An Explicit Index for Assessing the Accuracy of Cover-Class Areas

Guofan Shao, Wenchun We, Gang Wu, Xinhua Zhou, and Jianguo Wu

Abstract
We present a new index, called Relative Errors of Area (REA), for assessing the accuracy of cover-class areal percentage (%LAND) that is extracted from thematic maps after classifying remotely sensed data. We demonstrate how to derive REA from an error matrix and its relationship with user’s and producer’s accuracy. We compare the REA index with other accuracy indices in a hypothetical and two real case studies. The accuracy of cover-class areal estimates is highly correlated with the REA index, but not with other classification accuracy indices such as the overall classification accuracy. In general, users should be aware of using thematic maps with low REA values. Moreover, the estimates of cover-class area can be revised by using REA if cell values of the major diagonal in an error matrix are available.

Introduction
Cover-class areal percentage, or %LAND, commonly derived from thematic maps that are produced by classifying remotely sensed data, is broadly used to measure natural resources, evaluate ecological conditions, and quantify landscape change in space and time (O’Neill et al., 1988; McGarigal and Marks, 1994; Frohn, 1998). Unfortunately, all thematic maps contain classification errors. Thus, statistical analyses with %LAND inevitably contain errors. Furthermore, the uncertainties of %LAND are not readily estimable with the existing classification accuracy measures (Hess, 1994; Shao et al., 2001; Wu and Shao, 2002). Without knowing the uncertainties of landscape metrics, it is difficult to compare different landscapes, to detect landscape change over time, or to relate detected landscape pattern to ecological processes (Hess, 1994). Readers could be even wrongly impressed by the values of landscape metrics if an explicit accuracy estimate is unknown. Unless landscape indices can be explicitly assessed, quantitative analysis with landscape indices is still a controversial issue.

Studies have addressed the issues relevant to the problem of the accuracy or uncertainty in using landscape indices. The studied issues include correlations among landscape indices (Riitters et al., 1995; Cain et al., 1997), scale factors of landscape mapping with remote sensing data (Woodcock and Strahler, 1987; Moody and Woodcock, 1995), scale effects on landscape analysis (Wickham and Ritters, 1995; Jelinski and Wu, 1996; Qi and Wu, 1996; Wu et al., 1997; Fortin, 1999), sensitivity of landscape indices to land-cover misclassification and composition (Moody and Woodcock, 1995; Wickham et al., 1997; Shao et al., 2001), uncertainties of diversity indices derived from error matrices (Hess and Bay, 1997), and error propagation with environmental modeling (Jelinski and Wu, 1996; Heuvelink, 1998). However, little success has been achieved regarding the effect of classification errors on the accuracy of landscape indices.

With the advent of more advanced digital satellite remote sensing techniques, the necessity of performing accuracy assessments of image classification has received renewed interest (Congalton, 1991). Classification accuracy can be assessed with various measures or statistics derived from an error matrix (Congalton and Green, 1999). These statistics include overall accuracy, producer’s accuracy, user’s accuracy, Kappa statistic (including weighted Kappa statistic), and classification success indices (Koukoulas and Blackburn, 2001). On the one hand, each accuracy measure is appropriate for certain, but not for all, applications. On the other hand, the overall accuracy or Kappa statistic is more commonly used in practice than user’s and producer’s accuracy. In this paper, we compare these accuracy statistics with a new index, called Relative Errors of Area (REA), in estimating the uncertainty of %LAND, which is defined as the total area of a cover class divided by the total landscape area multiplied by 100 (McGurigal and Marks, 1994; http://flash.lakeheadu.ca/~remptlp/patch/; last accessed 13 January 2003).

An Illustration of Classification Errors with a Contrived Data Set
In a hypothetical situation (Figure 1), two classifications have the same overall accuracy of 90 percent. For the first classification, the producer’s accuracy is 79 percent for class 1 and 100 percent for class 2, and user’s accuracy is 100 percent for class 1 and 83 percent for class 2; for the second classification, the producer’s accuracies of class 1 and 2 are 100 percent and 80 percent, respectively, and user’s accuracy is 83 percent for class 1 and 100 percent for class 2. Although the overall accuracy is relatively high or satisfactory in practice, which is higher than the minimum level of 85 percent defined by the U.S. Geological Survey (Anderson et al., 1976), the area estimations based on the two classifications are obviously different. While the...
value of %LAND is about 50 percent for class 1 or 2 in the reference map, %LAND ranges from about 40 percent to 60 percent for each class in both classifications. The errors of %LAND are ±10 for each class in both classifications. The relative difference in area estimations between the two classifications is (29–19)/24 = 41 percent for class 1 and is (30–20)/25 = 40 percent for class 2.

