Effect of Alternative Splitting Rules on Image Processing Using Classification Tree Analysis

Michael Zambon, Rick Lawrence, Andrew Bunn, and Scott Powell

Abstract

Rule-based classification using classification tree analysis (CTA) is increasingly applied to remotely sensed data. CTA employs splitting rules to construct decision trees using training data as input. Results are then used for image classification. Software implementations of CTA offer different splitting rules and provide practitioners little guidance for their selection. We evaluated classification accuracy from four commonly used splitting rules and three types of imagery. Overall accuracies within data types varied less than 6 percent. Pairwise comparisons of kappa statistics indicated no significant differences (p-value > 0.05). Individual class accuracies, measured by user's and producer's accuracy, however, varied among methods. The entropy and twoing splitting rules most often accounted for the poorest performing classes. Based on analysis of the structure of the rules and the results from our three data sets, when the software provides the option, we recommend the gini and class probability rules for classification of remotely sensed data.

Introduction

The classification of digital imagery to extract useful thematic information is one of the main objectives of environmental remote sensing, and accurate classification is often essential for the effective use of remote sensing products by end users. A variety of classification algorithms are available to analysts for image classification, such as maximum likelihood, minimum distance to means, and parallelepiped (Jensen, 1996). Additionally, more sophisticated methods have been developed, including neural networks, machine learning, and expert classifiers that have helped analysts to markedly improve classification accuracy beyond that obtainable with other methods (Quinlan, 1993; Hansen *et al.*, 1996; Friedl and Brodley, 1997; Lawrence and Labus, 2003).

In this study, we compared multiple decision-tree-based classifications for three image types to assess the impacts of different decision tree development rules on classification results. This article is aimed at practitioners of remote sensing and is not intended as a comprehensive survey of statistical methods associated with decision tree analysis for rule-based classification. For a more theoretical discussion of decision tree methods, the reader is referred to the statistical and machine learning literature dedicated to this topic (e.g., Breiman *et al.*, 1984).

Rule-based classification of remotely sensed imagery using classification tree analysis (CTA, sometimes referred to as classification and regression tree analysis (CART), decision trees, or binary recursive partitioning) has recently received increasing attention. CTA methods have shown promise for improving classification accuracy in several recent studies (Hansen *et al.*, 1996; Friedl and Brodley, 1997; Lawrence and Wright, 2001; Lawrence and Labus, 2003). Decision trees represent nonparametric methods that are especially useful in analyzing large data sets with complex structure (Cappelli *et al.*, 2002). Decision tree analysis, such as CTA, holds several advantages over traditional supervised methods such as maximum likelihood classification (Friedl and Brodley, 1997). It does not depend on assumptions of distributions of the data, can easily handle nonlinear relationships and missing values, and can incorporate categorical ancillary data as well as continuous variables.

Several software packages are available for performing CTA, and the underlying implementations often vary among packages with some software offering a variety of splitting rules. Studies have been published addressing the choice of the CTA algorithm, with the consistent conclusion that selection of the CTA algorithm has little effect on overall accuracy (Breiman *et al.*, 1984; Pal and Mathers, 2003). These studies, however, fail to address variations in individual class accuracies and might lead an analyst to believe that choice of the algorithm is unimportant. We examined four widely available CTA splitting rules applied to three diverse remotely sensed data sets and found that, while overall accuracies might be similar, the effect on individual class accuracies was substantial. The choice of algorithm, therefore, might significantly affect the results of individual classes.

Four splitting rules that are widely available in current software implementations of CTA for growing a decision tree include: gini, class probability, twoing, and entropy. Each of the splitting rules attempts to segregate data using different approaches. The gini index is defined as:

$$Gini\,(t) = \sum_i p_i\,(1 - p_i) \tag{1}$$

where p_i is the relative frequency (determined by dividing the total number of observations of the class by the total

Michael Zambon and Rick Lawrence are with the Department of Land Resources and Environmental Sciences, Montana State University, 128 AJM Johnson Hall, P. O. Box 173490 Bozeman, MT 59717 (mikezambon@msn.com, rickl@montana.edu).

