Can Error Explain Map Differences Over Time?

Robert Gilmore Pontius Jr and Christopher D Lippitt

ABSTRACT: This paper presents methods to test whether map error can explain the observed differences between two points in time among categories of land cover in maps. Such differences may be due to two reasons: error in the maps and change on the ground. Our methods use matrix algebra: (1) to determine whether error can explain specific types of observed categorical transitions between two maps, (2) to represent visually the differences between the maps that error cannot explain, and (3) to examine how the results are sensitive to possible variation in map error. The methods complement conventional accuracy assessment because they rely on standard confusion matrices that use either a random or a stratified sampling design. We illustrate the methods with maps from 1971 and 1999, which show seven land-cover categories for central Massachusetts. The methods detect four transitions from agriculture, range, forest, and barren in 1971 to built in 1999, which a 15 percent error cannot explain. Sensitivity analysis reveals that if the accuracy of the maps were less than 77 percent, then error could explain virtually all of the observed differences between the maps. The paper discusses the assumptions behind the methods and articulates priorities for future research.

KEYWORDS: Accuracy, hypothesis, change, land, matrix, uncertainty

Introduction

Scientists typically attempt to determine whether historic land change has occurred by overlaying maps that have the same categories from two points in time, and assuming that the differences between the maps are attributable to changes on the ground (Yang and Lo 2002; Singh 1989; Liu and Zhou 2004). This procedure makes sense when the maps are perfectly accurate. However, scientists know that maps are not perfectly accurate and that the amount of error is frequently too large to ignore. This is a particularly important concern because the amount of change on the ground for many investigations is less than 15 percent, while the error in many maps may be as large as 15 percent. While the majority of research on accuracy assessment has focused on the assessment of error in a single map, researchers are becoming increasingly interested in assessing the accuracy of maps that show change over time (Foody 2002; Khorram 1999; Lunetta and Elvidge 1999).

This paper addresses this topic and illustrates it with a case from Massachusetts, USA, where residents are alarmed by the reported rate at which land in the State is being transformed (Breunig 2003). A tour of the region reveals that much of this land transformation has been occurring in the City of Worcester and its nine surrounding towns (Figure 1). Figure 2 shows the difference between maps of 1971 and 1999 for seven categories of land cover. A naïve interpretation of Figure 2 would lead to the conclusion that the observed differences indicate land change at the black patches. However, even if there were no land change between 1971 and 1999, possible errors in either of the maps would result in differences between the two maps. It is not immediately clear which portions of the observed differences in Figure 2 are attributable to map error versus real land change on the ground, because uncertain data form the foundation of both of the contributing maps. This paper establishes a general technique to examine maps of a mutual categorical variable at two points in time in order to attribute the differences to two possible sources: error in the maps and change on the ground.

The most common technique to compare two maps statistically between two points in time uses a crosstabulation matrix. The numbers in regular font in Table 1 constitute a crosstabulation matrix of two maps for the study area in central Massachusetts, with the 1971 categories as the rows and the 1999 categories as the columns, where each entry in the matrix is a percent of the study area. A crosstabulation matrix is sometimes called a transition matrix when it compares two maps

Robert Gilmore Pontius Jr and **Christopher D Lippitt**, Clark University, George Perkins Marsh Institute, School of Geography, Department of International Development, Community, and Environment, 950 Main Street, Worcester MA 01610-1477, USA. Phone: 508-793-7761; Fax: 508-793-8881. E-mail: <rpontius@clarku.edu>; <clippitt@clarku.edu>.

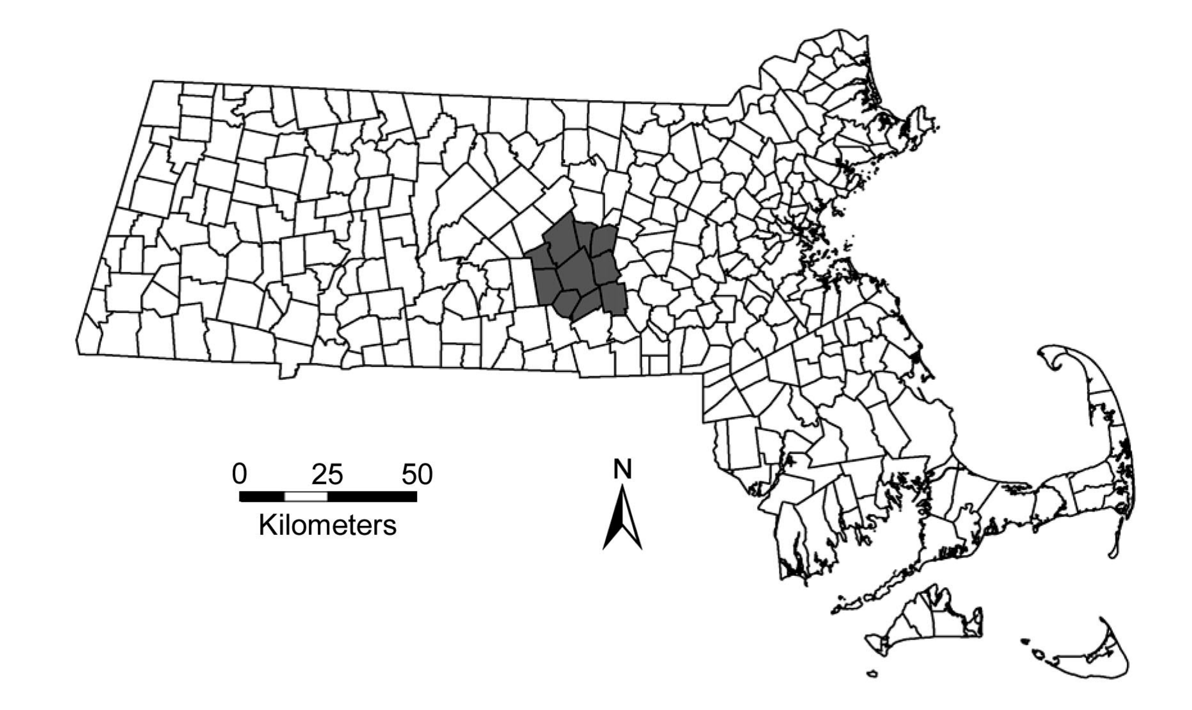

Figure 1. The study area in central Massachusetts consisting of the city of Worcester and its nine surrounding towns.

from different times (Pontius et al. 2004b). All values on the diagonal indicate agreement between the two maps and all values off the diagonal indicate disagreement.

Values on the diagonal are usually attributed to land persistence on the ground. In some cases, land on the diagonal has not persisted on the ground, but it appears on the diagonal due to map error. Values off the diagonal are usually attributed to land transitions on the ground between different categories. In some cases, land off the diagonal has not experienced change on the ground, but it appears off the diagonal due to map error. Error found in the 1971 and/or 1999 maps can induce the transition matrix to indicate more or less change than has actually occurred on the ground.

