Stan Aronoff*
Department of Forestry
University of California
Berkeley, CA 94720

An Approach to Optimized Labeling of Image Classes

A labeling algorithm is described which finds the label assignment which will minimize the loss function due to misclassification.

Abstract: *Labeling is approached as a problem of finding the best way to express the classes of one classification system in terms of the classes of a second system. An approach and procedure are described that define a loss function for the misclassifications resulting from a label assignment. This loss function takes into account the variable sample sizes for different classes and the different costs of each type of misclassification. An automated labeling system is then described which finds the label assignment which will minimize the loss function. A user set threshold value is included so that the more ambiguous classes are not assigned labels.*

Introduction

Classification algorithms are commonly used in the analysis of Landsat and other image data to generate image classes. The analyst then examines the correspondence between these image classes and the resource types (here termed resource classes) to be identified and assigns a label to each image class which best describes this correspondence.

Labeling can be approached as a problem of finding the best way to express the classes of one classification system in terms of a second system. In this paper a loss function is proposed that can be used to evaluate the optimality of a label assignment. A labeling algorithm is then described which finds the optimum assignment of labels to image classes by minimizing the classification loss function. The algorithm assumes that each image class must be assigned to a single resource class—the usual case where a map product is to be generated. Where the application is concerned with accurate area estimates and mapping is not required, a procedure to generate mixed class assignments may be preferred. Mixed class assignments are not considered here but can be developed from the methodology presented.

The labeling algorithm is not limited to remote sensing applications. It can be used whenever the objective is to express a set of units in one classification in terms of the classes of a second system. For example, a soils map might be used to derive a trafficability map by first defining the labels as high, moderate, and low trafficability, then acquiring a set of sample points for which both the soil and trafficability class was known, and finding the optimal labeling of each soil class.

Assessing the success of a classification and labeling procedure is essentially the same problem as map accuracy assessment. An approach to map accuracy testing is discussed in the following section which is then used to develop a loss function to optimize the labeling procedure.

Map Accuracy Testing

One method to test whether a map is of acceptable accuracy is to select a sample of map points, check the map classification against ground data, and then make a statement about the true accuracy of the map. Such a statement generally claims some minimum level of accuracy with some high level of confidence, e.g., a minimum of 85 percent accuracy at the 95 percent confidence level. The sampling

* Presently with DIPIX Systems Limited, 120 Colonnade Road, Ottawa, Ontario K2E 7J5, Canada.

0099-1112/84/5006-0719$02.25/0

problem is then one of determining the optimal number (N) of map samples to be compared with ground data, and an allowable number of misclassifications (X) of these samples. After these values are determined, N map samples are selected and their classifications are compared against the "true" classification of the sample point (e.g., ground data). If X or fewer points were misclassified, then the map is accepted as accurate at the specified level of precision.

In any statistical test there is a probability or risk that interpretation of the test results will lead to the wrong conclusion. The probabilities associated with the two types of erroneous conclusions may be termed consumer risk and producer risk.*

Consumer risk is the probability that a map of unacceptable accuracy will pass the accuracy test. It can be calculated from the binomial sampling distribution as follows:

$$\text{CRISK} = \sum_{Y=0}^{X} \frac{N!}{Y!(N-Y)!} Q_L^{(N-Y)} (1 - Q_L)^Y \qquad (1)$$

where CRISK = consumer risk,
Q_L = the minimum accuracy required,
X = number of allowable misclassifications,
N = total number of points sampled, and
Y = number of misclassifications.

The producer risk is the probability that a map of some acceptable accuracy Q_H will be rejected and is calculated as follows:

$$\text{PRISK} = \sum_{Y=X+1}^{N} Q_H^{(N-Y)} (1 - Q_H)^Y \qquad (2)$$

where PRISK = producer risk and
Q_H = a selected high accuracy level.

The selection of values for consumer and producer risks depend on the value of the information and cost of errors in a specific application.