Andrew Bunn is with Woods Hole Research Center and the Department of Land Resources and Environmental Sciences, Montana State University (abunn@whrc.org).

Scott Powell is with the USDA Forest Service, Pacific Northwest Research Station and the Department of Ecology, Montana State University (scott.powell@oregonstate.edu).

Photogrammetric Engineering & Remote Sensing
Vol. 72, No. 1, January 2006, pp. 25–30.

0099-1112/06/7201–0025/$3.00/0

© 2006 American Society for Photogrammetry and Remote Sensing

number of observations) of class i at node t, and node t represents any node (parent or child) at which a given split of the data is performed (Apte and Weiss, 1997). The gini index is a measure of impurity for a given node that is at a maximum when all observations are equally distributed among all classes. In general terms, the gini splitting rule attempts to find the largest homogeneous category within the dataset and isolate it from the remainder of the data. Subsequent nodes are then segregated in the same manner until further divisions are not possible.

An alternative measure of node impurity is the twoing index:

$$Twoing\ (t) = \frac{p_L\, p_R}{4} \left(\sum_i (|p(i|t_L) - p(i|t_R)|)\right)^2 \qquad (2)$$

where L and R refer to the left and right sides of a given split respectively, and $p(i|t)$ is the relative frequency of class i at node t (Breiman, 1996). Twoing attempts to segregate data more evenly than the gini rule, separating whole groups of data and identifying groups that make up 50 percent of the remaining data at each successive node.

Entropy, often referred to as the information rule, is a measure of homogeneity of a node and is defined as:

$$Entropy\ (t) = -\sum_i p_i \log p_i \qquad (3)$$

where p_i is the relative frequency of class i at node t (Apte and Weiss, 1997). The entropy rule attempts to identify splits where as many groups as possible are divided as precisely as possible and forms groups by minimizing the within group diversity (De'ath and Fabricius, 2000). This rule can be interpreted as the expected value of the minimized negative log-likelihood of a given split result and tends to identify rare classes more accurately than the previous rules.

The last rule examined in this study, class probability, is based on the gini Equation 1 and is not a different splitting rule in the same way that the gini rule is different than the towing rule or different than the entropy rule. Although the class probability algorithm is based on the gini equation, the results are focused on the probability structure of the tree rather than the classification structure or prediction success. The rule attempts to segregate the data based on probabilities of response and uses class probability trees to perform class assignment (Venables and Ripley, 1999). Decisions are made based on probabilities of observations belonging to a given class rather than node impurity like the gini rule. A new observation to be categorized is compared against all possible outcomes to calculate the resulting classification probabilities for each leaf of the tree. The observation is then assigned to the class that has the highest probability for the considered observation to belong to this class. It is possible to have a split with both terminal nodes belonging to the same target class but having a substantial difference in predicted class probabilities. Since the gini Equation 1 is always used to grow the class probability tree, resulting trees are generally very similar. Although the class probability algorithm could employ the towing (Equation 2) or entropy (Equation 3) instead of gini (Equation 1), we are not aware of any such commercial implementation (Breiman *et al.*, 1984 provided more information on the theory and implementation of this difference and the other rules described above).

As a general rule, in spite of previous studies, statistical software users are advised that the choice of splitting rule can profoundly affect the structure of a particular classification tree (Breiman *et al.*, 1984). Image classifications developed using different CTA splitting rules, therefore, might be dramatically different. Even if the overall accuracies between two trees grown by different splitting rules are similar, the data structure revealed might be very different. Previous studies applying CTA to remotely sensed image classification, however, have provided no guidance as to which methods, if any, might be superior for such applications.

We evaluated image classification accuracy using various splitting rules available in CTA using a diversity of remotely sensed imagery. Classification accuracy comparisons were made within image types to understand the effects of splitting rules on results for each type of imagery. The imagery selected had been previously used successfully with CTA and represented a diversity of available image options, including Landsat ETM+ imagery with ancillary data, Ikonos imagery with ancillary data, and PROBE-1 hyperspectral imagery.