This example has practical implications for areal estimations or determining changes in %LAND over time or across space. Unless the overall classification accuracy is 100 percent, which is impossible in reality, %LAND estimates based on remote sensing data may have sizable errors and, in turn, result in misinterpretations in further statistical or quantitative analysis.

The Derivation of Relative Errors of Area (REA)

To deal with the problem of classification errors associated with thematic maps based primarily on remote sensing data, we propose a new index that can be used to assess the accuracy of landscape area. This index provides information on the uncertainty in derived cover-class area or %LAND estimates.

If a thematic map contains $n$ classes, its accuracy can be assessed with the following error matrix (Congalton, 1991; Congalton and Green, 1999):

$$
\begin{array}{cccc}
\text{Reference} & f_{11} & f_{12} & \ldots & f_{1n} & f_{1+} \\
\text{Classification} & \vdots & \vdots & \ddots & \vdots & \vdots \\
\text{Column Total} & f_{+1} & f_{+2} & \ldots & f_{+n} & N \\
\end{array}
$$

where $N$ is the total number of sampling points, $f_{ij}$ ($i$ and $j = 1, 2, ..., n$) is the joint frequency of observations assigned to class $j$ by classification and to class $i$ by reference data, $f_{i+}$ is the total frequency of class $i$ as derived from the classification, and $f_{+j}$ is the total frequency of class $j$ as derived from the reference data.

For a given cover-class $k$ ($1 \leq k \leq n$), the reference value of %LAND ($LR_k$) is computed as

$$LR_k = \frac{\sum_{i=1}^{n} f_{ik}}{N} = \frac{\sum_{i=1}^{n} f_{ik} + f_{kk}}{N}. \quad (2)$$

The classification value of %LAND ($LC_k$) is derived as

$$LC_k = \frac{\sum_{i=1}^{n} f_{ik}}{N} = \frac{\sum_{i=1}^{n} f_{ik} + f_{kk}}{N}. \quad (3)$$

Thus, the difference between $LC_k$ and $LR_k$ is

$$LC_k - LR_k = \frac{\sum_{i=1}^{n} f_{ik} - \sum_{j=1}^{n} f_{jk} + \sum_{i=1}^{n} f_{ik} - \sum_{i=1}^{n} f_{ik}}{N} = \frac{\sum_{i=1}^{n} f_{ik} - \sum_{i=1}^{n} f_{ik}}{N}. \quad (4)$$

If $LC_k - LR_k = 0$, there are two possibilities: classification errors are zero, or commission errors ($CE_k$) and omission errors ($OE_k$) are the same for cover-class $k$. The first possibility is normally untrue in reality. In many situations, the second possibility is also untrue. If $CE_k > OE_k$ and $LC_k - LR_k > 0$, the value of %LAND of class $k$ is underestimated. Therefore, the components of $CE_k$ and $OE_k$ in Equation 4 determine the accuracy of %LAND for cover-class $k$.

Mathematically, $CE_k$ is given as

$$CE_k = \sum_{j=1}^{n} f_{kj} \quad (5)$$

and $OE_k$ is expressed as

$$OE_k = \sum_{i=1}^{n} f_{ik}. \quad (6)$$

The balance between $CE_k$ and $OE_k$ indicates the absolute errors of area estimate for cover-class $k$. The relative errors of area (REA) are then defined as

$$REA_k = \frac{\sum_{i=1}^{n} f_{ik} - \sum_{j=1}^{n} f_{jk}}{f_{kk}} \times 100 \quad (7)$$

where $f_{kk}$ is an element of the $k$th row and $k$th column in an error matrix. It represents the frequency of sample points that are correctly classified.