Minimum Accuracy Value

A hypothesis test leads to the conclusion that either the map is sufficiently accurate or it is not, depending on whether it passed or failed the test. However, a map which fails to pass a test for 85 percent accuracy might still be adequate for a user requiring only 80 percent accuracy. The minimum accuracy value is a measure of quality obtained by calculating the accuracy of the map (Q_L) for which the observed number of misclassifications would be the allowable number of misclassifications (X), given the specified consumer risk. (Producer risk is already specified by the sample size selected.) The minimum accuracy value can be interpreted to be that accuracy level which is so low that there is only a small chance (the consumer risk) that the test results could be as good as those observed. It is calculated from Equation 1 by setting X to the observed number of misclassifications, N to the sample size used, and CRISK to the consumer risk used in designing the accuracy test. The author developed a computer program to iteratively find a value of Q_L which satisfies the equation.

The normal approximation to the binomial with continuity correction was used for minimum accuracy value calculations in this paper. Where sample sizes are small and the minimum accuracy values are high (0.9), the difference from using a direct calculation may be significant. However, the sample sizes used here are large enough (over 100) and the minimum accuracies close enough to 0.5 (0.4 to 0.7) that this difference is negligible. The minimum accuracy values were calculated from the following equation (adapted from Snedecor and Cochran (1967)):

$$\text{CRISK} = Pr\left[Z > \frac{\frac{N-Y}{N} - Q_L - \frac{1}{2N}}{\sqrt{\frac{(1-Q_L)Q_L}{N-1}}} \right] \qquad (3)$$

where Pr is the "probability of" and other terms are as defined above.

The minimum map accuracy value is the highest map accuracy value for which the observed test results would constitute passing the map accuracy test for a selected level of consumer risk. The minimum accuracy value is a probabilistic estimate of the minimum expected accuracy of the map.

The reason for calculating this conservative accuracy value is that it is a measure which takes into account the degree of certainty of the estimate. Using a consumer risk of 0.05, a test result of 90 percent correct for a sample size of 100 has a minimum accuracy of 83.6 percent whereas, for a sample size of 10, the minimum accuracy value is 60.5 percent. The minimum accuracy value, by reflecting the level of uncertainty related to sample size, is a useful index for comparing accuracy test results in which the sample sizes are different.

Development of the Loss Function

The minimum accuracy value is an index of quality that incorporates the degree of uncertainty and the parameters of the accuracy assessment, i.e., the consumer risk and the sample size. If the consumer risk selected is 5 percent, the minimum accuracy value is interpreted to be the minimum expected accuracy of the map in 95 percent of the cases. The remaining 5 percent of the time is not

* (Aronoff (1982) showed that the consumer and producer risks could become either the Type I or Type II statistical error, depending on the way the null hypothesis is contructed. By using the terms consumer and producer risks, the confusion is avoided.)

TABLE 1. DATA MATRIX

Resource Classes	Image Classes 1	2	3	4	5	6	7	8	9	10	11	12	13	Sum
Pine	40	593	158	93	745	89	0	5	18	298	2	197	2	2240
W. Fir	3318	750	1795	187	130	696	34	66	100	21	28	186	182	7493
R. Fir	153	48	130	0	5	6	2	1	0	1	1	4	13	364
D. Fir	1900	1681	2159	204	354	658	24	28	61	54	0	349	224	7696
Brush	4	2	0	7	1	10	30	28	10	0	4	7	58	161
Sum	5415	3074	4242	491	1235	1459	90	128	189	374	35	743	479	17954

"accounted for" by the minimum accuracy value in the sense that there will be a 5 percent chance that a map could have a true accuracy lower than the minimum accuracy value. In constructing a loss function, the producer risk is not of interest because no alternative hypothesis is tested. Selecting a small value for the consumer risk gives a minimum accuracy that takes into account most of the uncertainty in the sample estimate of the map's accuracy. As the consumer risk is reduced, there is an increase in the difference between the minimum accuracy value and the sample proportion correct.

If the minimum accuracy value is treated as the minimum expected accuracy for a given resource class, then the complement (1.0 − minimum accuracy) can be interpreted as the maximum expected error for a class. This value can be used to assess the relative cost of assigning an image class (e.g., as produced by an unsupervised classification of Landsat data) to a resource class.