Methods

We used data from three previous studies in which CTA had been implemented using S-Plus® statistical software, which uses the class probability splitting rule. These data were chosen due to their availability and diversity of data types for comparison of CTA methods. These data types ranged from moderate spatial resolution multispectral imagery to fine spatial resolution hyperspectral imagery (Table 1).

The three data types and their associated study areas included four Landsat 7 ETM+ scenes from the Greater Yellowstone Ecosystem (Lawrence and Wright, 2001), a four-band Ikonos scene and topographic data from Sequoia National Park (Lawrence *et al.*, 2004), and a hyperspectral PROBE-1 scene of the Virginia City, Montana area (Driscoll, 2002). Each of these data sets, collected for purposes associated with previous studies, were used in this study only for relative land cover classification comparisons based on the methods described in this study.

The Landsat 7 ETM+ data, collected 13 July, 15 September, and 04 December 1999, and 29 June 2000 included six

Table 1. Digital Imagery Data Types and Characteristics

Data	Location	Dates	Resolution	Ancillary Data
Ikonos	Sequoia National Park	19 July 2001	4 meter spatial 4-band spectral (450–850 *n*m)	slope
Landsat	Greater Yellowstone Ecosystem	13 July, 15 September, 04 December 1999; and 29 June 2000	30 meter spatial 6-band spectral (450–2350 *n*m)	slope, aspect, elevation, tasseled cap components, tasseled cap difference components
PROBE-1	Virginia City, Montana	25 August 1999	5 meter spatial 128-band spectral (440–2500 *n*m)	None

spectral bands with 30 meter spatial resolution (the thermal band was not included). Also included in the Landsat analysis were tasseled cap components for each of the dates and tasseled cap difference components between the summer and fall dates and between the summer and winter dates. Ikonos data, collected on 19 July 2001, had four spectral bands, including a blue, green, red, and near infrared (NIR) band, and 4 m spatial resolution. The PROBE-1 hyperspectral data acquired on 25 August 1999 consisted of 128 spectral bands, covering the visible through shortwave infrared portions of the spectrum, with a spatial resolution of 5 m.

Topographic data were used with the Landsat and Ikonos imagery. Ancillary data for the area of the Landsat imagery included slope, aspect, and elevation, extracted from a 30 m digital elevation model (DEM) of this area. Ikonos data analysis included a slope layer extracted from a 10 m DEM of the area.

The number of classes and training pixels per class varied for each image type. The Landsat data were classified into six landcover types including conifer, hardwood, herbaceous, mixed conifer/hardwood, mixed conifer/herbaceous, and burned. The PROBE-1 classification also had six classes including conifer, deciduous, developed, range, water, and disturbed land. The Ikonos image was classified into four classes including tree, water, rock, and meadow. Training pixels used for the Ikonos classification were evenly distributed across each cover type and PROBE-1 training data were relatively even across classes, while the training data for the Landsat scene varied greatly between cover types with more rare classes having considerably fewer pixels (Table 2).

We used CART software, version 5.0, which included each of the splitting rules being tested, for our analysis. Classification trees were developed for each data set using each of the splitting rules. By default, these trees were initially grown to over-fit the data and required pruning to develop more robust and parsimonious trees for image classification purposes (Venables and Ripley, 1999; Siciliano and Mola, 2000).

Automated cross-validation methods in CART were used to determine the optimal tree size for each run. This was accomplished by dividing the data into ten groups with a similar distribution of the dependant variable for each group. Maximal tees were constructed using nine of the groups, and the tenth group was used to estimate initial error rates. This process was repeated until each of the ten groups was used as the test sample. The results of the ten tests were then combined into overall error rates and applied to the tree using all the data. This provided reliable estimates of the independent predictive accuracy of each tree for selection of final tree results. The final trees were used as a set of decision rules and incorporated into four classifications for each of the data sets described.