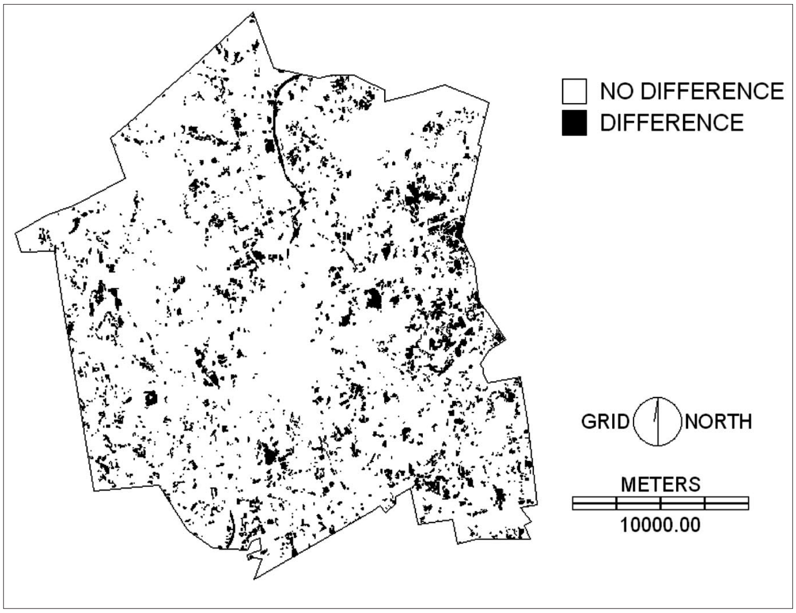

Figure 2. Differences in seven land-cover categories between maps of 1971 and 1999.

This problem is especially troubling when the percent of error in the maps is larger than the percent of change on the ground. It is common for research concerning land transformation to examine landscapes that experience between 5 and 25 percent change between two points in time (Pontius et al. 2004b; Yang and Lo 2002). A standard acceptable overall accuracy for land cover mapping studies has been set between 85 and 90 percent (Anderson et al. 1976; Lins and Kleckner 1996). In practice, land cover classification accuracies can be even lower. This implies

		1999							Total
		Built	Agriculture	Range	Forest	Water	Wetland	Barren	
1971	Built	29 **25** *30*	0 **1** *0*	0 **0** *0*	1 **3** *3*	0 **1** *1*	0 **0** *0*	0 **0** *0*	30 **30** *36*
	Agriculture	1 **1** *1*	3 **2** *2*	0 **0** *0*	0 **1** *1*	0 **0** *0*	— **0** *0*	— **0** *0*	5 **5** *4*
	Range	1 **1** *1*	0 **0** *0*	1 **1** *1*	0 **1** *1*	— **0** *0*	0 **0** *0*	0 **0** *0*	2 **2** *2*
	Forest	7 **4** *4*	0 **1** *1*	0 **1** *1*	49 **48** *43*	0 **1** *1*	0 **1** *1*	0 **0** *0*	56 **56** *51*
	Water	0 **1** *1*	— **0** *0*	— **0** *0*	— **1** *1*	5 **3** *3*	0 **0** *0*	— **0** *0*	5 **5** *5*
	Wetland	0 **0** *0*	— **0** *0*	0 **0** *0*	0 **0** *0*	— **0** *0*	1 **0** *0*	— **0** *0*	1 **1** *1*
	Barren	0 **0** *0*	— **0** *0*	0 **0** *0*	0 **0** *0*	— **0** *0*	0 **0** *0*	0 **0** *0*	0 **0** *0*
	Total	38 **32** *38*	3 **4** *3*	2 **2** *2*	50 **55** *50*	5 **5** *5*	1 **1** *1*	0 **0** *0*	100 **100** *100*

Table 1. Matrix **D** in regular font, matrix **F1** in bold, and matrix **F2** in italics.[1]

that if two different scientists each make a map of the same location for the same time, and they make errors independently, then we could expect substantial difference between the maps (Heuvelink and Burrough 1993).

There are established methods to compare the accuracies of different map makers when they attempt to make maps of the same location for the same time (Congalton et al. 1983; Stehman 1997; Paine and Kiser 2003). If two scientists make a map of the same location for different times, the change-map product derived from the two classifications may exhibit accuracy similar to the accuracy obtained by multiplying the accuracies of each individual classification (Fuller et al. 2003; Mas 1999; Stow 1999). Thus, if two producers were to make a map of different times, each with an accuracy rate of 85 percent, the accuracy rate of the resulting change-map would be approximately 85 percent × 85 percent = 72 percent, leading to a possibly poor estimate of land change if the uncertainty is ignored.

Yang and Lo (2002), for example, report accuracies of between 85 and 90 percent for classifications used in a study of the greater Atlanta metropolitan area between 1973 and 1997/1998. They also report a total transition to urban of approximately 18 percent of the land. It is not clear how to interpret the reported 18 percent disagreement between the two maps because it is not immediately obvious that the maps possess accuracy sufficient to detect 18 percent land change with certainty.

It could be possible to compute the expected accuracy of the map-change product based on simplifying assumptions combined with information from a systematic accuracy assessment. However, many of the available maps have no reported assessment of accuracy, and it is impossible to acquire ground information when at least one of the maps is from the distant past. Nevertheless, scientists should perform some type of analysis concerning accuracy even in situations where precise information does not exist, because knowledge about uncertainty is crucial for proper interpretation. Lack of precise information concerning accuracy does not warrant ignoring this potentially serious issue.

This paper proposes methods to examine the implications of map error for the assessment of land change on the ground and to flag transitions that map error can explain, even in situations where precise information concerning the accuracy of the maps is not available. The paper has three specific goals: (1) to determine whether error can explain specific types of observed categorical transitions between two maps; (2) to represent visually the differences between the maps that error cannot explain, and (3) to examine how sensitive the results are to possible variation in map error.

[1] Each number is expressed as a percent of the study area to the nearest integer, where a zero means a positive number less than one half and a dash means that no pixels were observed in the particular transition. The user's accuracy in both maps is assumed to be 85 percent for each category.

Methods

Data

The Resource Mapping Project (RMP) of the University of Massachusetts at Amherst performed photo interpretation to create land cover maps for Massachusetts in 1971 and 1999. These maps constitute the basis for much of the research on land-use change in Massachusetts, because the maps are freely available from the State's geographic information system website (MassGIS 2002). For the 1971 layer, the RMP aggregated the original 104 classes and eventually digitized 21 categories. The authors of this paper performed further aggregation to seven categories defined by Anderson et al.'s (1976) level 1. Table 1 lists these seven categories in its column and row headers. Category aggregation tends to decrease both the error in individual maps and the difference between maps from two points in time (Pontius and Malizia 2004). Aerial photography is the basis of the 1971 information. The RMP digitized the 1999 data from digital orthophotos. Check plots were examined on the 1971 and 1999 data to compare the final digitized land-cover map to the photographic information.

Statistical information concerning accuracy does not exist in a systematically quantified form. Assessment using ground information has never been performed for the maps, and it is now impossible to obtain ground information from 1971 and 1999. Consequently, we will make an initial assessment of the maps based on the assumption that the accuracy for both maps is 85 percent in order to illustrate the proposed methods. We then perform sensitivity analysis to examine how the results are sensitive to variation in the assumed level of accuracy.