Consider an unsupervised classification of Landsat data to be used for mapping a resource such as forest cover. Visual inspection is commonly used to label each image class. Test sites are selected, the image classs is determined from the Landsat classification results, and the corresponding resource class is determined from verification data (e.g., field inspection). Then the image class is assigned a label that best describes its correspondence with the verification data.

The data matrix in Table 1 is from an unsupervised classification of Landsat-3 data for the mixed conifer forest type in a portion of the Plumas National Forest, California. The table shows the correspondence between 13 image classes and five resource classes, namely PINE (ponderosa pine), W. FIR (white fir), R. FIR (red fir), D. FIR (Douglas fir), and BRUSH (areas of brush). It was produced by selecting a number of test areas, identifying their resource class designation from ground data, finding the test areas on the Landsat image, and then counting the number of pixels in each resource class that was assigned to each image class. Each column of the matrix shows the number of pixels in an image class that corresponds to each resource class.

Labeling is the process of deciding which image classes will be used to represent (i.e., be assigned to) a given resource class. Table 2 shows the results of a subjective assignment of image classes to resource classes which was done by an interpreter comparing the unsupervised Landsat classification to ground data. The image class data were then combined according to the assignments of Table 2 by adding the respective columns from the data matrix (Table 1). The resulting evaluation matrix is shown in Table 3. For example, the first column in Table 3 was the result of adding corresponding elements of columns 5 and 10 from Table 1. So, Table 3 represents the success of predicting the resource classes using the Landsat classification label assignments shown in Table 3. Those image classes that the analyst considered to be poorly related to all resource classes were left unassigned and are shown assigned to the "OUT" class in Table 2.

TABLE 2. IMAGE CLASS ASSIGNMENTS

Resource Class	Image Classes				
Pine	5	10			
W. Fir	1	6			
R. Fir					
D. Fir	2	3	4	12	13
Brush					
Out	7	8	9	11	

TABLE 3. EVALUATION MATRIX

		Predicted Pine	W. Fir	R. Fir	D. Fir	Brush	Out	Sum
Verified	Pine	1043	129	0	1043	0	25	2240
	W. Fir	151	4014	0	3100	0	228	7493
	R. Fir	6	159	0	195	0	4	364
	D. Fir	408	2558	0	4617	0	113	7696
	Brush	1	14	0	74	0	72	161

The evaluation matrix is in effect a contingency table that shows how many pixels were correctly classified and the number of each type of misclassification. Each row of the evaluation matrix can be treated as an accuracy test for the resource class with the row sum as the sample size and the diagonal element as the number of correct assignments. Minimum accuracy values can be calculated for each row. A maximum expected loss value can be calculated as one minus the minimum accuracy. In addition, a weight representing the cost of a misclassification can be assigned to each row. In the case of a forestry application, the per pixel value of the timber could be used as a weight. The maximum expected loss for each row can then be calculated as follows:

$$ML_i = MM_i \, W_i \, n_i \tag{4}$$

where ML_i = maximum expected loss for resource class i,
MM_i = maximum expected error for resource class i,
n_i = number of pixels of resource class i in the test areas, and
W_i = weight assigned to resource class i.

The expected loss for all resource classes (TL) is the sum of the maximum expected loss values for each resource class:

$$TL = \sum_{i=1}^{m} ML_i \tag{5}$$

where m = number of resource classes, and
TL = total maximum expected loss.

Table 4 illustrates an assessment of the maximum expected loss for the class assignments shown in Table 2, using the evaluation matrix in Table 3. (A consumer risk of 0.001 was found empirically to be effective for image classes on the order of several thousand pixels or less.) The weights have been set to 1, which would be appropriate where, for example, the accurate identification of each resource class is equally important.