Producer's, user's, and overall accuracy for each of these classifications was assessed using randomly selected, independent reference data and standard error matrix procedures (Congalton and Green, 1999). Kappa statistics were compared pairwise for significant differences using the Delta method for Kappa comparisons from classification of remotely sensed imagery. Variance of the Kappa statistics was computed for each classification and Student's t-tests were performed for all pairs of Kappas within each data type both at a standard p-value of 0.05 and at a more conservative Bonferroni corrected p-value of 0.008. The Bonferroni correction is appropriate for multiple comparisons tests as were conducted between our Kappas. Comparisons of class accuracies and Kappas were made only within each data type and not between data types to assess the relative effect of the various splitting rules for each image classification. In this way, only the splitting rules changed for each comparison, holding all other factors constant within an image type.

Results

The overall accuracy results for the Ikonos data varied only 2 percent from the poorest performing splitting rule, entropy at 82 percent, to the best performing splitting rule, and class probability at 84 percent (Table 3). Overall accuracy for the Landsat data results had a slightly larger range of 4 percent, from 60 percent for the twoing rule to 64 percent for the class probability rule. PROBE-1 overall accuracies had the highest variability, from 78 percent for entropy to 84 percent for the both gini and twoing splitting rules, a range of 6 percent.

Although overall accuracies were consistent, class accuracies varied considerably (Tables 4 and 5). Producer's and user's accuracies for the Landsat data performed poorest in the mixed vegetation categories of conifer/herbaceous (≤49 percent) and conifer/hardwood (≤34 percent), respectively for all of the splitting rules. Also user's accuracies for the Landsat data were generally lower than producer's for most classes.

TABLE 2. REFERENCE DATA PER CLASS USED FOR FINAL CLASSIFICATIONS FOR EACH DATA TYPE

Data Type and Class	Training Pixels	Independent Validation Pixels
Landsat Data		
Conifer	172	88
Conifer/herbaceous	594	308
Burned	85	44
Conifer/hardwood	33	17
Hardwood	58	29
Herbaceous	180	92
Total	1122	578
Ikonos Data		
Tree	1390	49
Water	1390	50
Meadow	1390	53
Rock	1390	106
Total	5560	258
PROBE-1 Data		
Water	264	8
Conifer	376	46
Deciduous	305	16
Developed	354	22
Range	447	35
Disturbed	201	23
Total	1947	150

TABLE 3. CLASSIFICATION RESULTS FOR DATA TYPES AND SPLITTING RULES

Data	Rule	Overall Accuracy (%)	Kappa
Ikonos	Gini	83.3	0.772
	Entropy	82.2	0.756
	Class Prob.	83.7	0.778
	Twoing	83.3	0.768
Landsat	Gini	63.2	0.521
	Entropy	60.6	0.485
	Class Prob.	63.8	0.519
	Twoing	60.4	0.489
PROBE-1	Gini	84.0	0.798
	Entropy	78.0	0.721
	Class Prob.	83.3	0.788
	Twoing	84.0	0.796

TABLE 4. PRODUCER'S PERCENT ACCURACY FOR EACH DATA TYPE AND SPLITTING RULE

Producer's Accuracy	Gini	Entropy	Class Probability	Twoing
Landsat Data				
Conifer	92	94	90	98
Conifer/herbaceous	45	44	49	42
Burned	82	82	68	86
Conifer/hardwood	88	88	88	88
Hardwood	72	66	59	52
Herbaceous	80	67	83	72
Ikonos Data				
Tree	84	78	84	81
Water	94	96	96	94
Meadow	87	85	87	72
Rock	76	76	76	86
PROBE-1 Data				
Water	71	71	71	71
Conifer	96	93	96	96
Deciduous	50	44	50	50
Developed	73	86	95	91
Range	86	74	74	83
Disturbed	96	74	91	87