Strategy

Figure 3 illustrates our methodological approach, which tackles the general problem of the comparison of maps from two points in time. For our example, time 1 is 1971 and time 2 is 1999. A systematic accuracy assessment would produce for maps of time 1 and time 2 the matrices **C1** and **C2**, respectively. These matrices are the standard confusion matrices that compare sampled map information in the rows to sampled ground information in the columns (Congalton and Green 1999; Rosenfield and Fitzpatrick-Lins 1986; Stehman and Czaplewski 1998).

For clarity, we refer to map maker X as the one who makes errors specified by **C1** and map maker Y as the one who makes errors specified by **C2**. In situations where matrices **C1** and **C2** are not available, we make an educated guess concerning the map accuracy based on similar studies and discussions with the map makers. For our particular case of land change in central Massachusetts, we assume initially that matrix **C1** gives a user's accuracy of 85 percent for each category at time 1 and that matrix **C2** gives a user's accuracy of 85 percent for each category at time 2.

The user's accuracy for a particular category is the probability that a sampled pixel is classified accurately, given that it is classified as that particular category in the map. Thus each category has its own user's accuracy. If the user's accuracy is 85 percent for a category, then that category has 15 percent commission error. We assume that the 15 percent commission error for each category is distributed evenly among the other six categories, such that the percent of a given category that is confused with a different category is 2.5 percent for each of the other six categories. Eventually, we perform sensitivity analysis to see how the results vary as a function of variation in the assumptions concerning accuracy.

The matrix **D** indicates the observed difference between the maps of time 1 and time 2. The rows of matrix **D** show the categories of the map of time 1 and its columns show the categories of the map of time 2. Our approach to analyze **D** is similar to statistical hypothesis testing, where the null hypothesis is that there is no land change between the two points in time. If the null hypothesis were true, then any positive numbers among the off-diagonal entries of **D** would be attributable to error. If an off-diagonal entry of **D** is larger than what we would expect due solely to error, then we reject the null hypothesis and conclude that there is evidence of land change on the ground for the transition that corresponds to the particular entry. The burden of proof is on the data to show that an off-diagonal entry in **D** is larger than the magnitude that error could explain. Consequently, we need to compute a matrix that shows how **D** would appear if the null hypothesis were true; therefore, we need to consider how the landscape would appear, if the null hypothesis were true.

We consider two cases: 1) the landscape on the ground at both times is the same as the landscape estimated at time 1, and 2) the landscape on the ground at both times is the same as the landscape estimated at time 2. For each of the cases, we compute the matrix that we would expect for

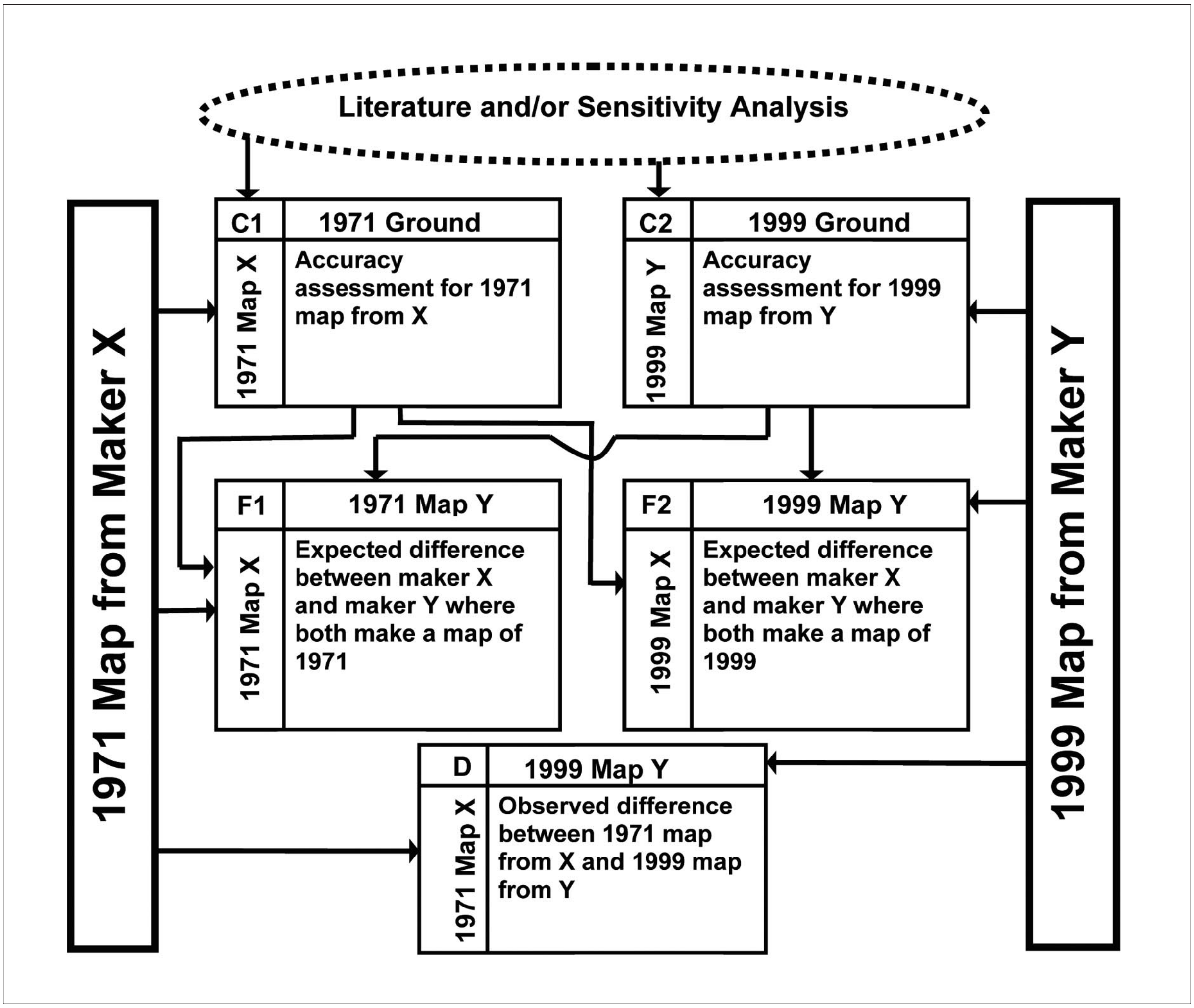

Figure 3. Logic of methods showing flows of information among five matrices: **C1**, **C2**, **F1**, **F2**, and **D**. If matrices **C1** and **C2** are not available, they can be based on values from similar studies and/or analyzed with sensitivity analysis.

matrix **D**, given the types of errors specified in matrices **C1** and **C2**.

For the first case, if there were no errors in the maps, then matrix **D** would be a diagonal matrix, where the column and row totals would be equal to the distribution of categories at time 1 as estimated by the ground information for time 1. A diagonal matrix is a matrix for which all the off-diagonal entries are zero. If there were errors in the maps specified by **C1** and **C2**, then matrix **D** would probably contain some positive off-diagonal entries due solely to error. We compute matrix **F1** such that it is the matrix that we would expect under the assumption that the landscape on the ground at both times is the same as the landscape estimated at time 1, and that matrixes **C1** and **C2** describe the errors in the maps. We compare matrix **D** to **F1** to find whether observed off-diagonal entries in **D** are larger than expected off-diagonal entries in **F1**.