From Table 4 it can be seen that the minimum accuracy values are lower than the percent correct values. This difference is larger as the sample size becomes smaller because the level of uncertainty is greater for smaller sample sizes. Compare, for example, the minimum accuracy for the PINE class, 0.432 with 46.4 percent correct (i.e., 0.466), a difference of 0.028, and the minimum accuracy for D. FIR, 0.582 with 60 percent correct, a difference of 0.018.

More Complex Weighting

In some applications it may be important to minimize specific misclassifications, or, if the necessary information is available, it may be desirable to distinguish the losses incurred by each type of misclassification. Using the previous forestry example, the cost of each misclassification could be the absolute difference between the value of the timber as predicted by the classification and the value as given by the verification data.

Calculation of the maximum expected loss for a class differs from the simpler case described above in that the expected loss must first be partitioned among the misclassified categories. Consider the PINE class in Table 8. From the accuracy assessment it can be seen that 1043 pixels out of 2240 were correctly classified. The minimum accuracy for this class is 0.433 or 43.3 percent, somewhat less than the percentage correct. The maximum expected error for the class will be 1 − 0.433 = 0.567 or 56.7 percent. Using the simple weighting scheme, this loss value was multiplied by the row sum to give an expected number of misclassified pixels, and then multiplied by the weight to give a maximum expected loss (see Equation 4).

However, when weights are assigned to each type of misclassification, the maximum expected error must be further broken down, i.e., partitioned, to obtain the portion of the error to be associated with each weight. This is done by distributing the maximum expected error (0.567) among the types of misclassifications (shown in a row of the evaluation matrix) in the same proportion as observed in the

Table 4. Accuracy Assessment Table Using Image Class Assignments of Table 3 and Weights Set to 1

Resource Class	Number Correct	Unassigned	Row Sum	% Correct	% Omission	Column Sum	% Commission	Consumer Risk	Minimum Accuracy	Maximum Loss
Pine	1043	25	2240	46.6	53.4	1609	35.2	0.001	0.433	1270.1
W. Fir	4014	228	7493	53.6	46.4	6874	41.6	0.001	0.517	3619.1
R. Fir	0	4	364	0	100.0	0	0	0.001	0	364.0
D. Fir	4617	113	7696	60.0	40.0	9029	48.9	0.001	0.582	3216.9
Brush	0	72	161	0	100.0	0	0	0.001	0	161.0
For Entire Matrix	9674	442	17954	53.9	46.1	17512	44.8			8631.1

TABLE 5. NORTHERN SIERRA STUMPAGE VALUES IN $ PER 1000 BD. FT.

Species		Value
Douglas Fir	(D. Fir)	40
Red Fir	(R. Fir)	20
White Fir	(W. Fir)	20
Ponderosa Pine	(Pine)	60

(Source: Anonymous, 1982)

sample. In this case, the maximum expected error for the resource class PINE is distributed among the four types of errors, namely W. FIR, R. FIR, D. FIR, and BRUSH, in the proportion 154:0:1043:0. In this way the maximum expected error values for each type of misclassification of a resource class will sum to the maximum expected error for the resource class as a whole. The maximum expected loss, partitioned among the different types of misclassifications, is calculated as follows:

$$PL_{i,j} = \left(\frac{A_{i,j}}{\sum_{\substack{i=1 \\ i \neq j}}^{m} A_{i,j}} \right) W_{i,j} MM_i n_i \tag{6}$$

where $PL_{i,j}$ = partitioned maximum expected loss for misclassifying pixels into re source class j that were class i,
$A_{i,j}$ = number of pixels assigned to class j that were class i, and
$W_{i,j}$ = weight for misclassifying pixels into resource class j that were class i.

The maximum expected loss for the evaluation matrix is then given by

$$TL = \sum_{i=1}^{m} \sum_{\substack{j=1 \\ j \neq i}}^{m} PL_{i,j} \tag{7}$$

where $PL_{i,j}$ = maximum expected loss for misclassifying a pixel as resource class j when it was class i,
TL = maximum expected loss for the evaluation matrix, and
m = number of resource classes.