TABLE 5. USER'S PERCENT ACCURACY FOR EACH DATA TYPE AND SPLITTING RULE

User's Accuracy	Gini	Entropy	Class Probability	Twoing
Landsat Data				
Conifer	60	58	59	57
Conifer/herbaceous	91	85	88	91
Burned	60	44	67	59
Conifer/hardwood	34	34	32	31
Hardwood	58	68	89	56
Herbaceous	49	50	48	45
Ikonos Data				
Tree	82	84	82	71
Water	100	100	100	100
Meadow	61	58	61	73
Rock	94	92	95	88
PROBE-1 Data				
Water	100	100	100	100
Conifer	85	86	85	85
Deciduous	89	88	89	89
Developed	80	58	78	69
Range	94	90	96	91
Disturbed	88	68	70	87

Class accuracies for the Ikonos data also varied considerably, although less so than the Landsat results. The poorest performing classes for producer's and user's were the rock (≤76 percent) and meadow (≤73 percent) classes, respectively. The results for remaining classes were considerably higher for both producer's and user's accuracies.

Class accuracies for the PROBE-1 data were generally high for both producer's and user's accuracy, with the main exception of producer's accuracy for the deciduous class, which ranged from 50 percent to 44 percent. The poorest performing classes for user's accuracy were the developed class (≤80 percent) for most splitting rules and the disturbed class (70 percent) for the class probability rule.

Kappa statistics for the Ikonos classifications covered a similarly narrow range from 0.76 to 0.78, while Kappa statistics for the Landsat data ranged from 0.49 to 0.52. Kappa statistics for PROBE-1 covered the widest range, from 0.72 to 0.80. Results of the pairwise comparisons of the kappa statistics for each data type (Table 6) indicated no statistically significant differences between any of the kappa statistics for all data types at both the 0.05 and Bonferroni corrected alpha levels (all p-values > 0.05).

TABLE 6. P-VALUES FOR PAIRWISE STATISTICAL COMPARISON OF KAPPA RESULTS. NO P-VALUES WERE SIGNIFICANT AT $\alpha = 0.05$ OR THE MULTIPLE COMPARISON BONFERRONI CORRECTED $\alpha = 0.008$

	Entropy	Class Prob.	Twoing
Ikonos Data			
Gini	0.72	0.90	0.93
Entropy	—	0.63	0.79
Class Prob.		—	0.83
Landsat Data			
Gini	0.90	0.32	0.83
Entropy	—	0.39	0.92
Class Prob.		—	0.45
PROBE-1 Data			
Gini	0.17	0.86	0.98
Entropy	—	0.23	0.18
Class Prob.		—	0.88

Discussion

Our results provided evidence that the selection of splitting rules in CTA is not a critical factor for overall accuracy of decision-tree-based classification of remote sensing imagery supporting previously reported findings (Breiman *et al.*, 1984; Pal and Mathers, 2003). These results might be counter intuitive because the different splitting rules partition the data at each split in a different manner. Intuitively, this should produce dichotomous trees with different structures and result in differing classification results. Examination of the dichotomous tree results for the PROBE-1 data set showed that the main branches of these trees were very similar, often using the same variables and values for data separation. The differences in tree structures within a data type exist mainly in the lower branches, further from the root node, where less variation in the data was accounted for and more subtle distinctions between classes were made. Data at these tree levels tend to be observations that are not easily separated into specific classes due to potential overlap of data distributions for given class assignments. Since these differences existed in the lower branches, the effect of these differences on image classifications were not substantial and thereby result in similar classification accuracies and kappa statistics among results from the different splitting rules for the hyperspectral data.

Tree results for the Landsat and Ikonos multispectral data sets showed greater variation in variable selection for the upper branches of their respective trees. Although trees for these data were overall quite similar among the different splitting rules, the twoing rule for Ikonos used slope for the first split compared to the other three splitting rules, which used information from the blue spectral band. The consistency of the overall accuracy results from these different trees indicated that there is more than one pathway in CTA that can produce similar results, especially with multispectral data sets. Also, the class probability rule tended to produce trees with fewer variables and simpler structure than the other rules for these multispectral data sets, which was not the case with the hyperspectral data. The fewer number of spectral bands available in multispectral data sets and the appropriateness of selected ancillary data could account for this simpler tree structure. The greater number of spectral bands in the hyperspectral data provides more information for class separation and can result in more complex tree structures to arrive at specific class separation. Overall though, results from this study showed that, in terms of overall accuracy, CTA methods are reasonably robust to variable selection for each of the data types examined.