For the second case, we compute matrix **F2** such that it is the matrix that we would expect under the assumption that the landscape at both times is the same as the landscape estimated at time 2 and that matrixes **C1** and **C2** describe the errors in the maps. We compare matrix **D** to **F2** to find whether observed off-diagonal entries in **D** are larger than expected off-diagonal entries in **F2**.

The most important results derive from the comparison of **D** versus **F1** and **D** versus **F2**. If an off-diagonal entry in matrix **D** is greater than the corresponding entry in matrices **F1** or **F2**, then there is evidence to reject the null hypothesis in favor of the conclusion that there is land change on the ground, because map error cannot explain completely the particular type of transition. If an off-diagonal entry in matrix **D** is less than the corresponding entry in matrices **F1** or **F2**, then there is not sufficient evidence to reject the null hypothesis because map error can explain completely the particular type of difference. The next subsection gives the mathematical details to compute **F1** and **F2** as a function of the information in **D**, **C1**, and **C2**.

Matrix Algebra

The methods require raster maps from two points in time, where each pixel is classified as a member of exactly one category of a shared categorical variable. The methods also require *J*-by-*J* confusion matrices **C1** and **C2** that describe the accuracy of the maps from time 1 and time 2 respectively, based on pixels sampled on the ground, where the sample in time 1 may be selected independently from the sample in time 2. With these requirements, let:

- t = time such that t = 1 or t = 2;
- J = number of categories in the analysis;
- i = index for a category in a map;
- j = index for a category on the ground;
- k = index for a category in a map;
- n_{tij} = entry in row i and column j of matrix **Ct** that gives the sampled number of pixels that are classified as category i in the map of time t and observed as category j on the ground at time t;
- n_{ti+} = sampled number of pixels that are classified as category i in the map of time t;
- n_{t+j} = sampled number of pixels observed as category j on the ground at time t;
- n_{t++} = sampled number of pixels for the accuracy assessment at time t;
- N_{ti+} = population number of pixels that are category i in the map of time t;
- N_{t+j} = unknown population number of pixels that are category j on the ground at time t;
- N_{t++} = population number of pixels in the study area at time t where $N_{1++} = N_{2++}$;
- w_{tij} = probability of a pixel being classified as category i in a map given that it is category j on the ground when **Ct** describes the map's errors;
- d_{ik} = entry in row i and column k of matrix **D**;
- f_{tik}= entry in row i and column k of matrix **Ft**; and
- h_{tik} = entry in row i and column k of matrix **Ht**.

The first task is to use the information from the accuracy assessment to estimate the population number of pixels that are category *j* on the ground at time *t*. We must know the method of sampling of pixels for the accuracy assessment in order to interpret the information in **Ct** properly (Congalton 1988b; Stehman 2001). We consider two cases: simple random sampling and stratified sampling. If simple random sampling were used across the study area to select the n_{t++} observations, then N_{t+j} would be estimated by Equation (1).

$$\hat{N}_{t+j} = \left(\frac{n_{t+j}}{n_{t++}}\right) \times N_{t++} \tag{1}$$

If stratified sampling were used, then the estimate of N_{t+j} would be slightly more elaborate. In stratified sampling, we treat each group of pixels classified as category *i* as a stratum in the map of time *t*. The n_{ti+} observations are sampled randomly from among the N_{ti+} pixels in stratum *i*. For this type of sampling, N_{t+j} is estimated by Equation (2).

$$\hat{N}_{t+j} = \sum_{i=1}^{J}\left[\left(\frac{n_{tij}}{n_{ti+}}\right) \times N_{ti+}\right] \tag{2}$$

Next, we would like to compute the probability that a pixel selected randomly from the study area is classified as category *i* by a map maker, given that it is category *j* on the ground. In probability theory notation, this conditional probability is expressed as P(class = *i* | ground = *j*). Equation (3) works from the definition of conditional probability.

$$\mathrm{P}(\text{class} = i \mid \text{ground} = j) = \frac{\mathrm{P}(\text{class} = i \cap \text{ground} = j)}{\mathrm{P}(\text{ground} = j)} = \frac{\mathrm{P}(\text{ground} = j \mid \text{class} = i) \times \mathrm{P}(\text{class} = i)}{\mathrm{P}(\text{ground} = j)} \tag{3}$$

Equation (4) computes the entries of a matrix **Wt** according to the concept of Equation (3) and the entries in confusion matrix **Ct**. For $t = 1$, each w_{1ij} gives the estimated probability of a pixel being classified as category *i* by map maker X, given that it is category *j* on the ground. For $t = 2$, each w_{2ij} gives the estimated probability of a pixel being classified as category *i* by map maker Y, given that it is category *j* on the ground. In particular, w_{tjj} is the producer's accuracy for category j for the map maker whose errors are specified by **Ct**.

$$w_{tij} = \frac{\left(\frac{n_{tij}}{n_{ti+}}\right) \times \left(\frac{N_{ti+}}{N_{t++}}\right)}{\left(\frac{\hat{N}_{t+j}}{N_{t++}}\right)} = \frac{n_{tij} \times N_{ti+}}{n_{ti+} \times \hat{N}_{t+j}} \quad (4)$$

The producer's accuracy for a particular category is the probability that a pixel is classified correctly, given it is that particular category on the ground. Each category can have its own producer's accuracy. It is useful to express matrix **Wt** as J column vectors, denoted as $\mathbf{Vt_j}$ where each column vector has dimension J-by-1 and corresponds to a particular category j. The sum of the entries in each column vector $\mathbf{Vt_j}$ is 1.

If the null hypothesis were true, then any pixel that is a member of category j on the ground at time 1 would also be a member of category j on the ground at time 2. However, this pixel would not necessarily appear on the diagonal of matrix **D** because the classification might be wrong at either or both times. If the pixel is category j on the ground, then w_{1ij} is the probability that X would classify it as category i and w_{2kj} is the probability that Y would classify it as category k. Therefore w_{1ij} multiplied by w_{2kj} is the probability that the randomly selected pixel of category j on the ground is classified as category i by maker X and as category k by maker Y, assuming independence of the classification errors of the two map makers. So, if a pixel truly persists as category j on the ground, then the probability it would appear in row i and column k of the difference matrix **D** is expressed by matrix $\mathbf{E_j}$ in Equation (5).

$$\mathbf{E_j} = \mathbf{V1_j}\mathbf{V2_j}^{\mathrm{T}} \quad (5)$$

The superscript T denotes the transpose of vector $\mathbf{V2_j}$. Therefore, each matrix $\mathbf{E_j}$ is a J-by-J matrix, where $j = 1, \ldots, J$. Matrix $\mathbf{E_j}$ shows the classification according to map maker X in the rows and according to map maker Y in the columns. The entries of matrix $\mathbf{E_j}$ give the conditional probability of a pixel selected randomly from a map being classified as category i by maker X and category k by maker Y, given that it is category j on the ground.

