Predicting Soil Drainage Class Using Remotely Sensed and Digital Elevation Data

A.T. Cialella, R. Dubayah, W. Lawrence, and E. Levine

Abstract

Soil drainage strongly affects the patterns and processes of ecosystems, including biomass production, vegetation community distribution, soil development, aeration, hydrologic processes, and trace gas fluxes. To obviate the need for extensive field surveys, we present a technique to use a remotely sensed optical image and digital elevation data to predict soil drainage class at a 6- by 4-km research site in a mixed conifer forest in Howland, Maine. Elevation, detrended elevation, slope, aspect, and flow accumulation were determined from a 10-m resolution digital elevation model (DEM) of the site. Normalized Difference Vegetation Index (NDVI) data derived from the Advanced Visible and Infrared Imaging Spectrometer (AVIRIS) were used to represent differences in vegetation cover. Classification tree analysis predicted soil drainage class with an average of 78 percent accuracy.

Introduction and Background

Natural soil drainage is an important factor influencing many biophysical processes, such as biomass production and pattern, nutrient cycling, and trace gas fluxes (e.g. see Jenny, 1941; Ferwerda and Young, 1981; Harley and Smith, 1983; Chapin and Shaver, 1985; Sahrawat and Keeney, 1986; Brady, 1990; Meng and Seymour, 1992; MacDonald et al., 1993; Levine et al., 1994). Natural soil drainage is the frequency and duration of periods when the soil is free of saturation or partial saturation (Brady, 1990). State and county soil surveys which report this soil characteristic refer to it qualitatively as a drainage class with categories ranging from "very-poorly" to "excessively-well" drained.

Determining soil drainage class is laborious, time consuming, and costly because it requires field examination of a soil profile. Typically, a soil pit must be excavated and filled with water to produce saturation, allowed to drain naturally, then refilled in order to measure the time it takes for the water to recede. A level of subjectivity is also included because the drainage class is determined based on this direct observation and on inference from other local landscape observations (see Soil Survey Staff (1993) for further information on determining soil drainage class). These difficulties have led investigators to seek alternate methods of mapping drainage class.

Jenny (1941), in a classic study, identified the five essential components of soil formation as topography, vegetation, parent material, climate, and time. Information on two of

Jenny's soil forming factors, topography and vegetation, is available without resorting to *in situ* observations. Both of these in turn are related to soil drainage. Digital topographic databases, such as the 30-m digital elevation model series (DEM), exist for most of the U.S. Various studies have shown that data derivable from DEMs, such as landscape shape (e.g., level, low-lying), slope, aspect, and flow accumulation (where and how water will accumulate over a landscape), are related to soil drainage class (Troeh, 1964; Kirkby, 1978; O'Loughlin, 1986; Wood et al., 1990; Bell et al., 1992; Meng and Seymour, 1992).

Information about vegetation biophysical properties can be obtained using remote sensing techniques (Curran, 1983; Asrar et al., 1984; Tucker et al., 1985; Prince, 1991; Townshend et al., 1991; Daughtry et al., 1992; Price, 1993; Dubayah, 1994). A general relationship between grouped soil series and phytomass development was reported by Lozano-Garcia et al. (1991) using a remotely sensed vegetation index. Levine et al. (1994) observed statistical differences between soil drainage class and variation in species composition, as expressed through a vegetation index.

A statistical technique, classification tree analysis, can (as one of its applications) utilize these factors to arrive at predictions of soil drainage class. Classification tree analysis is based on binary recursive partitioning of the predictor variable data to arrive at homogeneous groupings of the response variable data for each hierarchical level (Chambers and Hastie, 1992). (Here, classification trees are differentiated from regression trees in that the former may have a categorical response variable, such as drainage class, while the response variable in the latter is numeric, such as would be produced by a least-squares regression (Breiman et al., 1993).) The process creates a hierarchical tree whose branch points correspond to partitions of the data, and the leaves at the end of branches are the final classes. Classification tree analysis often may provide insight into the physical structure of the data. It also has an advantage in its consideration of categorical data for both dependent and/or independent variables, and independent variables which may be severely non-normal. Nonadditive behavior is more easily captured in tree-based models as compared to standard linear models, where interactions between variables must be pre-specified and are of a particular multiplicative form (Breiman et al., 1993).

Recently, tree-based models have been used as a method of classification (e.g., Hockman et al., 1990; Grubb and King, 1991; Baker et al., 1993). Michaelsen et al. (1987) investigated factors controlling the establishment of Blue oak seedlings in the central Coast Ranges of California using tree

Department of Applied Science, Brookhaven National Laboratory, Bldg. 490D, Upton, NY 11973-5000.

R.C. Dubayah and W.T. Lawrence are presently with the Department of Geography, University of Maryland, LeFrak Hall, Room 2137, College Park, MD 20742.