Although there was little difference in overall accuracies, the various rules achieve those accuracies in diverse ways though individual class accuracies, a finding not previously reported and a phenomenon that strongly affects the final classification results. An examination of a portion of the Landsat study area near Bozeman, Montana, for example, demonstrated very different classification patterns from the class probability and twoing splitting rules, notwithstanding very similar overall accuracies (Figure 1). This figure demonstrates the variation in results obtained by different splitting rules. In particular, rare classes, such as hardwood and conifer/hardwood, showed tremendous differences between the two images. While the Landsat producer's accuracy for the hardwood class was similar between splitting rules (59 percent for class probability and 52 percent for twoing), the user's accuracy was strikingly different (89 percent for class probability and 56 percent for twoing).

These substantial variations in class accuracies present the remote sensing practitioner with the obvious question of how to select from among the alternative rules. An examination of the relative class accuracies shows that, with few exceptions, the twoing and entropy rules resulted in the lowest accuracies by class, although these results are sometimes balanced by particularly high class accuracies. When results are ranked 1 through 4 in order of performance by class, with the lowest ranking indicating the highest accuracy, twoing and entropy resulted in higher overall average rankings. For both producer's and user's accuracies, gini demonstrated the lowest average rankings (1.69 and 1.56, respectively), followed closely by class probability (1.75 and 1.81, respectively). Entropy and twoing demonstrated substantially higher average rankings for producer's (2.69 and 2.19, respectively) and user's accuracies (2.63 and 2.63, respectively).

Although our sample size for testing the differences in algorithm performance by class was small (three data sets), these results are consistent with the structure of the respective rules. The gini and class probability rules operate similarly by identifying records belonging to the single largest homogeneous class at each split and combining the remainder to be considered at the next split. Although, gini focuses on the classification structure and prediction success, and class probability focuses on the probability structure of the tree, both produce similar results by identifying a single class at each split. Although statistical theory