In order to test our null hypothesis for each of the two cases, we need to know the probability that a pixel selected randomly from the map is classified as category i by maker X and category k by maker Y, given each of our two cases, which are: 1) the landscape on the ground at both times is the same as the landscape estimated at time 1; and 2) the landscape on the ground at both times is the same as the landscape estimated at time 2. Equation (6) gives the desired J-by-J matrix for each case t. Equation (6) does this by computing the weighted average of all J matrices ($\mathbf{E_1}, \ldots, \mathbf{E_J}$), where the weight is the estimated proportion of each category on the ground at time t.

$$\mathbf{Ft} = \frac{\sum_{j=1}^{J}\left(\hat{N}_{t+j} \times \mathbf{E_j}\right)}{\sum_{j=1}^{J}\hat{N}_{t+j}} \quad (6)$$

The entry in row i and column k of matrix **Ft** is the probability that a pixel selected randomly from the study area is classified as category i by X and category k by Y, given that both X and Y make a map of the landscape at time t. The only difference between **F1** and **F2** is that Equation (6) uses the estimated category proportions of time 1 to compute **F1** and the estimated category proportions of time 2 to compute **F2**. Table 1 shows the entries for **F1** and **F2**.

Simple subtraction reveals how the off-diagonal entries of **F1** and **F2** compare to the corresponding entries in matrix **D**. If the entry in row i and column k in **D** is greater than the corresponding entry in **Ft**, then there is evidence of a measurable transition from category i in time 1 to category k in time 2. If the entry in row i and column k in **D** is less than the corresponding entry in **Ft**, then error can explain the apparent transition from category i to category k.

For cases where error cannot explain the difference in the map completely, it is useful to compute the remaining proportion of the difference that error cannot explain. Equation (7) computes the proportion of the observed transition from category i to category k that error cannot explain, given the entries in matrices **D** and **Ft**.

$$h_{tik} = \mathrm{MAX}\left[\frac{d_{ik} - f_{tik}}{d_{ik}}, 0\right] \quad for\ i \neq k \quad (7)$$

The entries denoted as h_{tik} form a matrix **Ht** for each case t as shown in Table 2. The goal is to examine the differences between the maps, so h_{tik} is interesting for only non-diagonal entries. If there is no observed difference for the transition from i to k in matrix **D**, then h_{tik} is undefined.Equation (8) defines G_t as the total amount of difference

		1999						
		Built	Agriculture	Range	Forest	Water	Wetland	Barren
1971	Built		**0** *0*	**0** *0*	**0** *0*	**0** *0*	**0** *0*	**0** *0*
	Agriculture	**15** *22*		**0** *0*	**0** *0*	**0** *0*	— —	— —
	Range	**33** *40*	**0** *0*		**0** *0*	— —	**0** *0*	**0** *0*
	Forest	**39** *37*	**0** *0*	**0** *0*		**0** *0*	**0** *0*	**0** *0*
	Water	**0** *0*	— —	— —	— —		**0** *0*	— —
	Wetland	**0** *0*	— —	**0** *0*	**0** *0*	— —		— —
	Barren	**37** *40*	— —	**0** *2*	**0** *0*	— —	**0** *0*	

Table 2. Matrix **H1** in bold and matrix **H2** in italics.[2]

between the maps that error cannot explain. It computes how much larger each off-diagonal entry in matrix **D** is compared to the corresponding entry in matrix **Ft**, then sums those differences and expresses the sum as a proportion of the study area. Equation (8) excludes the diagonal entries of the matrices by not allowing i to equal k.

$$G_t = \sum_{i=1}^{J} \left\{ \sum_{\substack{k=1 \\ i \neq k}}^{J} \mathrm{MAX}\left[\left(d_{ik} - f_{tik} \right), 0 \right] \right\} \tag{8}$$

Visualization

It is desirable to visualize the amount of difference that error cannot explain by modifying Figure 2. The modification reduces the darkness of each black patch in Figure 2 by the degree to which error can explain the particular type of difference. If error can explain completely the observed difference for a particular patch, then the patch becomes white. If error can explain partially the observed difference for a particular patch, then the patch becomes gray, where the darkness of the gray relates directly to the percent of the transition that error cannot explain as given by Table 2. If error cannot explain any of the observed difference for a particular patch, then the patch remains black. Figure 4 shows the implications of Equation (7) when it uses matrix **F1**, and Figure 5 shows the implications when Equation (7) uses matrix **F2**.

Sensitivity to Unknown User's Accuracies

We examine the sensitivity of the results to various assumptions concerning the possible levels of accuracy. Sensitivity analysis compares the results for 31 levels of accuracy between 70 percent and 100 percent, in increments of 1 percent. At each level, the specified accuracy is the assumed user's accuracy for all categories, where the commission error for each category is distributed evenly among the other six categories. Sensitivity analysis provides an approach to determine the accuracy required to detect a specified amount of land change on the ground, given the observed difference in the maps.

Figure 4 shows how map accuracy on the horizontal axis influences the amount of difference in the maps that error cannot explain on the vertical axis. As accuracy increases, each f_{tik} shrinks in Equation (8), so G_t grows. When accuracy is 100 percent, each f_{tik} is zero; thus G_t is the amount of difference observed in the maps as specified by **D**. Figure 6 compares the results based on matrix **F1** to the results based on matrix **F2**.

Results

Table 2 shows evidence of four transitions that error cannot explain when the assumed accuracy of the maps is 85 percent. The transitions are from agriculture, range, forest, and barren in 1971 to built in 1999. We reject the null hypothesis for those four transitions because the entries in both **H1** and **H2** are positive. The zeroes in Table 2 show that error can explain nearly all of the other observed differences between the maps when the assumed error in the maps is 15 percent.

Figures 4 and 5 show the implications visually, at the assumed user's accuracy of 85 percent for

[2] Each number is expressed as a percent of the particular transition that error cannot explain, where a zero means the entry in **D** is less than the corresponding entry in **Ft** and a dash means undefined due to division by zero. The user's accuracy in both maps is assumed to be 85 percent for each category. The diagonal is blank because the method focuses on change, not persistence.