E.R. Levine is presently with the Biospheric Sciences Branch, NASA/Goddard Space Flight Center, Greenbelt, MD 20771.
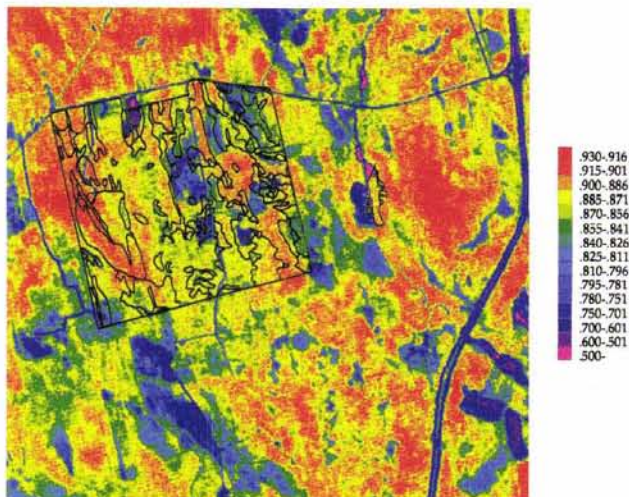
Plate 1. Broad-band NDVI calculated from AVIRIS for an 8.5- by 8-km area within the International Paper Company's Northern Experimental Forest (NEF), near Howland, Maine. Soil associations are outlined in black for the 6- by 4-km study area.

structures. Davis *et al.* (1990) predicted a continuous vegetation index as a site classifier from digital terrain and land-use data and Davis and Dozier (1990) applied tree techniques to develop an ecological land classification to classify terrain for a complex region in southern California. Most recently, Michaelsen *et al.* (1994) applied a tree-based model to classify a tall grass prairie landscape using digital terrain, land-use, and land-cover data. They concluded that tree-based model analyses are useful for analyzing complex hierarchical relationships often found in multivariate data sets.

Inspired by these results and Jenny's (1941) soil forming factors, we have applied classification tree analysis to predict soil drainage class. This technique incorporates physical environmental factors that control soil drainage that are determined from remotely sensed and digital elevation data. In particular, we use remote sensing to derive a vegetation index, and digital elevation data for finding topographic properties including slope, aspect, and flow accumulation. From these data, a classification tree is constructed and used to produce maps of soil drainage class.

Methodology

Study Area
The field site is a 6- by 4-km section of boreal forest located near Howland, Maine (Latitude: 45.2°N; Longitude: 68.7°W) within the International Paper Company's Northern Experimental Forest (NEF). This area of central Maine has a cold and humid climate with annual precipitation between 80 and 100 cm. The temperatures range from −20°C in the winter to less than 30°C in the summer. The site is low lying and relatively flat, with a maximum relief of 110 metres from east to west. Soil drainage ranges from excessively drained to very-poorly drained soils. Glacial activity played a critical role in the formation of the soil and its characteristics which now exist in the area (Levine *et al.*, 1994).

Data Acquisition
A DEM was created for the 6- by 4-km site at a 10-metre grid spacing. The DEM was interpolated from hand digitized con-

tours of a 7.5-minute USGS topographic map at a scale of 1:24,000. An image from the Advanced Visible and Infrared Imaging Spectrometer (AVIRIS), a 224-band high spectral resolution instrument (Lillesand and Kiefer, 1994), was used to derive an atmospherically corrected vegetation index — the Normalized Difference Vegetation Index (NDVI) — at a 20-m resolution (Plate 1) for bands 0.626 to 0.694μm (red) and 0.776 to 0.901μm (infrared) (Levine *et al.*, 1994). A soil survey identifying major soil series and soil associations at a scale of 1:12,000 was conducted with the aid of color aerial photographs and SPOT satellite imagery for the 6- by 4-km site (Levine *et al.*, 1994). Soil mapping units (composed of ten soil associations) were created at a 10-m resolution and spatially represented as soil polygons in a geographic information system (GIS). A corresponding soil drainage map was also produced from profile description information collected during the soil survey (Plate 2). Soil drainage class per soil mapping unit was derived by weighted means of drainage classes assigned to soil series within a soil mapping unit. A complete description of the methodology used to prepare the soils database can be found in Levine *et al.* (1994).

Data Development
From the initial DEM raster data, several other variables were derived. A grid of flow accumulation was produced using an algorithm for calculating flow direction and accumulation (Jenson and Domingue, 1988). Slope and aspect grids were also computed from the DEM.

The initial NDVI grid included areas of human disturbance, such as roads and managed clear cuts, as well as undisturbed vegetation. The disturbed areas were masked from the grid to reduce the possibility of introducing error into the relationship between the vegetation index, NDVI, and soil drainage class.

The terrain of the site has a gradual rise in elevation from the southeast corner to the northwest corner. This rise overwhelms any local variations in topography. The overall elevation trend was removed from the surface to leave only local variations. Therefore, grids of flow direction and flow
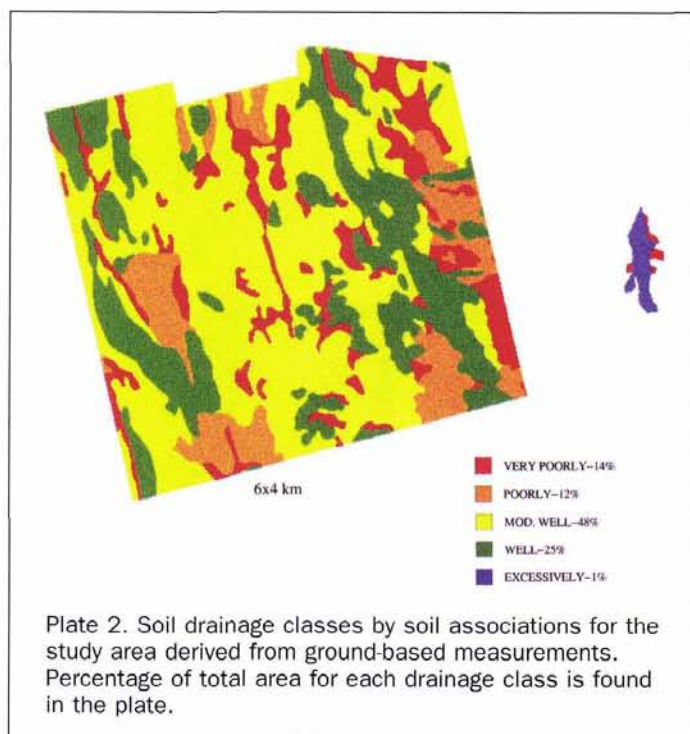

Plate 2. Soil drainage classes by soil associations for the study area derived from ground-based measurements. Percentage of total area for each drainage class is found in the plate.