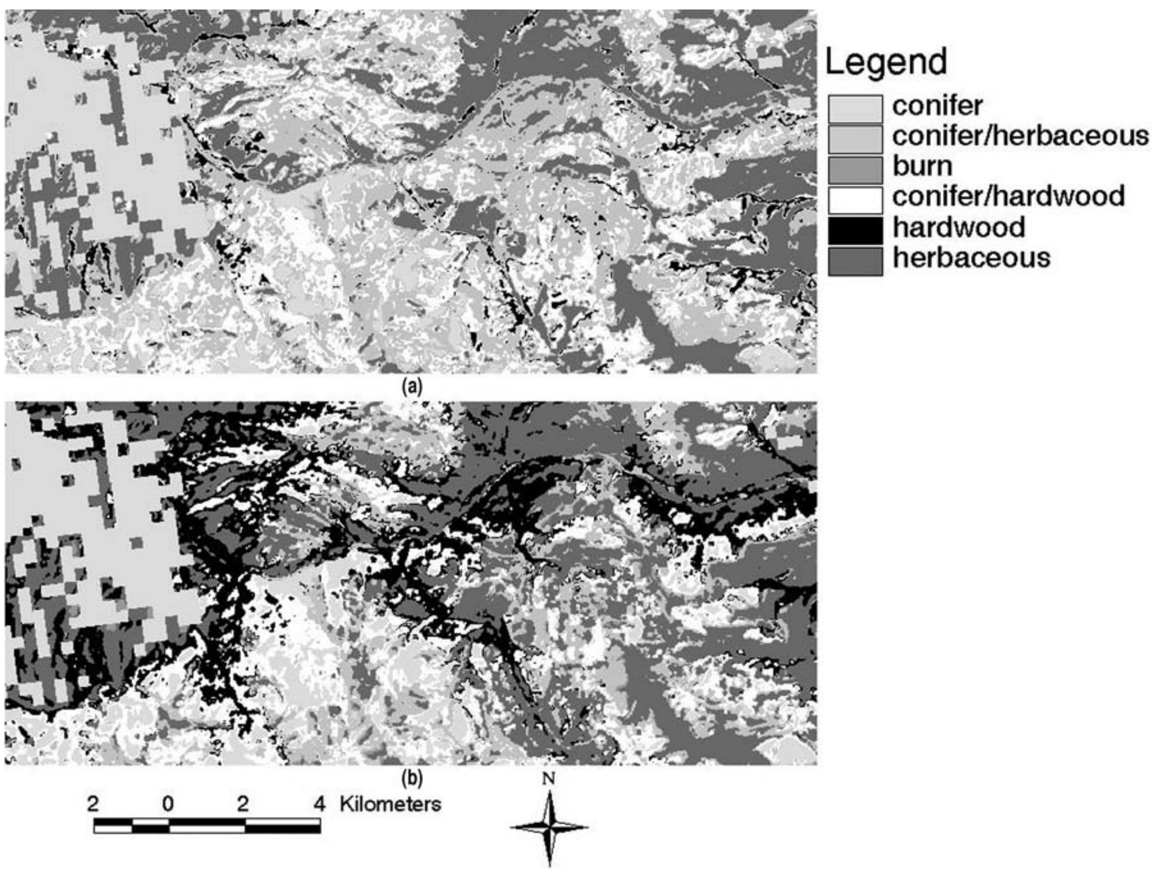

Figure 1. This image shows a portion of the Landsat study area near Bozeman, MT. Image (a) was classified using the class probability splitting rule while image (b) was classified using the twoing splitting rule. This graphic demonstrates the different results obtained by different splitting rules. In particular, rare classes, such as hardwood and conifer/hardwood, show tremendous differences between the two images.

on these rules is scarce, we believe this approach effectively isolates the conditions defining each class. Twoing and entropy, on the other hand, seek to identify the structure of the data by segregating groups that are not necessarily homogeneous. This has increased potential of leaving similar classes grouped (which can be advantageous for other types of statistical analyses where similarities among classes are explored) and likely results in certain classes with especially low accuracies. While twoing and entropy can uncover valuable information about the data structure not evident from gini or class probability, we believe that the approach of these latter rules is more appropriate for classification of remotely sensed data and that, when software provides the option, gini or class probability should be selected for such applications.

The observed variation in class accuracies for these data sets and splitting rules might also be the result of the difficulty in distinguishing certain classes from each other using spectral and selected ancillary data. For example, results of class accuracies for the Landsat data indicated a fundamental difficulty in separating the mixed classes of conifer/herbaceous and conifer/hardwood from each other and from pure classes containing cover components of these mixed classes. Results for the Ikonos image showed that CTA generally performed well with these data but had difficulty separating the meadow and rock classes. At the time of data acquisition in late July, the high meadows in the scene were composed predominantly of sparse, senescent grasses. These meadows appear spectrally similar to the granitic rock outcrops and sands in this region of the Sierra Nevada Mountains, making them difficult to distinguish from each other. The difficulty in separating the deciduous, developed, and disturbed classes for the PROBE-1 data might stem from much of the developed area in Virginia City containing considerable deciduous vegetation. In addition, mining disturbed areas are difficult to distinguish from developed gravel streets, roads, and parking lots, lending to the spectral confusion between these particular classes. The subtle differences in the splitting rule methods, therefore, might cause each rule to handle these class confusions differently, resulting in the variation observed among rules.

The variation of classification accuracies also might be reduced with different tree selection methods resulting from different pruning procedures for these data. Since this study was an effort to compare multiple tree growing procedures through the use of splitting rules, multiple pruning methods between and within data sets might confound the results of the intended analysis. Consistent tree selection and pruning procedures, therefore, were applied to all data sets. Pruning and tree selection are critical to reliable results and should be evaluated separately to determine optimum performance of CTA methods (Pal and Mathers, 2003).