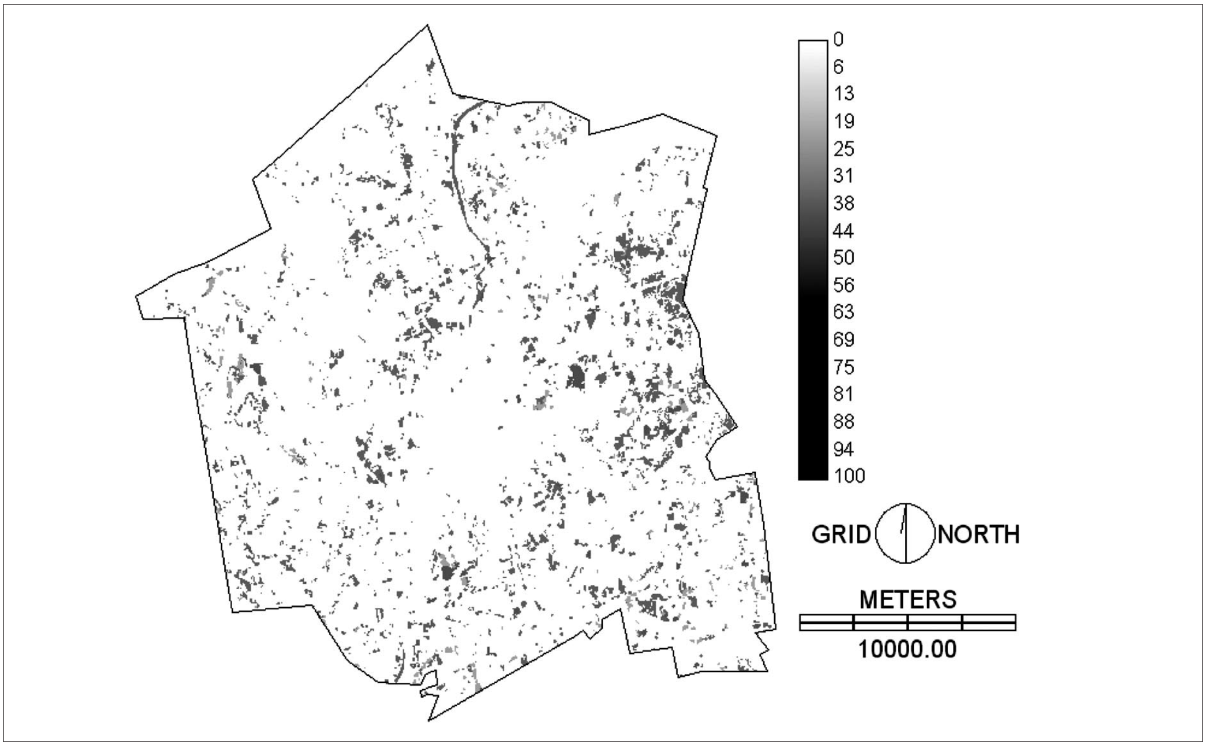

Figure 4. Percent of observed difference that error cannot explain according to matrix **H1**, assuming the user's accuracy in both maps is 85 percent for each category. The darkness of the patch relates directly to the percent of the transition that error cannot explain.

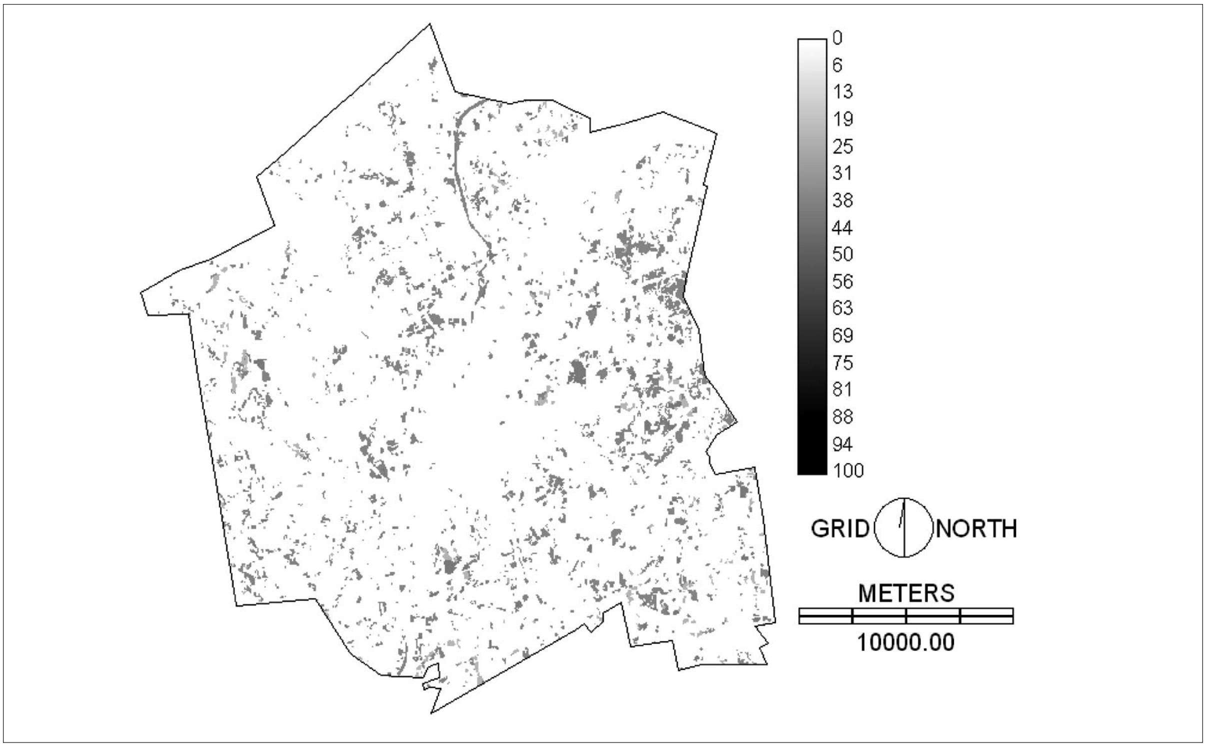

Figure 5. Percent of observed difference that error cannot explain according to matrix **H2**, assuming the user's accuracy in both maps is 85 percent for each category. The darkness of the patch relates directly to the percent of the transition that error cannot explain.

every category. The assumed error can explain more than half of the transitions from agriculture, range, forest, and barren to built, so those transitions are light gray. Error can explain all other observed differences completely according to **H1**, and all others except the transition from barren to range according to **H2**. Consequently, Figures 4 and 5 indicate less land change than Figure 2.

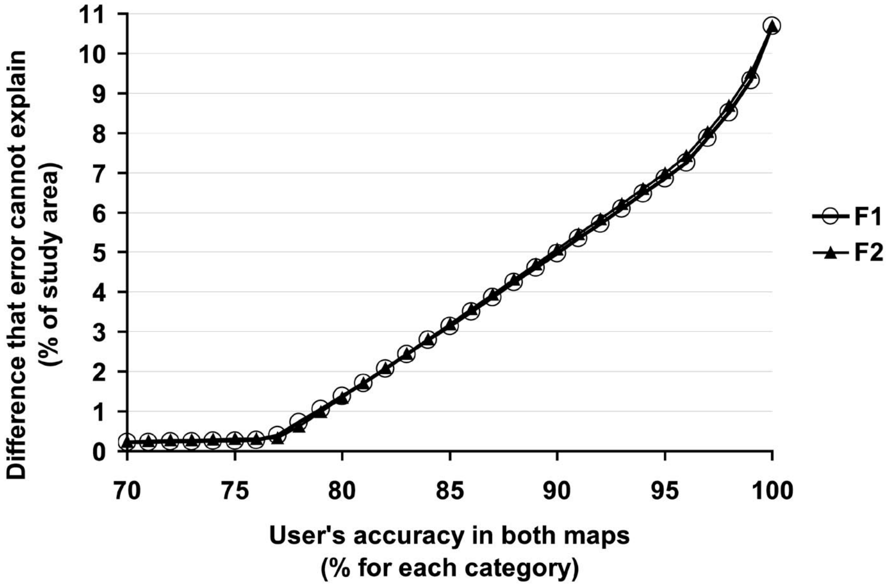

Figure 6. Observed difference between maps of 1971 and 1999 that error cannot explain as a function of map accuracy. Results based on matrix F1 shown by the circles are nearly identical to the results based on matrix **F2** shown by the triangles.