81.0% Accuracy

VERY POORLY–60.7%
POORLY–91.3%
MOD. WELL–86.1%
WELL–77.0%
EXCESSIVELY–100.0%

6x4 km

(a)

75.3% Accuracy

VERY POORLY–27.1%
POORLY–62.7%
MOD. WELL–91.6%
WELL–76.2%
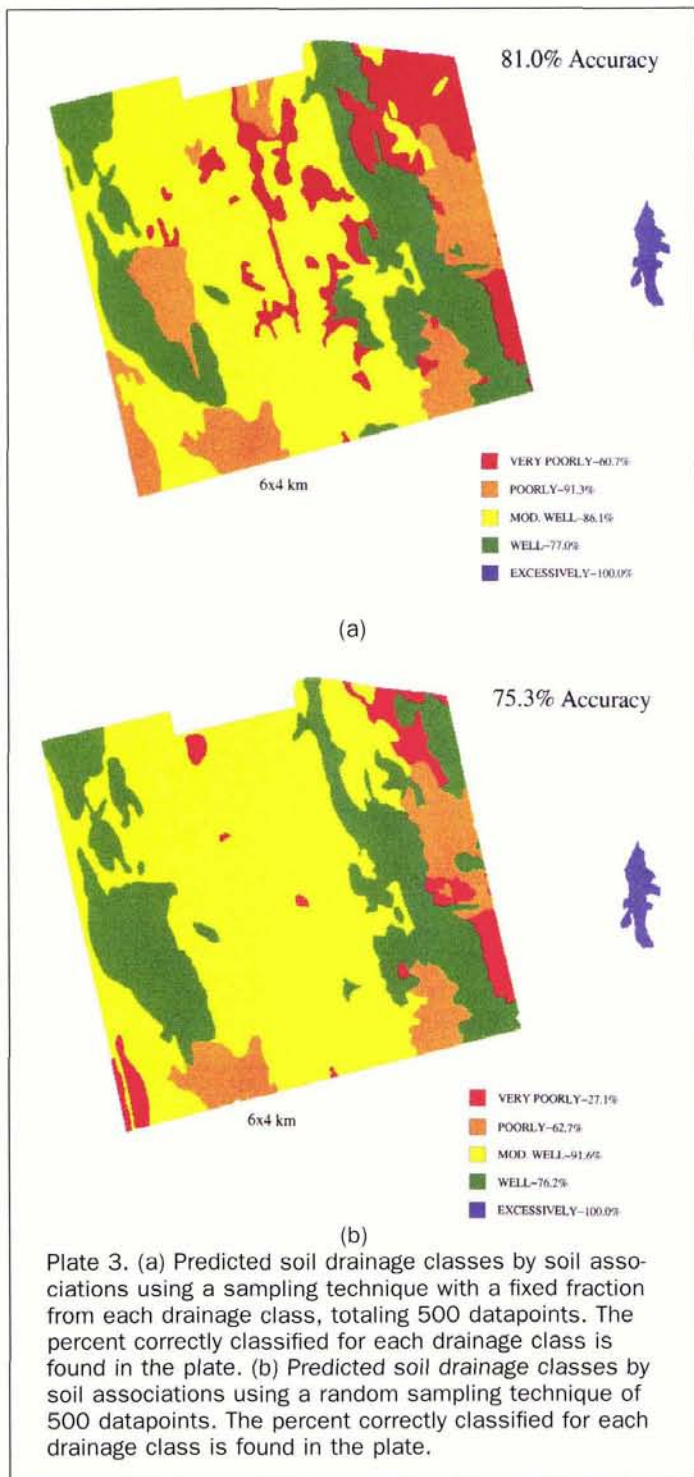EXCESSIVELY–100.0%

6x4 km

(b)

Plate 3. (a) Predicted soil drainage classes by soil associations using a sampling technique with a fixed fraction from each drainage class, totaling 500 datapoints. The percent correctly classified for each drainage class is found in the plate. (b) Predicted soil drainage classes by soil associations using a random sampling technique of 500 datapoints. The percent correctly classified for each drainage class is found in the plate.

accumulation (hereafter referred to as localized flow direction and localized flow accumulation) were produced to reflect this detrended elevation surface.

Correlational analyses using Spearman's Rank Order Correlation were conducted between soil drainage class and the physical variables of elevation, slope, aspect, NDVI, flow accumulation and direction, detrended elevation surface, and localized flow direction and flow accumulation. All data were aggregated by soil polygons because the smallest ground-based measurements available were at the soil polygon level.

Classification Tree Analysis

Our analyses were performed using the S-Plus statistical software (Statistical Sciences, Inc., Seattle, Washington). Classification tree analysis requires a set of classification or predictor variables and a single-response variable, any or all of which may be categorical. The initial tree-growing methodology employs an algorithm based on binary recursive partitioning of the predictor dataset to arrive at increasingly homogeneous regions of the space of predictor variables (X), such that within each region the conditional distribution of y given x, $f(y|x)$, no longer depends on x. At an individual node, j, xj is partitioned into xj, LEFT and xj, RIGHT such that $f(yj|xj$, LEFT) and $f(yj|xj$, RIGHT) are most different. The best splits are determined for each individual predictor variable by this recursive looping. The best of all splits is chosen as the split for that node. At each node all variables contribute some amount to the total predictability, but a node is assigned based on the one variable that is the most predictive. The tree branching continues until the number of datapoints is too small to split and still remain reliable (Chambers and Hastie, 1992).