According to Congalton and Green (1999), User’s Accuracy of class $k$ ($UA_k$) can be expressed as

$$UA_k = \frac{f_{kk}}{\sum_{j=1}^{n} f_{kj}} = \frac{f_{kk}}{f_{kk} + \sum_{j=1}^{n} f_{kj}} \quad (8)$$

and Producer’s Accuracy of class $k$ ($PA_k$) can be expressed as

$$PA_k = \frac{f_{kk}}{\sum_{i=1}^{n} f_{ik}} = \frac{f_{kk}}{f_{kk} + \sum_{i=1}^{n} f_{ik}}. \quad (9)$$
By substituting Equations 8 and 9 into Equation 7, it is easily derived that

\[
REA_k = \left( \frac{1}{UA_k} - \frac{1}{PA_k} \right) \times 100.
\]  

(10)

Thus, \( REA \) can be obtained using information on the error matrix with Equation 7 or the user and producer’s accuracy with Equation 10. If \( REA_k = 0 \), \( %LAND \) has no errors; \( REA_k > 0 \), \( %LAND \) is underestimated; if \( REA_k < 0 \), \( %LAND \) is overestimated. In the hypothetical example above, the first classification underestimates class 1 (\( REA_1 = -26.6\% \)) and overestimates class 2 (\( REA_2 = 20.5\% \)), and the second classification overestimates class 1 (\( REA_1 = 20.5\% \)) and underestimates class 2 (\( REA_2 = -25.0\% \)).

To calibrate areal estimates with \( REA \), we introduce a constant \( K_k \) for class k as follows:

\[
%LAND_{k,c} = %LAND_k + K_k \times REA_k
\]  

(11)

where \( %LAND_{k,c} \) is calibrated \( %LAND_k \). Assuming that \( LR_k \) (see Equation 2) is the unbiased estimate of \( %LAND_k \), \( K_k \) can be derived as

\[
K_k = \frac{-\sum_{i=1}^{n} f_{ik} + \sum_{j=1}^{n} f_{jk}}{N} = \frac{K_{kk}}{N}.
\]  

(12)

\( K_k \) can be computed as long as cell values of the major diagonal in an error matrix are available. In the hypothetical example above, \( K_1 = -19/49 = -0.39 \) and \( K_2 = -25/49 = -0.51 \) for classification 1 and \( K_1 = -24/49 = -0.49 \) and \( K_2 = -25/49 = -0.41 \) for classification 2. Both classifications result in the same values of calibrated \( %LAND_k \).

Case Studies

Case 1: Mapping a Two-Class Landscape

Two sub-scenes of Landsat Thematic Mapper (TM) data, path 116 and row 31 (128° and 42° N), were acquired on 12 September 1987 and 04 September 1993. The study site is covered with typical old-growth, broadleaved-coniferous forest and clear-cut areas. Two sub-scenes of Landsat Thematic Mapper (TM) data, path 22 and row 32, acquired on 05 October 1992, was used to map land-use and land-cover classes of an area of Tippecanoe County, Indiana. Based on the classification system of the U.S. Geological Survey (Anderson et al., 1976), four Level-I land-use and land-cover classes of urban, agricultural land, forest, and water were defined from the image (Shao et al., 2001).

Color infrared aerial photographs acquired in 1993 at a scale of 1:10,000 for the same area were scanned and rectified to the UTM projection coordinates with a 1-m resolution. The aerial images were used as background images while displaying the TM image. The comparison between the aerial and TM images on the computer helped assign names to training areas and reference samples. The classification was carried out by 23 students (18 graduates and five undergraduates) from a remote sensing class. The students could choose any of the following techniques: supervised classification, unsupervised classification, original TM bands, PCA (Principal Components Analysis) transformation, or Tasseled Cap transformation. Following a 3- by 3-pixel majority filtering to classified images, the accuracy of the 23 four-class thematic maps was evaluated. To assure that all the maps share the same standard of classification accuracy, the instructor provided all the students with a single set of 250 reference samples, which was determined two fields and each field was sized around 50 pixels. Three classifiers — Minimum Distance (MD) that takes only the first-order statistics, the class mean, Quadratic Likelihood (QL) that includes both the first- and second-order statistics, and Extraction and Classification of Homogeneous Objects (ECHO) — were used for supervised classifications. MD and QL are referred to as spectral classifiers because they consider only spectral information, whereas ECHO is a spectral-spatial classifier or sample-based classifier (Kettig and Landgrebe, 1976). After initial classifications, spectral classes were grouped into two information classes: forest and clear-cut. The classification experiment resulted in 18 thematic maps (2 data × 3 training samples × 3 classifiers). A manually digitized thematic map from 1:20,000-scale infrared color aerial photographs covering an area of 7 by 12 km was used as reference data for accuracy assessment. The reference map was re-rectified and re-sampled to match the resolution of the Landsat TM data used for this study. A set of 1,000 simple-random samples (pixels) was located on the reference map, and the values for all the sample locations were used to build error matrix tables.