Figure 6 shows the results of the sensitivity analysis. Results based on **F1** are nearly identical to those based on **F2**. When the user's accuracies are below 77 percent, then error can explain nearly all of the observed difference between the maps. When the user's accuracies are 85 percent, then error can explain 8 percentage points of the 11 percentage points of observed difference. When the user's accuracies are 91 percent, then error can explain about half of the observed difference.

Discussion

Assumption Concerning Independence of Errors

The proposed method assumes that the probability of error by maker X is independent of the probability of error by maker Y. The assumption is unlikely to be perfectly correct due to two possible interrelated types of autocorrelation in the error. First, error is likely to exhibit spatial autocorrelation due to the fact that some regions of a study area are usually easier to classify than others, and these regions are frequently clustered spatially (Congalton 1988a). For example, the error in a large homogeneous patch of forest is likely to be relatively small, compared to the error in transition zones where many edges and small patches exist.

Second, the error is likely to exhibit temporal autocorrelation if some locations are systematically easier to classify accurately than others over time. For example, if categories on flat slopes are more accurate than those on steep slopes, then classification errors will likely exhibit temporal autocorrelation because slopes usually persist over time. If the characteristics that cause spatial autocorrelation persist over time, then the spatial autocorrelation will tend to cause temporal autocorrelation. For example, if a large homogenous patch of forest is protected from disturbance over time, then it will be relatively easy to classify accurately over time. For these reasons, it may be reasonable to suspect that two map makers would make errors in similar locations when they attempt to make maps of the same place and time (Steele et al. 1998).

If the error of maker X is correlated positively with the error of maker Y, then our methods' assumption concerning independence of errors would induce matrices **F1** and **F2** to overestimate the proportion of the difference that error could explain. In order to improve the methods, future research should focus on spatial and temporal autocorrelation among the errors (Muchoney and Strahler 2002). Inclusion of either type of autocorrelation would certainly make the calculations of matrices **F1** and **F2** more complicated.

Assumptions Concerning the Confusion Matrices

When interpreting these results, we must remember that the example applies two important assumptions concerning the structure of the confusion matrices: 1) both maps have the same user's accuracy for every category; and 2) each category is likely to be confused with equal probability among the other six categories. We make these assumptions and combine them with sensitivity analysis because accuracy assessment information does not exist for our data. However, if confusion matrices **C1** and **C2** were to exist, then they might indicate that some classes are more likely to be accurate than others. Furthermore, some subsets of categories are more likely to be confused with each other than other subsets (Rogan and Chen 2003). For example, it is usually easier to confuse wetland with forest than wetland with built, because wetland and forest can appear similar in aerial photographs and share similar spectral response patterns. If the necessary information concerning accuracy were available, then the proposed methods would account for variation in the accuracy of individual classes, because the mathematics of our methods can use category-specific information that **C1** and **C2** would contain.

The structure of the confusion matrices also assumes that each pixel belongs entirely to exactly one category. This assumption considers neither the possibility of multiple classes occurring in a single pixel nor the possibility for ambiguity in the definitions of the categories. In order to allow for these possibilities, the confusion matrices could be designed using methods for mixed pixels (Pontius 2002; Pontius and Suedmeyer 2004; Pontius and Cheuk 2006) and/or perhaps according to the logic of fuzzy sets (Woodcock and Gopal 2000; Green and Congalton 2004).

Estimate of Land Change and Persistence

Proper interpretation requires that the reader be cognizant of what the proposed methods can and cannot do. The proposed methods test whether the observed differences between the maps are inconsistent with a null hypothesis of persistence. They are designed to flag particular transitions for additional examination by considering whether error can explain the observed differences between maps. Therefore, the methods can be used to determine whether the accuracy of the maps is sufficiently high to detect with certainty the suspected amount and type of land transition on the ground.

The proposed methods are not designed to estimate the amounts and types of land change on the ground. The methods do not test whether the amounts and types of observed categories in the maps are systematically different than the amounts and types of categories on the ground. This point is especially important when we consider the possibility that the error might cause the maps to show more agreement than exists on the ground. One would need a different approach to test whether error can explain the apparent persistence (Fuller et al. 2003). The interpreter should be careful to avoid the potential trap of accepting the null hypothesis of persistence when error can potentially explain the difference between the two maps.

Given the information concerning the maps and their errors, it is not immediately obvious what should be our best estimate for each category's amount of persistence and for each type of transition on the ground. An important next step in the development of these methods would be to design an estimator for the amount of each type of land transition on the ground, based on matrices **D**, **C1**, and **C2**.

Application to Land Change Modeling

The methods in this paper should be useful to scientists who develop models that predict land change. Typically, land change models begin with an initial map of time 1, produce a prediction map for time 2, and then are assessed with a validation map of time 2 (Pontius et al. 2004a). If a model considers J categories, then there are $J \times J$ possible transitions to consider between two points in time. Some models consider as many as 15 different categories, so such models can become extremely complex and can require substantial computing power in order to consider many small transitions. If error can explain some of the apparent transitions, then it might not be worth for the model to attempt to predict those transitions. Our methods enable the modeler to examine which apparent transitions could potentially be ignored. An important next step for the development of these methods in the context of land change modeling would be to design techniques to compute three pairwise comparisons among three maps: the initial map of time 1, the prediction map of time 2, and the validation map of time 2 (Petrova and Pontius 2005).

Conclusions

This paper proposes methods that are designed to allow a scientist to measure the degree to which error can explain observed differences between maps of land cover at two points in time, based on the confusion matrices for the two maps. The null hypothesis assumes no change on the ground between the two points in time, so if error cannot explain the observed differences, then there is strong evidence to reject the null hypothesis and to conclude that there exists land change on the ground. In the case of central Massachusetts, we observe 11 percent difference between maps from 1971 and 1999 for seven categories, while detailed information concerning map accuracy does not exist. Sensitivity analysis shows that if the maps were 91 percent correct, the error could explain about half of the observed differences. If the maps were 85 percent correct, then error could explain nearly all the observed differences between the maps, except four transitions that indicate gains to the built category. If the maps were less than 77 percent correct, then error could explain virtually all the observed differences between the maps. When interpreting these results, the reader must be cognizant of simplifying assumptions concerning the error structure in the maps. We hope that other scientists will use and extend the methods of this paper to examine how information concerning map accuracy can help to improve the level of sophistication in the measurement of land-use and land-cover change.

ACKNOWLEDGEMENTS

Data and financial support was provided by the Central Massachusetts Human-Environment Regional Observatory (CM-HERO), which is part of the HERO Network, funded by the National Science Foundation through grant # 9978052. We thank anonymous reviewers and the judges for the 2004 competition in the Remote Sensing Specialty Group of the Association of American Geographers, who offered helpful suggestions. Clarklabs facilitated this work through the creation of the GIS software Idrisi®.