Creating Samples

Because the data in our study were composed of ten 6- by 4-km grids (one each for vegetation, soil drainage class, elevation, detrended elevation, slope, aspect, flow accumulation, flow direction, and localized flow accumulation and flow direction), each 548 by 404 pixels, there was an abundance of data points with which to grow and test the tree. However, these data were sampled to minimize the possibility of spatial autocorrelation (Cliff and Ord, 1981). A tree's accuracy was investigated with a test dataset chosen from a set of points reserved from the total pool. Exploratory trees were created to test both the sample size of the dataset used to create the tree and the sample size of dataset used to test the tree.

Pruning

Once the tree was grown in the manner described above, a misclassification error rate for the tree was reported, based on the average misclassification error rates for each node. The error rate reported an estimate of the number of points that were incorrectly classified out of the total number of points (e.g., $110/500 = 0.22$). This error rate is determined within the classification tree analysis based on the inverse of the probability of classifying a pixel correctly. The more splits there are in the tree, the smaller the probability of misclassification; however, the predictive capability of the tree is reduced with a larger tree. The tree was grown as large as possible and then selectively pruned upward, producing a sequence of trees of decreasing complexity. Each tree's predictability was tested with an independent test dataset. Test samples were tested on pruned trees of various sizes to choose the sub-tree having the lowest estimated misclassification error rate.

Testing the Accuracy

While the misclassification error rate is a measure of the predictive power of the tree, that rate is calculated on a per-pixel basis. The smallest unit of ground-based measured soil drainage class data available to us was a soil association, or soil mapping unit, spatially represented by soil polygons across the site. Thus, the only test of predictability that could be verified was by soil associations. The prediction data output from each tree were grouped by soil polygons and compared to the ground-based measurements. A weighted-by-area count of correctly classified pixels (within polygons) showed which tree's classification compared best

Size of Sample (pixels)	Mean Misclassification Error Rate (%)	Number of Endnodes
40000	31.1	113
20000	30.9	114
15000	30.8	115
1000	24.6	88
500	22.0	64
100	23.0	13
32677*	34.0	107
1000*	23.0	59
500*	20.4	50
100*	25.0	13

*fixed fraction from each drainage class

with the ground-based measurements. However, because soil polygons are not pure with respect to drainage class, yet they are treated as such, the validation map itself will have errors. A pixel falling within a given polygon may be assigned a class different from the polygon and be labeled as misclassified. In reality, the pixel may be correctly classified, but without sub-polygon scale validation data, this is impossible to tell. Hence, we cannot distinguish map errors from model errors, and our accuracy test is somewhat weakened.

Predictor Variable Analyses
Datasets used to create the trees were varied as to the combination of predictor variables included. Some variables showed more predictive power than others. One predictor variable was removed from the dataset at a time and a new tree created to ascertain its influence on predictability. Mean misclassification error rates were noted with each new tree.

Cost versus Complexity
The number of terminal nodes in a classification tree has a direct effect on the mean misclassification error rates produced by the tree. Too many nodes increases the complexity and the likelihood of misclassification. Too few nodes creates poorer fits, though the cost of the tree measured in number of terminal nodes is reduced (Breiman *et al.*, 1993). An analysis was conducted where a cost complexity parameter, k, (Breiman *et al.*, 1993) was varied from 1 to 10. The gradual reduction and then increase in mean misclassification error rates was noted as the number of terminal nodes decreased. The most cost-effective tree was determined by the lowest mean misclassification error rate for a given cost complexity parameter.

Results
Several trees were produced with varying sample sizes and two sampling techniques, and their mean misclassification error rates were compared (Table 1). The sample sizes varied from 100 to 40,000 points. The two sampling techniques were (1) random sample and (2) a fixed fraction from each soil drainage class, selected randomly. For both sampling techniques and varying sample sizes, the smallest mean misclassification error rate was reported with a sample size of 500 pixels. Thus, two trees created with 500 pixels by each sampling technique were chosen for analysis.

When sample size of a test dataset (data used to test the accuracy of the tree) was varied from 15,000 pixels to 1,000 pixels, the misclassification error rate increased by only 0.3 percent for the smaller test sample. The larger sample size was chosen to estimate the accuracy of the tree for both sampling techniques, so that a representative sample of pixels within each soil polygon would be available.

An analysis was conducted to estimate the best pruned

sub-tree based on the cost versus complexity associated with the number of terminal nodes (Breiman *et al.*, 1993). The cost-complexity analysis is reported in Table 2 for both trees: one created with 500 random sample points (TREE-A) and the other created with a fixed fraction from each drainage class, selected randomly, totaling 500 points (TREE-B). The table lists the mean misclassification error rates of trees pruned for a range of cost complexity parameter values, k, for both the pruned tree and a test data sample. The k values of 4 and 5 yielded the same lowest misclassification error rates for a test sample for TREE-A (a k value of 5 was chosen arbitrarily). Although k values of 1 and 2 yielded the lowest misclassification error rates for TREE-B, a k value of 6 was nearly equivalent in terms of error rates and was chosen due to its fewer number of terminal nodes. Two trees were pruned using those k values, resulting in structures with 51 and 43 terminal nodes. Characteristically, the gradual reduction in estimated error rate observed in Table 2 is caused by a tradeoff between too few splits in a tree, creating poor fits, and too many splits, increasing the likelihood of predicting many of the wrong class (Breiman *et al.*, 1993).