The overall accuracy for the 18 thematic maps ranged from 82.6 to 93.2 percent whereas their KHAT (Congalton and Green, 1999) values ranged from 0.63 to 0.86 (Figure 2). The range of producer’s accuracy is between 82 and 98 percent for forest and 64 and 95 percent for clear-cut; the range of user’s accuracy is between 79 and 96 percent for forest and 80 and 96 percent for clear-cut. The %LAND is estimated to be between 49.7 and 70.4 percent for forest and 29.6 and 50.3 percent for clear-cut.

\( REA \) ranges from -17 to +23 percent for forest and from -46 to +20 percent for clear-cut. No simple relationships are found between %LAND and overall accuracy. The estimations of %LAND are related to producer’s accuracy but almost perfect relationships are found between \( REA \) and %LAND of forest (Figure 3). At the 95 percent confidence level, the slope of the regression is 0.52 between \( REA (x) \) and %LAND (y), which is about the same as the mean of K values (\( KB = 0.53 \)) of class forest among the 18 maps. When Equation 11 is used to perform calibrations, the range of %LAND of forest is reduced from 49.7 to 70.4 percent to 55.5 to 59.1 percent (Figure 3).
with the stratified random sampling technique provided by Erdas Imagine software (http://gis.leica-geosystems.com/Products; last accessed 13 January 2003). The sample unit was a pixel. The student could repeat classifications with various classification techniques until they reached a satisfactory overall accuracy. This resembled closely the common practice in image classification.

Each map was different in terms of classification accuracy and area estimation (Table 1). The overall accuracy of the 23 thematic maps ranged from 78 to 89 percent. The range of producer’s accuracy is 32 to 92 percent for urban, 73 to 97 percent for agriculture, 47 to 92 percent for forest, and 78 to 100 percent for water; the range of user’s accuracy is 62 to 96 percent for urban, 72 to 93 percent for agriculture, 62 to 100 percent for forest, and 84 to 100 percent for water. The %LAND is estimated to be 5.3 to 24.0 percent for urban, 49.7 to 78.3 percent for agriculture, 13.1 to 31.5 percent for forest, and 0.8 to 1.4 percent for water.

The REA ranges from −200 to +52 percent for urban, from −30 to +37 percent for agriculture, from −100 percent to +43 percent for forest, and from −23 to +18 percent for water. The closest relationship is found between REA and %LAND of each land-use and land-cover class among six classification accuracy indices (Table 2). When Equation 11 is used to perform calibrations, the range of %LAND is reduced to 11.2 to 17.4 percent for urban, 61.5 to...
ships with %LAND but not the overall classification accuracy. This explains why REA has close relationships with the average of user’s and producer’s accuracy for individual cover classes. This promotes a change from explicit because it can be directly computed based on user’s and producer’s accuracy of individual cover classes.

The implications of REA are obvious different from those of the overall classification accuracy because the former is related to the difference between user’s and producer’s accuracy values and the latter is approximately the average of user’s and producer’s accuracy for individual cover classes. This explains why REA has close relationships with %LAND but not the overall classification accuracy. However, the overall accuracy is still useful for revealing the potential uncertainty of %LAND. Figure 4a indicates that the variation in estimated %LAND decreases if overall accuracy increases. Only if overall accuracy is 100 percent, will %LAND have no variations at all and, therefore, is perfectly accurate. If thematic maps have relatively low overall classification accuracy, the differences between user’s and producer’s accuracy could be high. In this case, REA is particularly meaningful. If user’s and producer’s accuracy is unavailable, maps with higher values of the overall classification accuracy are still potentially preferable to those with lower values of the overall classification accuracy. This is because a small difference in overall accuracy may result in a big difference in user’s and producer’s accuracy, which, in turn, results in a big difference in cover-class areal estimates.