The stumpage values for timber in the Plumas National Forest are given in Table 5. BRUSH is given a timber value of 0. The weighting matrix (Table 6) was calculated by taking the absolute difference between the stumpage value for the predicted tree species and the verified species for each type of misclassification. For example, misclassifying PINE, valued at $60/1000 board feet, as Douglas Fir, valued at $40/1000 board feet, is given a loss value of 20. The same spectral class assignments assessed in Table 4 were re-evaluated using the weighting matrix of Table 6 to give the assessment shown in Table 7.

TABLE 6. WEIGHTING MATRIX

	Pine	W. Fir	R. Fir	D. Fir	Brush	Out
Pine	0	40	40	20	60	60
W. Fir	40	0	0	20	20	20
R. Fir	40	0	0	20	20	20
D. Fir	20	20	20	0	40	40
Brush	60	20	20	40	0	0

OPTIMAL LABEL ASSIGNMENT

Having defined a loss function, an optimal labeling algorithm can be used to find the label assignments that will minimize the loss. The method was used to test each image class separately. By assigning each image class the label (in this case, the resource class name) which minimizes the loss for that image class, the loss for the entire evaluation matrix is also minimized.

The following decision rule was used: For each image class (j), the loss ($L_{k,j}$) produced by assigning each of the (m) labels to the class was calculated; i.e.,

$$L_{k,j} = \sum_{\substack{i=1 \\ i \neq k}}^{m} kd_{i,j} \, W_{i,k} \tag{8}$$

where $L_{k,j}$ = loss if the pixels in image class j were assigned label k,
$kd_{i,j}$ = number of pixels in image class j that were in resource class i,
$W_{i,k}$ = weight for a misclassification of resource class i as resource class k, and
m = total number of resource classes.

Then image class j is assigned the label k such that $L_{k,j}$ is a minimum.

Note that minimum accuracy values are not calculated when a single image class is being considered because the sample size for each test is the same.

In the event that two or more labels give the lowest loss value, then the tie is broken by assigning the label that would give the largest number of correctly identified pixels.

Table 8 shows the results of an optimal label assignment using the data set of the previous examples. In this case, all weights have been set equal to 1. This results in a plurality assignment, i.e., the image class is assigned to the resource class with the largest number of pixels in that image class. This algorithm assigns a label to every image class, so none can be assigned to the OUT class. The maximum expected loss for the matrix (TL) is 8401.1, which is less than the value for Table 4 of 8633.4. This indicates that the automated assignment is "better" (in terms of minimizing the expected loss) than the analyst's assignments evaluated in Table 4. However, this comparison is unfair in that the analyst was, in effect, using a threshold. If an image

TABLE 7. ACCURACY ASSESSMENT TABLE USING ASSIGNMENTS FROM TABLE 3 AND WEIGHTING MATRIX

Resource Class	Number Correct	Unassigned	Row Sum	% Correct	% Omission	Column Sum	% Commission	Consumer Risk	Minimum Accuracy	Maximum Loss
Pine	1043	25	2240	46.6	53.4	1609	35.2	0.001	0.433	29200.2
W. Fir	4014	228	7493	53.6	46.4	6874	41.6	0.001	0.517	75524.0
R. Fir	0	4	364	0	100.0	0	0	0.001	0	4220.0
D. Fir	4617	113	7696	60.0	40.0	9029	48.9	0.001	0.582	66699.8
Brush	0	72	161	0	100.0	0	0	0.001	0	3300.0
For Entire Matrix	9674	442	17954	53.9	46.1	17512	44.8			178944.0

class was judged not to show a high enough level of correspondence to the resource class, it was not assigned. The loss function will register a benefit whenever the number of correctly classified pixels is increased—no matter how small the increase. The problem is that there is a second implicit objective in labeling, namely that the levels must also have some minimum level of "class purity," especially when the labels will be used for map production. The problem of incorporating a threshold into the algorithm is discussed in a separate section below.