Two prediction trees were created using the best pruned trees and a predictor dataset. These prediction trees provided the data necessary to produce soil drainage class maps. Plate 3 shows the drainage class maps created from the two prediction trees as compared to the soil drainage class map created from ground based measurements (Plate 2). PRED-B, a map created from the pruned tree with a fixed fraction from each soil drainage class, was 7 percent more accurate (Plate 3a) than the map created from a random sample, PRED-A (Plate 3b). In particular, PRED-B more accurately estimated the very-poorly drained and poorly drained soil polygons than did PRED-A. Both maps predicted the single excessively drained soil polygon accurately. The percentages of correctly classified polygons (by pixel) per drainage class are reported in Table 3. The weighted mean percent accuracy was 75.3 percent for PRED-A and 81.0 percent for PRED-B.

The percentages of misclassification for each drainage class (errors of omission) are reported for PRED-A and PRED-B in Table 4. When PRED-A misclassified a soil polygon, it was classified as either the largest or second largest soil drainage class of the map, moderately well drained and well drained. However, misclassifications of soil polygons in PRED-B were

Classification Tree	k	Number of terminal nodes	Mean Misclassification Error Rates (%) Pruned Tree	Test Sample
TREE-A	1	61	22.2	45.9
	2	57	22.8	45.5
	3	55	23.0	44.5
	4	51	23.8	43.9
	5	51	23.8	43.9
	6	48	24.6	44.5
	7	45	26.0	45.9
	8	36	29.8	45.5
	9	32	30.6	44.8
	10	27	31.8	44.1
TREE-B	1	49	20.4	46.9
	2	49	20.4	46.9
	3	47	20.4	47.1
	4	46	20.6	47.3
	5	44	21.0	47.1
	6	43	21.2	47.0
	7	39	23.0	48.4
	8	30	25.0	47.3
	9	30	25.0	47.3
	10	26	26.6	49.7

TABLE 3. PREDICTION ACCURACY.

Prediction Map	Drainage Class	% Area	% Correct
PRED-A	Very Poorly	14	27.1
	Poorly	12	62.7
	Moderately Well	48	91.6
	Well	25	76.2
	Excessively	1	100.0
	Percent overall accuracy		75.3
PRED-B	Very Poorly	14	60.7
	Poorly	12	91.3
	Moderately Well	48	86.1
	Well	25	77.0
	Excessively	1	100.0
	Percent overall accuracy		81.0

most often classified as a drainage class with similar characteristics.

Partial dendrograms of pruned TREE-A (Figure 1a) and pruned TREE-B (Figure 1b) show comparisons of the nodes and class predictions for each tree. Both trees initially split on the predictor variable of elevation. The left and right splits at the next level were also made on the same predictor variables for both trees; detrended elevation and NDVI, respectively. Below those splits, the trees begin to deviate in the number of nodes and the variables on which the splits occurred.

Table 5 gives the mean misclassification error rates for TREE-A and TREE-B, with individual predictor variables removed singly. The greatest change in misclassification rates was produced for both trees when the elevation was excluded from the predictor variables dataset. Excluding detrended elevation surface, reflecting local topographic position, produced the second greatest change in misclassification rates for both TREE-A and TREE-B. Omitting vegetation, as expressed by NDVI, produced the next greatest change in misclassification rates for TREE-B. The next largest change for TREE-A was produced when aspect was excluded, although excluding NDVI produced similar results.

Discussion

Accuracy of the Estimation
The prediction map created from a sample composed of a fixed fraction from each soil drainage class produced a more accurate estimate than did the map created from a random sampling of the data. Because the greater part of the area was moderately well drained, the training set used to grow the tree from a random sample incorporated more moderately well drained data in the tree. Fewer pixels from the drainage classes with smaller percent areas were included in the training set. Hence, the prediction tree created from a random sample was biased toward the moderately well drained class. Using a simplified analytical model, Breiman et al. (1993) suggest that constructing a test sample from a fixed fraction from each class produces a more accurate estimation, as we found.

Spatial analysis of soil drainage misclassification suggests that the data manipulation procedures used to create the validation (ground truth) map may have contributed to some misclassification. First, the process of aggregating soil drainage class data from the soil series level to the soil association level forced a single drainage class to be assigned to a soil association, as discussed earlier, leading to possible validation map errors. Further, this assignment was the weighted average of each drainage class for each series within the association, with these averages then rounded to the nearest

whole number (Levine et al., 1994), leading to more generalization. In addition, some soil series were too small to map, although they may have still shown variation in NDVI and terrain variables.

Explanatory Properties of the Classification Tree
One advantage of this classification method is that it may provide insight into the structure of the data. The underlying physical mechanisms that generated the predictor variable data were interpreted through tree dendrograms and split competitions at the top of the trees. Topography had the greatest correlation with soil drainage class. In particular, an elevation of approximately 240 ft was consistently assigned the first split in both trees. This data split value can be explained through physical features of the landscape. Broad riparian swamps (including alder swamps, beaver swamps, flowage, and bogs) occurred below that contour interval. The detrended elevation surface, reflecting local topographic position, was also highly influential. Removing the horizontal

TABLE 4a. ACCURACY OF PREDICTION VS. GROUND-BASED MEASUREMENTS (PRED-A).