The relationship between REA and %LAND theoretically follows a linear function. In reality, R^2 is less than 1. This is because the source, an error matrix, from which REA is derived, contains sampling errors (Congalton, 1988). In this paper, Case 1 has 1,000 simple-random samples for a two-class map, whereas Case 2 has 250 points for a four-class map. The larger sample size results in stronger relationships between REA and %LAND than does the smaller sample size. To assure unbiased representatives of an error matrix, sample points should be located with the simple random sampling scheme. If stratified sampling methods are used, the producer’s accuracy should be computed by weighting the cell proportions by the proportion of each land cover on the map (Stehman and Czaplewski, 1998). In this case, the constant K in Equation 11 should be computed by using the weighted value of cell f_{k}. Anderson et al. (1976) proposed the minimum level of classification accuracy was at least 85 percent. Based on what we have learned from the hypothetical and real examples, 85 percent of overall accuracy is not high enough for assuring accurate estimations of cover-class area. At the continental or global scale, the practical level of classification accuracy is even lower than 85 percent (Scepan, 1999; Vogelmann et al., 2001). The effects of the low classification accuracy on downstream applications are still unknown. In many situations, users have no choice but to simply believe that classification accuracy measures, particularly overall accuracy, provide sufficient needed information. Such interpretation is dangerous because it may help produce misleading results or conclusions when a thematic map’s low classification accuracy is used. One of advantages of REA is that it combines the user’s and producer’s accuracy into one index and, therefore, is readily interpretable. In this case, classification accuracy can be translated into an explicit measure of map quality for map users. If an error matrix is available, REA can be used to revise areal estimates of individual cover classes. This promotes a change from “merely referring to” to “actually using” the information of classification accuracy.

**Acknowledgments**

The authors would like to thank their home institutions for their support for conducting this study. Remote sensing data and ground data were provided by the Changbai Mountain Forest Ecosystem Research Station at the Chinese Academy of Sciences and the Laboratory for Application of Remote Sensing (LARS) at Purdue University. The thematic maps for the second case study were provided by 23 students in FNR 558 of 1999. Anonymous reviewers provided useful comments for revising the manuscript. The final stage of ground checking was supported by the “100 people” project from the Chinese Academy of Sciences.

**Table 1. The Minimum, Maximum, and Mean Values of Classification Accuracy and %LAND of Individual Classes**

<table>
<thead>
<tr>
<th>Accuracy Index</th>
<th>Urban</th>
<th>Agriculture</th>
<th>Forest</th>
<th>Water</th>
</tr>
</thead>
<tbody>
<tr>
<td>OA</td>
<td>77.6</td>
<td>32.0</td>
<td>72.8</td>
<td>48</td>
</tr>
<tr>
<td>Max.</td>
<td>92.9</td>
<td>90.0</td>
<td>97.6</td>
<td>92</td>
</tr>
<tr>
<td>Mean</td>
<td>83.3</td>
<td>70.1</td>
<td>89.3</td>
<td>74.2</td>
</tr>
</tbody>
</table>

**Table 2. A Summary of R^2 Values Following a Linear Regression Analysis between Accuracy Indices and %LAND of Individual Classes**

<table>
<thead>
<tr>
<th>Accuracy Index</th>
<th>Urban</th>
<th>Agriculture</th>
<th>Forest</th>
<th>Water</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overall Accuracy</td>
<td>0.07</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>Producer’s Accuracy</td>
<td>0.74</td>
<td>0.80</td>
<td>0.68</td>
<td>0.62</td>
</tr>
<tr>
<td>User’s Accuracy</td>
<td>0.74</td>
<td>0.74</td>
<td>0.72</td>
<td>0.39</td>
</tr>
<tr>
<td>Relative Errors of Area (REA)</td>
<td>0.86</td>
<td>0.94</td>
<td>0.87</td>
<td>0.75</td>
</tr>
<tr>
<td>Classification Success Index*</td>
<td>0.03</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>Individual Classification*</td>
<td>0.04</td>
<td>0.02</td>
<td>0.02</td>
<td>0.00</td>
</tr>
<tr>
<td>Success Index</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*From Koukoulas and Blackburn (2001).

65.4 percent for agriculture, 21.2 to 25.6 percent for forest, and 0.9 to 1.3 percent for water.
Figure 4. The relationships of %LAND values (Y axis) of forest with various accuracy indices.

References


(Received 13 November 2001; accepted 24 October 2002; revised 04 November 2002)