Table 9 is an optimal assignment of the data set using the weighting matrix of Table 6. Note that the image class assignments are slightly different; the labels for classes 6 and 13 have been changed. Image class 6 has changed from an assignment of W. FIR to D. FIR. Because the weight for D. FIR is twice that of W. FIR and the number of pixels belonging to each resource class is about the same (see Table 1), the change appears to be appropriate. Image class 13 has changed from D. FIR to W. FIR, a change that does not seem appropriate because the plurality assignment for class 13 is D. FIR, and that resource class is also higher valued than W. FIR.

The calculation of the loss values ($L_{k,j}$) are shown

TABLE 8. OPTIMIZED ASSIGNMENT WITH WEIGHTS SET TO 1

Evaluation Matrix

Verified \ Predicted	Pine	W. Fir	R. Fir	D. Fir	Brush	Out	Sum
Pine	1043	154	0	1043	0	0	2240
W. Fir	151	4242	0	3100	0	0	7493
R. Fir	6	163	0	195	0	0	364
D. Fir	408	2671	0	4617	0	0	7696
Brush	1	86	0	74	0	0	161

Image Class Assignments

Resource Class	Image Classes					
Pine	5	10				
W. Fir	1	6	7	8	9	11
R. Fir						
D. Fir	2	3	4	12	13	
Brush						
Out						

Accuracy Assessment Table

Resource Class	Number Correct	Unassigned	Row Sum	% Correct	% Omission	Column Sum	% Commission	Consumer Risk	Minimum Accuracy	Maximum Loss
Pine	1043	0	2240	46.6	53.4	1609	35.2	0.001	0.433	1270.1
W. Fir	4242	0	7493	56.6	43.4	7316	42.0	0.001	0.548	3386.8
R. Fir	0	0	364	0	100.0	0	0	0.001	0	364.0
D. Fir	4617	0	7696	60.0	40.0	9029	48.9	0.001	0.582	3216.9
Brush	0	0	161	0	100.0	0	0	0.001	0	161.0
For Entire Matrix	9902	0	17954	55.2	44.8	17954	44.8			8398.8

TABLE 9. OPTIMIZED ASSIGNMENT USING WEIGHTING MATRIX

Evaluation Matrix

Verified	Predicted: Pine	W. Fir	R. Fir	D. Fir	Brush	Out	Sum
Pine	1043	67	0	1130	0	0	2240
W. Fir	151	3728	0	3614	0	0	7493
R. Fir	6	170	0	188	0	0	364
D. Fir	408	2237	0	5051	0	0	7696
Brush	1	134	0	26	0	0	161

Image Class Assignments

Resource Class	Image Classes					
Pine	5	10				
W. Fir	1	7	8	9	11	13
R. Fir						
D. Fir	2	3	4	6	12	
Brush						
Out						

Accuracy Assessment Table

Resource Class	Number Correct	Unassigned	Row Sum	% Correct	% Omission	Column Sum	% Commission	Consumer Risk	Minimum Accuracy	Maximum Loss
Pine	1043	0	2240	46.6	53.4	1609	35.2	0.001	0.433	26823.4
W. Fir	3728	0	7493	49.8	50.2	6336	41.2	0.001	0.479	81208.5
R. Fir	0	0	364	0	100.0	0	0	0.001	0	4000.0
D. Fir	5051	0	7696	65.6	34.4	10009	49.5	0.001	0.639	55565.1
Brush	0	0	161	0	100.0	0	0	0.001	0	3780.0
For Entire Matrix	9822	0	17954	54.7	45.3	17954	45.3			171377.0

in Table 10. It can be seen that the large loss for the misclassified D. FIR pixels is offset by the lower losses for BRUSH and R. FIR. The weighting matrix used here represents the difference between the actual and predicted stumpage values. The label assignment has, therefore, optimized the following relation: If image class 13 must be assigned a single label, which label will give a predicted value closest to the actual total value of those pixels as defined by the weights? From the data matrix (Table 1) and the stumpage values used to generate the weights (Table 5), the actual value of the pixels for image class 13 is calculated as $(2 \times 60) + (182 \times 20) + (13 \times 20) + (224 \times 40) + (58 \times 0) = 12980$. If all 479 pixels in class 13 were considered W. FIR, the predicted value would be $479 \times 20 = 9580$, an underestimate of 3400. Considering the class to be D. FIR would give an estimate of $479 \times 40 = 19160$, an overestimate of 6180. So the assignment algorithm did select the best label, given that only one could be selected.