Ground-Based Measurements	Prediction (%)				
	very-poorly drained	poorly drained	moderately well drained	well drained	excessively drained
very-poorly	**27.1**	6.1	51.5	13.5	1.6
poorly	0.0	**62.7**	12.5	24.7	0.0
moderately well	0.8	0.0	**91.6**	7.6	0.0
well	4.7	2.2	16.9	**76.2**	0.0
excessively	0.0	0.0	0.0	0.0	**100.0**

TABLE 4b. ACCURACY OF PREDICTION VS. GROUND-BASED MEASUREMENTS (PRED-B).

Ground-Based Measurements	Prediction (%)				
	very-poorly drained	poorly drained	moderately well drained	well drained	excessively drained
very-poorly	**60.7**	15.9	11.2	10.5	1.6
poorly	0.0	**91.3**	4.7	4.0	0.0
moderately well	9.1	0.7	**86.1**	4.0	0.0
well	4.1	7.4	11.4	**77.0**	0.0
excessively	0.0	0.0	0.0	0.0	**100.0**

TABLE 5. MEAN MISCLASSIFICATION ERROR RATES FOR EXCLUDED PREDICTOR VARIABLES.

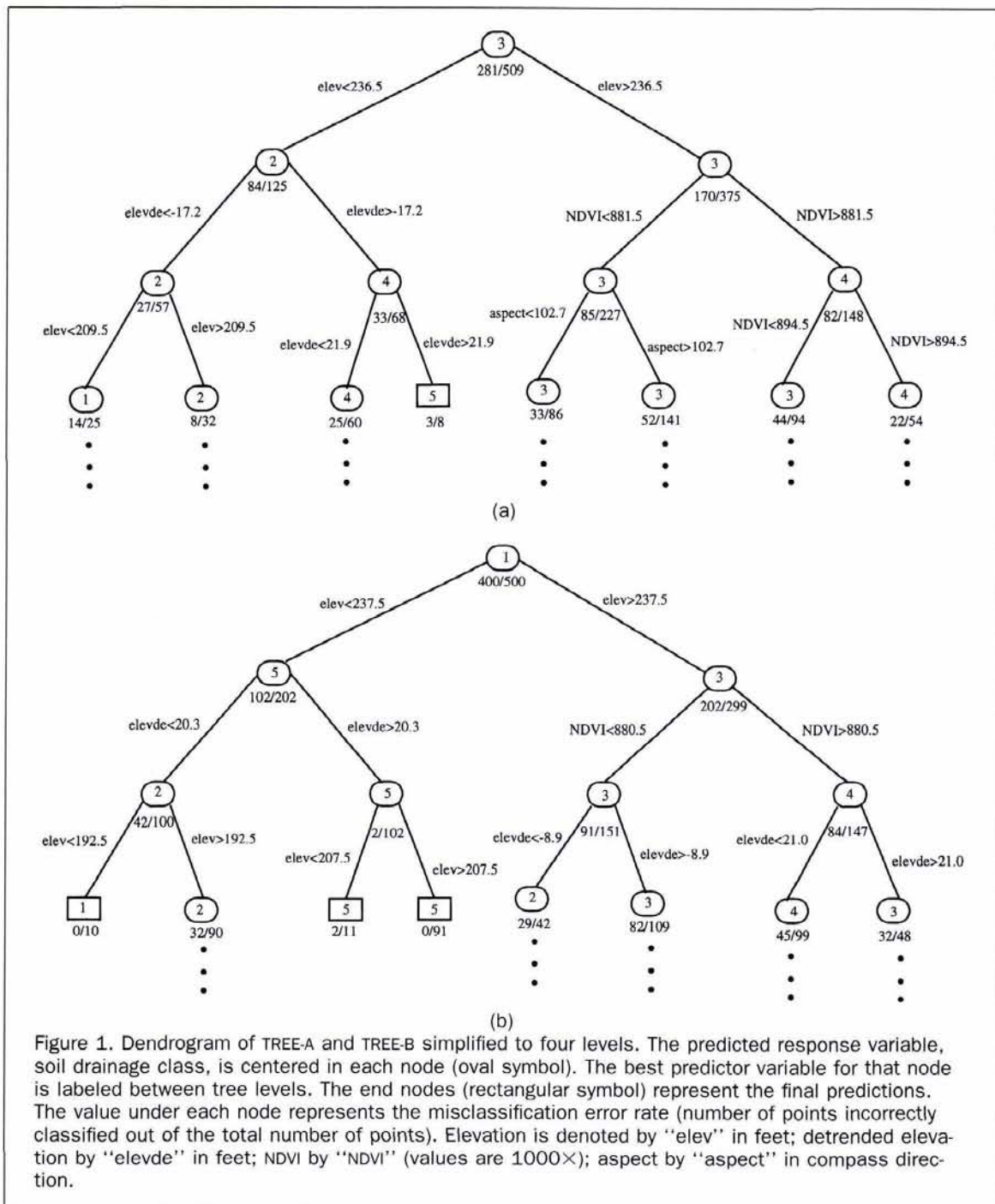| Classification Tree	Predictor Variable Excluded	Misclassification Error Rate (%)
TREE-A	(none)	22.0
	elevation	27.0
	slope	22.2
	aspect	23.8
	detrended DEM	25.2
	NDVI	23.4
	accumulation	22.2
	local accum.	21.2
TREE-B	(none)	20.4
	elevation	25.4
	slope	20.2
	aspect	19.8
	detrended DEM	25.0
	NDVI	21.4
	accumulation	20.8
	local accum.	21.2

Figure 1. Dendrogram of TREE-A and TREE-B simplified to four levels. The predicted response variable, soil drainage class, is centered in each node (oval symbol). The best predictor variable for that node is labeled between tree levels. The end nodes (rectangular symbol) represent the final predictions. The value under each node represents the misclassification error rate (number of points incorrectly classified out of the total number of points). Elevation is denoted by "elev" in feet; detrended elevation by "elevde" in feet; NDVI by "NDVI" (values are 1000×); aspect by "aspect" in compass direction.

trend in slope enhanced the local relief, exposing upland slopes, for example, and the hummocky terrain of the study site.