This study provides a degree of comfort to an image analyst using CTA that use of a particular splitting rule should not be a substantial concern for overall accuracy in the classification process in spite of the variability observed in individual class accuracies. The analyst is cautioned, however, that the resulting classifications from different rules are likely to be markedly different. In general, our results indicate that gini or class probability are the more appropriate rules for image classification, all other factors being equal. If particular classes are important to the analyst, however, splitting rule selection might become very important, in which case multiple classifications using several of the splitting options available could be examined to arrive at the best classification for the given study objectives. Finally, other software options available might well outweigh the differences in splitting rules, such as the availability of boosting with the See5 software package (Lawrence *et al.*, 2004).

Acknowledgments
We would like to thank the many Montana State University students involved over several years with the collection of field data that accompany each of the imagery data sets. We would also like to thank the reviewers who provided exceptional insights and thoughtful comments that greatly improved this manuscript from its original form.

References

Apte, C., and S. Weiss, 1997. Data mining with decision tress and decision rules, *Future Generation Computer Systems*, 13:197–210.

Breiman, L., 1996. Some properties of splitting criteria, *Machine Learning*, 24:41–47.

Breiman, L., J.H. Friedman, R.A. Olshen, and C.J. Stone, 1984, *Classification and Regression Trees*, Wadsworth International Group, Belmont, California, 358 p.

Cappelli, C., F. Mola, and R. Siciliano, 2002. A statistical approach to growing a reliable honest tree, *Computational Statistics & Data Analysis*, 38:285–299.

Congalton, R.G., and K. Green, 1999. *Assessing the Accuracy of Remotely Sensed Data: Principles and Practices*, CRC Press, Boca Raton, Florida, 137 p.

De'ath, G., and K. Fabricius, 2000. Classification and regression trees: a powerful yet simple technique for ecological data analysis, *Ecology*, 81(11):3178–3192.

Driscoll, S.G., 2002. *Detecting and Mapping Leafy Spurge (Euphorbia Esula) and Spotted Knapweed (Centaurea Maculosa) in Rangeland Ecosystems Using Airborne Digital Imagery*, Masters Thesis, Montana State University, 111 p.

Friedl, M.A., and C.E., Brodley, 1997. Decision tree classification of land cover from remotely sensed data, *Remote Sensing of Environment*, 61:399–409.

Hansen, M., R. Dubayah, and R. Defries, 1996. Classification trees: an alternative to traditional landcover classifiers, *International Journal of Remote Sensing*, 17:1075–1081.

Jensen, J.R., 1996. *Introductory Digital Image Processing: A Remote Sensing Perspective*, Second Edition, Prentice Hall, Upper Saddle River, New Jersey, 316 p.

Lawrence, R.L., A. Bunn, S. Powell, and M. Zambon, 2004. Classification of remotely sensed imagery using stochastic gradient boosting as a refinement of classification tree analysis, *Remote Sensing of Environment*, 90:331–336.

Lawrence, R.L., and M.A. Labus, 2003. Early detection of Douglas-fir beetle infestation with sub-canopy resolution hyperspectral imagery, *Western Journal of Applied Forestry*, 18:202–206.

Lawrence, R.L., and A. Wright, 2001. Rule-based classification systems using classification and regression tree (CART) analysis, *Photogrammetric Engineering & Remote Sensing*, 67:1137–1142.

Pal, M., and P. Mathers, 2003. An assessment of the effectiveness of decision tree methods for land cover classification, *Remote Sensing of Environment*, 86:554–565.

Quinlan, J.R., 1993. *C4.5: Programs for Machine Learning*, Morgan Kaufmann Publishers, San Mateo, California, 302 p.

Siciliano, R., and F. Mola, 2000. Multivariate data analysis and modeling through classification and regression trees, *Computational Statistics & Data Analysis*, 32:285–301.

Venables, W.N., and B.D. Ripley, 1999. *Modern Applied Statistics with S-Plus*, Third Edition, Springer-Verlag, New York, New York, 501 p.

(Received 20 May 2004; accepted 12 November 2004; revised 18 January 2005)