A further refinement of the labeling procedure, in the case where a map is not required, would be to allow the splitting of image classes. In this way the pixels in an image class could be apportioned among a number of resource classes.

A THRESHOLD OF AMBIGUITY

As noted in the previous section, the loss function will register a benefit whenever the number of correctly classified pixels is increased. As a result, the loss value will decrease by assigning any label to an image class instead of listing it as an OUT class, as long as there is some increase in the number of correctly identified pixels. Even if the pixels of an image class were evenly divided among the resource classes, i.e., complete ambiguity, assigning a label would still decrease the loss value.

It is often desirable to flag those image classes which are "too" ambiguous and should either be

TABLE 10. COMPARISON OF LOSS VALUES FROM LABELLING IMAGE CLASS 13 AS WHITE FIR AND DOUGLAS FIR

	As White Fir	As Douglas Fir
Pine	$40 \times 20 = 80$	$2 \times 20 = 40$
W. Fir		$182 \times 20 = 3640$
R. Fir	$13 \times 0 = 0$	$13 \times 20 = 260$
D. Fir	$224 \times 20 = 4480$	
Brush	$58 \times 20 = 1160$	$58 \times 40 = 2320$
	L = 5720	L = 6260

TABLE 11. MARGINAL BENEFITS FOR OPTIMIZED ASSIGNMENT WITH WEIGHTS SET TO ONE

Image Class	Marginal Benefit
1	0.605
2	0.546
3	0.506
4	0.423
5	0.582
6	0.477
7	0.416
8	0.527
9	0.555
10	0.779
11	0.856
12	0.477
13	0.466

TABLE 12. MARGINAL BENEFITS FOR OPTIMIZED LABEL ASSIGNMENT USING WEIGHTING MATRIX

Image Class	Marginal Benefit
1	20.05
2	30.43
3	22.25
4	24.66
5	40.06
6	20.79
7	5.99
8	11.74
9	18.03
10	50.67
11	16.65
12	30.25
13	15.53

labeled by an analyst or not assigned a label at all. This requires that the ambiguity in a class be measured. The marginal benefit to include a class can be used as a measure of ambiguity: i.e.,

$$MB_j = \frac{TL - TLout_j}{n_j} \tag{9}$$

where MB_j = marginal benefit to include image class j as assigned by the algorithm,
TL = maximum expected loss for a given automated label assignment,
$TLout_j$ = maximum expected loss for a given automated label assignment when image class j is assigned to the OUT class, and
n_j = number of pixels in image class j.

Table 11 shows the marginal benefits for the optimal assignment shown in Table 8 which used weights all set to 1. Referring to the data matrix (Table 1), it can be seen that the more ambiguous image classes 4, 7, and 13 have relatively low marginal benefits whereas those image classes strongly associated with a single resource class, such as classes 10 and 11, have high values. Table 12 shows the marginal benefits when the weight matrix of Table 6 is used. Note that the relative order of image classes from highest to lowest marginal benefits has changed to reflect the different costs of misclassifications, as expressed by the weighting matrix.

Using the marginal benefit, a threshold can be set so that any class giving a benefit less than the threshold is assigned to the OUT class. For convenience, the threshold was made proportional to the average per pixel loss of the automated assignment as follows:

$$\mathrm{TEST}_j = MB_j - \mathrm{THRESH} - \frac{TL}{N} \tag{10}$$

where TEST_j = test value for image class j,
THRESH = user selected threshold value,
TL = maximum expected loss value, and
N = total number of pixels over all image classes.

Then image class j is assigned to the OUT class if TEST_j is less than or equal to zero.