Vegetation, as expressed by NDVI, was the next strongest predictor of soil drainage class. The NDVI values for this study area may have effectively separated conifers from the hardwoods, with conifers generally having a lower NDVI. Conifers are known to grow on more poorly drained soils than hardwoods.

Slope, flow accumulation, and localized flow had little correlation with soil drainage class for this study area. If these data were used to construct a stream network and distance from the stream was added as a dataset, they may have

had a higher correlation. Aspect also may be more important in an area with steeper slopes than the study area because, in general, aspect influences soil formation and vegetation with slopes greater than 15 percent (Klingebiel *et al.*, 1988).

Finally, it should be noted that, although previous studies have shown the appropriateness of a classification tree approach (Davis *et al.*, 1990; Michaelsen *et al.*, 1994), we are imposing a hierarchichal structure as a model assumption when no such structure may exist. Examination of our tree structures shows the importance of elevation; many of the splits are on this variable and detrended elevation. Successive splits on a single variable, even if non-monotonic in nature, suggest that a non-hierarchical structure may be present

in the data. In particular, a class of relatively new models known as general additive models may be appropriate and warrant further study (Chambers and Hastie, 1992).

Future Applicability

The primary limitation to applying classification tree analysis (and to all statistical tree-based models) in other study areas is that some representative sampling of soil drainage class data is required to produce a classification tree. The number of soil drainage class training data points required to produce accurate estimates is not known. Our results suggest that a minimum of 100 point pedon samples would be required for a reasonably accurate estimate (see Table 1). An important future study would be to test a larger region where the soil drainage class data were available at specific point locations, rather than aggregated to the soil polygon level. The amount of point data required to make accurate estimates could then be investigated, as well as the overall applicability of the technique at a regional scale.

Conclusion

A technique was developed that incorporates a physically based statistical procedure to estimate soil drainage class from physical variables that can be derived from remotely sensed and digital elevation data. Our results suggest that soil drainage class can be predicted with reasonable accuracy from these data for a boreal forest study site. In addition, the technique can provide insight into the physical factors that exert the strongest influence over soil drainage class for an area. In the future, with minimal amounts of soil drainage class training data, this technique might be applied to produce a continuous map of soil drainage class at a regional scale that would be valuable for investigations of a wide range of ecosystems processes.

Acknowledgments

References

Asrar, G., M. Fuchs, E.T. Kanemasu, and J.L. Hatfield, 1984. Estimating absorbed photosynthetic radiation and leaf area index from spectral reflectance in wheat, *Agronomy Journal*, 76:300–306.

Baker, F.A., D.L. Verbyla, and C.S. Hodges, Jr., 1993. Classification and regression tree analysis for assessing hazard of pine mortality caused by heterobasidion annosum., *Plant Disease*, 77(2):136–139.

Bell, J.C., R.L. Cunningham, and M.W. Havens, 1992. Calibration and validation of a soil-landscape model for predicting soil drainage class, *Soil Science Society of America Journal*, 56:1860–1866.

Brady, N., 1990. *The Nature and Properties of Soils*, Macmillan, New York, 621 p.

Breiman, L., J. Friedman, R. Olshen, and C. Stone, 1993. *Classification and Regression Trees*, Chapman and Hall, New York, 358 p.

Chambers, J.M., and T.J. Hastie (editors), 1992. *Statistical Models in S*, Wadsworth, Pacific Grove, California, 587 p.

Chapin, F.S., and G.R. Shaver, 1985. Arctic, *Physiological Ecology of North American Plant Communities* (B.F. Chabot and H.A. Mooney, editors), Chapman and Hall, New York, pp. 16–40.

Cliff, A.D., and J.K. Ord, 1981. *Spatial Processes: Models and Applications*, Pion Limited, London, 266 p.

Curran, P.J., 1983. Multispectral remote sensing for estimate of green leaf area index, *Philosophical Transactions of the Royal Society, Series A*, 309:257–270.

Daughtry, C.S., K.P. Gallo, S.N. Goward, S.D. Prince, and W.P. Kustas, 1992. Spectral estimates of absorbed radiation and phytomass production in corn and soybean canopies, *Remote Sensing of Environment*, 39(2):141–152.

Davis, F.W., and J. Dozier, 1990. Information analysis of a spatial database for ecological land classification, *Photogrammetric Engineering & Remote Sensing*, 56(5):605–613.

Davis, F.W., J. Michaelsen, R. Dubayah, and J. Dozier, 1990. Optimal terrain stratification for integrating ground data from FIFE, *Symposium on FIFE, First ISLSCP Field Experiment*, pp. 11–15.

Dubayah, R.C., 1994. Modeling a solar radiation topoclimatology for the Rio Grande River Basin, *Journal of Vegetation Science*, 5:627–640.

Ferwerda, J.A., and H.E. Young, 1981. The relationship between spruce and fir biomass production and four soil series of a major soil catena in Maine, *Proceedings of the 1981 IUFRO World Congress*, Kyoto, Japan, 26 p.

