. *International Journal of Remote Sensing*, **14**, 3417–3441.
- WANG, F., 1990, Fuzzy supervised classification of remote sensing images. *IEEE Transactions on Geoscience and Remote Sensing*, **28**, 194–201.
- ZHU, Z., and EVANS, D. L., 1994, US forest types and predicted percent forest cover from AVHRR data. *Photogrammetric Engineering & Remote Sensing*, **60**, 525–531.