Table 13 shows the results of an optimized assignment using a threshold of 7. Image classes 7, 8, and 13 were not assigned labels. The threshold value provides an explicit, consistent method to identify classes that are ambiguous, as defined by the user's selection of a threshold value and a weighting matrix. In this way, the relative importance of distinguishing between resource classes is taken into account.

CONCLUSION

In this paper the problem of assigning labels to image classes was addressed. A labeling procedure was proposed that takes into account the relative certainty of accurately identifying a pixel in a given image class and the cost of each type of misclassification. In this way, the relative suitability of label assignments for a specific application can be measured. An optimal assignment algorithm was developed for this labeling procedure by finding the label assignments that would give the lowest expected loss value. This algorithm assigned a label to every image class regardless of its degree of ambiguity. However, the user could select a threshold value so that classes exhibiting a level of ambiguity exceeding the threshold were not labeled. The loss value calculated for the entire evaluation matrix (TL) can be used to compare the relative success of different labeling assignments on the same or different

TABLE 13. OPTIMIZED ASSIGNMENT USING WEIGHTING MATRIX AND THRESHOLD

Evaluation Matrix

		Predicted						
		Pine	W. Fir	R. Fir	D. Fir	Brush	Out	Sum
Verified	Pine	1043	60	0	1130	0	7	2240
	W. Fir	151	3446	0	3614	0	282	7493
	R. Fir	6	154	0	188	0	16	364
	D. Fir	408	1961	0	5051	0	276	7696
	Brush	1	18	0	26	0	116	161

Image Class Assignments per Pixel Threshold to Include = 7.0

Resource Class	Image Classes				
Pine	5	10			
W. Fir	1	9	11		
R. Fir					
D. Fir	2	3	4	6	12
Brush					
Out	7	8	13		

Accuracy Assessment Table

Resource Class	Number Correct	Unassigned	Row Sum	% Correct	% Omission	Column Sum	% Commission	Consumer Risk	Minimum Accuracy	Maximum Loss
Pine	1043	7	2240	46.6	53.4	1609	35.2	0.001	0.433	26972.0
W. Fir	3446	282	7493	46.0	54.0	5639	38.9	0.001	0.442	86742.0
R. Fir	0	16	364	0	100.0	0	0	0.001	0	4320.0
D. Fir	5051	276	7696	65.6	34.4	10009	49.5	0.001	0.639	61363.2
Brush	0	116	161	0	100.0	0	0	0.001	0	1460.0
For Entire Matrix	9540	697	17954	53.1	46.9	17257	44.7			180857.2

data sets and the relative success of different classification algorithms.

The labeling algorithm is not limited to remote sensing applications. It can be used whenever the objective is to express a set of units from one classification in terms of the classes of a second system. This application arises whenever a map is "re-interpreted" to provide a different analysis, e.g., generating a trafficability map from a soils map or a wildlife habitat map from a vegetation map. The algorithm presented here is able to optimize label assignments while incorporating the uncertainty inherent in drawing a sample, the costs of each type of misclassification, and the need for an assignment to have some minimum value as a predictor of the class identity of a pixel, i.e., attain some minimum level of "class purity."

ACKNOWLEDGMENTS

The author gratefully acknowledges the critical reviews and support of Dr. Randall Thomas and Mr. Andrew Benson of the Remote Sensing Research Program, Space Sciences Laboratory, and the assistance provided by Dr. Allan Stangenberger of the Department of Forestry, University of California, Berkeley. Preparation of this paper was partially funded by the Space Sciences Laboratory, University of California, Berkeley.

REFERENCES

Anonymous, 1982. *November 23, 1982 Harvest Value Schedules*. State of California Board of Equalization. Sacramento, California.

Aronoff, S., 1982. Classification Accuracy: A User Approach. *Photogrm. Engin. and Remote Sensing* 48(8):1229-1307.

Cochran, W. G., 1977. *Sampling Techniques*, third edition. John Wiley and Sons. New York, N.Y.

Snedcor, G. W., and W. G. Cochran, 1967. *Statistical Methods*, 6th edition. Iowa State University Press. Ames, Iowa.

(Received 21 July 1983; revised and accepted 5 February 1984)