Grubb, T.G., and R.M. King, 1991. Assessing human disturbance of breeding bald eagles with classification tree models, *Journal of Wildlife Management*, 55(3):500–511.

Harley, J.L., and S.E. Smith, 1983. *Mycorrhizal Symbiosis*, Academic Press, London, England, 483 p.

Hockman, J.N., G.A. Burger, and D.W. Smith, 1990. Classification model to predict Fraser fir christmas tree growth, *Forest Science*, 36(1):45–53.

Jenny, H., 1941. *Factors of Soil Formation — A System of Quantitative Pedology*, McGraw-Hill, New York, 281 p.

Jenson, S., and J. Domingue, 1988. Extracting topographic structure from digital elevation data for geographic information system analysis, *Photogrammetric Engineering & Remote Sensing*, 54(11):1593–1600.

Kirkby, M.J. (editor), 1978. *Hillslope Hydrology*, John Wiley, New York, 389 p.

Klingebiel, A.A., E.H. Horvath, W.U. Reybold, D.G. Moore, E.A. Fosnight, and T.R. Loveland, 1988. [title], Open File Report 88-102, U. S. Geological Survey, Souix Falls, South Dakota, 18 p.

Levine, E., R. Knox, and W. Lawrence, 1994. Relationships between soil properties and vegetation at the Northern Experimental Forest, Howland, Maine, *Remote Sensing of Environment*, 47:231–241.

Lillesand, T.M., and R.W. Kiefer, 1994. *Remote Sensing and Image Interpretation*, John Wiley and Sons, New York, pp. 417–420.

Lozano-Garcia, D., R. Fernandez, and C. Johannsen, 1991. Assessment of regional biomass-soil relationships using vegetation indexes, *IEEE Transactions on Geosciences and Remote Sensing*, 29(2):331–338.

MacDonald, J.D., L.R. Costello, and T. Berger, 1993. An evaluation of soil aeration status around healthy and declining oaks in an urban environment in California, *Journal of Arboriculture*, 19(4):209–219.

Meng, X., and R.S. Seymour, 1992. Influence of soil drainage on early development and biomass production of young, herbicide-released fir-spruce stands in north central Maine, *Canadian Journal of Forest Resources*, 22:955–967.

Michaelsen, J., F. Davis, and M. Borchert, 1987. A non-parametric method for analyzing hierarchical relationships in ecological data, *Coenoses*, 2(1):39–48.

Michaelsen, J., D. Schimel, M. Friedl, F. Davis, and R. Dubayah, 1994. Regression tree analysis of satellite and terrain data to guide vegetation sampling and surveys, *Journal of Vegetation Science*, 5:673:686.

O'Loughlin, E.M., 1986. Prediction of surface saturation zones in natural catchments by topographic analysis, *Water Resources Research*, 22(5):795–804.

Price, J.C., 1993. Estimating leaf area index from satellite data, *IEEE Transactions in Geosciences and Remote Sensing*, 31(3):727–734.

Prince, S.D., 1991. A model of regional primary production for use with coarse-resolution satellite data, *International Journal of Remote Sensing*, 12(6):1313–1330.

Sahrawat, K.L., and D.R. Keeney, 1986. Nitrous oxide emissions

from soil, *Advances in Soil Science* (B.A. Stewart, editor), Springer-Verlag, New York, pp. 103–148.

Soil Survey Staff, 1993. *Soil Survey Manual*, U. S. Department of Agriculture, U. S. Government Printing Office, Washington, D.C.

Townshend, J.G.R., C. Justice, W. Li, C. Gurney, and J. McNamus, 1991. Global land cover classification by remote sensing: present capabilities and future possibilities, *Remote Sensing of Environment*, 35:243–255.
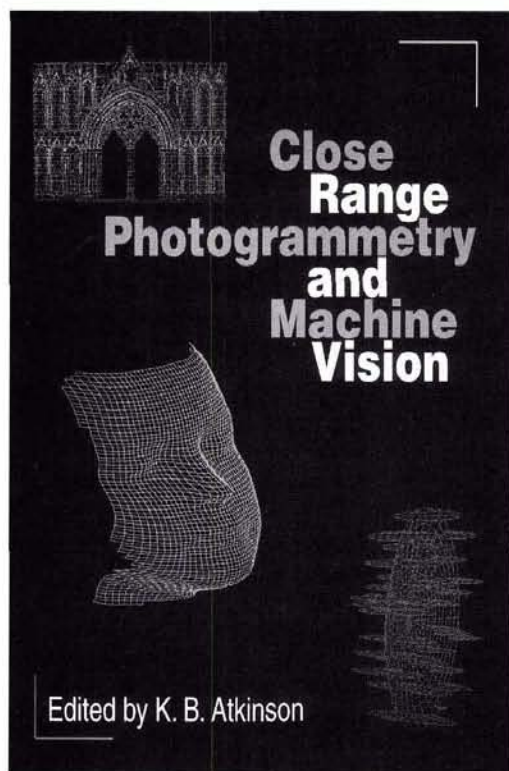
Troeh, F.R., 1964. Landform parameters correlated to soil drainage, *Soil Science Society of America Proceedings*, 28:808–812.

Tucker, C.J., J.G.R. Townshend, and T.E. Goff, 1985. African land-cover classification using satellite data, *Science*, 227:369–375.

Wood, E.F., M. Sivapalan, and K. Beven, 1990. Similarity and scale in catchment storm response, *Reviews of Geophysics*, 28(1):1–18.