REFERENCES

Anderson, J.R., E. Hardy, J. Roach, and R. Witmer. 1976. A land use and land cover classification system for use with remote sensor data. *Professional Paper* 964. United States Geological Survey.

Breunig, K. 2003. Losing ground: At what cost? Lincoln, Massachusetts: Massachusetts Audubon Society.

Congalton, R.G. 1988a. Using spatial autocorrelation analysis to explore the errors in maps generated from remotely sensed data. *Photogrammetric Engineering & Remote Sensing* 54(5):587-92.

Congalton, R.G. 1988b. A comparison of sampling schemes used in generating error matrices for assessing the accuracy of maps generated from remotely sensed data. *Photogrammetric Engineering & Remote Sensing* 54(5):593-600.

Congalton, R.G., and K. Green. 1999. *Assessing the accuracy of remotely sensed data: Principles and practices*. Boca Raton, Florida: Lewis Publishers. 137p.

Congalton, R.G., R.G. Oderwald, and R.A. Mead. 1983. Assessing Landsat classification accuracy using discrete multivariate statistical techniques. *Photogrammetric Engineering & Remote Sensing* 49(12):1671-8.

Foody, G.M. 2002. Status of land cover classification accuracy assessment. *Remote Sensing of Environment* 80(1):185-201.

Fuller, R.M., G.M. Smith, and B.J. Devereux. 2003. The characterization and measurement of land cover change through remote sensing: Problems in operational applications? *International Journal of Applied Earth Observation and Geoinformation* 4: 243-53.

Green, K., and R.G. Congalton. 2004. An error matrix approach to fuzzy accuracy assessment: The NIMA Geocover Project. In: R.S. Lunetta and J.G. Lyon (eds), *Remote sensing and GIS accuracy assessment*. Boca Raton, Florida: Lewis Publishers. pp. 163-72.

Heuvelink, G.B.M., and P.A. Burrough. 1993. Error propagation in cartographic modeling using Boolean logic and continuous classification. *International Journal of Geographical Information Systems* 3(7):231-46.

Khorram, S. (ed.) 1999. Accuracy assessment of remote sensing-derived change detection. *ASPRS Monograph Series*. American Society for Photogrammetry and Remote Sensing (ASPRS), Bethesda, Maryland. 64p.

Lins, K.S., and R.I. Kleckner. 1996. Land cover mapping: An overview and history of the concepts. In: J.M. Scott, T.H. Tear, and F. Davis (eds), Gap analysis: A landscape approach to biodiversity planning. American Society for Photogrammetry and Remote Sensing, Bethesda, Maryland. pp. 57-65.

Liu, H., and Q. Zhou. 2004. Accuracy analysis of remote sensing change detection by rule-based rationality evaluation with post-classification comparison. *International Journal of Remote Sensing* 5(25):1037-50.

Lunetta, R. L., and C. D. Elvidge (eds). 1999. *Remote sensing change detection: Environmental monitoring* *methods and applications*. London, U.K.: Taylor and Francis.

Mas, J.F. 1999. Monitoring land-cover changes: A comparison of change detection techniques. *International Journal of Remote Sensing* 20(1):139-52.

MassGIS (Massachusetts Geographic Information System). 2002. Accessible at www.state.ma.us/mgis/ (accessed 31 January 2002).

Muchoney, D.M., and A.H. Strahler. 2002. Pixel- and site-based calibration and validation methods for evaluating supervised classification of remotely sensed data. *Remote Sensing of Environment* 81(2-3): 290-9.

Paine, D.P., and J.D. Kiser. 2003. *Aerial photography and image interpretation*, 2nd ed. Hoboken, New Jersey: John Wiley & Sons. 632p.

Petrova, S., and R.G. Pontius Jr. 2005. Using imperfect information to validate a model of land change. *Conference Proceedings of the Meeting of the American Society for Photogrammetry and Remote Sensing, Baltimore, Maryland*. 12p.

Pontius Jr, R.G. 2002. Statistical methods to partition effects of quantity and location during comparison of categorical maps at multiple resolutions. *Photogrammetric Engineering & Remote Sensing* 68(10):1041-9.

Pontius Jr, R.G., D. Huffaker, and K. Denman. 2004a. Useful techniques of validation for spatially explicit land-change models. *Ecological Modelling* 179(4): 445-61.

Pontius Jr, R.G., and N.R. Malizia. 2004. Effect of category aggregation on map comparison. Lecture Notes in Computer Science 3234. In: M.J. Egenhofer, C. Freksa, and H. J. Miller (eds), *GIScience2004*. pp. 251-68.

Pontius Jr, R.G., E. Shusas, and M. McEachern. 2004b. Detecting important categorical land changes while accounting for persistence. *Agriculture, Ecosystems & Environment* 101(2-3):251-68.

Pontius Jr, R.G., and B. Suedmeyer. 2004. Components of agreement in categorical maps at multiple resolutions. In: R.S. Lunetta and J.G. Lyon (eds), *Remote sensing and GIS accuracy assessment*. Boca Raton, Florida: Lewis Publishers. pp. 233-51.

Pontius Jr, R.G., and M.L. Cheuk. 2006. A generalized cross-tabulation matrix for comparing soft-classified maps at multiple resolutions. *International Journal of Geographical Information Science* 20(1):1-30.

Rogan, J., and D. Chen. 2003. Remote sensing technology for mapping and monitoring land-cover and land-use change. *Progress in Planning* 61(4):301-25.

Rosenfield, G., and K. Fitzpatrick-Lins. 1986. A coefficient of agreement as a measure of thematic classification accuracy. *Photogrammetric Engineering & Remote Sensing* 52(2):223-7.

Singh, A. 1989. Digital change detection techniques using remotely sensed data. *International Journal of Remote Sensing* 10(6):989-1003.

Steele, B.M., J.C. Winne, and R.L. Redmond. 1998. Estimation and mapping of misclassification probabilities for thematic land cover maps. *Remote Sensing of Environment* 66(2):192-202.

Stehman, S.V. 1997. Selecting and interpreting measures of thematic classification accuracy. *Remote Sensing of Environment* 62(1):77-89.

Stehman, S.V. 2001. Statistical rigor and practical utility in thematic map accuracy assessment. *Photogrammetric Engineering & Remote Sensing* 67(6): 727-34.

Stehman, S.V., and R.L. Czaplewski. 1998. Design and analysis for thematic map accuracy assessment: Fundamental principles. *Remote Sensing of Environment* 64(3):331-44.

Stow, D.A. 1999. Reducing misregistration effects for pixel-level change detection. *International Journal of Remote Sensing* 20(12):2477-83.

Woodcock, C. E., and S. Gopal. 2000. Fuzzy set theory and thematic maps: Accuracy assessment and area estimation. *International Journal of Geographical Information Science* 14:153-72.

Yang, X., and C.P Lo. 2002. Using a time series of satellite imagery to detect land use and land cover changes in the Atlanta, Georgia metropolitan area. *International Journal of Remote Sensing* 23(9):1775-98.