Available online at www.sciencedirect.com
SCIENCE DIRECT

Agriculture, Ecosystems and Environment 101 (2004) 251–268

Agriculture Ecosystems & Environment

www.elsevier.com/locate/agee

Detecting important categorical land changes while accounting for persistence

Robert G. Pontius, Jr.*, Emily Shusas, Menzie McEachern

Department of International Development, Community and Environment, Graduate School of Geography, Clark University, 950 Main Street, Worcester, MA 01610-1477, USA

Abstract

The cross-tabulation matrix is a fundamental starting point in the analysis of land change, but many scientists fail to analyze the matrix according to its various components and thus fail to gain as much insight as possible concerning the potential processes that determine a pattern of land change. This paper examines the cross-tabulation matrix to assess the total change of land categories according to two pairs of components: net change and swap, as well as gross gains and gross losses. Analysis of these components can distinguish between a clearly systematic landscape transition and a seemingly random landscape transition. Multiple resolution analysis provides additional information concerning the distances over which land change occurs. An example of change among four land categories in central Massachusetts illustrates the methods. These methods enable scientists to focus on the strongest signals of systematic landscape transitions, which is necessary ultimately to link pattern to process.

Keywords: Land change; LUCC; Massachusetts; Matrix; Model; Pattern; Process; Scale; Transition

1. Introduction

1.1. A fundamental problem

If scientists judge a model's success by its ability to predict change correctly, then it appears that many land-use change models are failing. Typically, models extrapolate the change among land categories from time 1 to time 2. A validation procedure compares the models' predictions of time 2 to a reference map from time 2. Usually the percent correct is high, which engenders a naïve confidence in the model's predictive abilities. Closer inspection reveals the high percent correct is attributable primarily to the static state of the landscape between time 1 and time 2; if the model predicts persistence, then the model is usually accurate. Furthermore, a null model that predicts no change is often better than a model that predicts change, as the agreement between the reference map of time 1 and the reference map of time 2 is often greater than the agreement between a model's prediction of time 2 and the reference map of time 2. A model's prediction of time 2 can, upon visual inspection, appear to be an excellent fit due to the dominance of persistence, whereas the same model usually predicts incorrectly when it tries to predict change. Published literature shows that this phenomenon is more the rule than the exception (Wear and Bolstad, 1998; Mertens and Lambin, 2000; Geoghegan et al., 2001; Schneider and Pontius, 2001; Brown et al., 2002; Chen et al., 2002; Lo and Yang, 2002; Manson, 2002).

* Corresponding author. Tel.: +1-508-793-7761; fax: +1-508-793-8881.
E-mail address: rpontius@clarku.edu (R.G. Pontius Jr.).

0167-8809/$ – see front matter © 2003 Elsevier B.V. All rights reserved.
doi:10.1016/j.agee.2003.09.008

These failures derive in part from the fact that the most commonly used statistical methods lack fundamental concepts important for land change analysis. The methods taught in most university programs fail to account for land persistence in a manner capable of detecting important signals of land change. If scientists fail to detect the most prominent signals of land change, then they can neither research nor model land change accurately. This paper introduces novel statistical methods to help identify signals of systematic processes within a pattern of land change. These methods will help our profession focus research on the strongest signals of systematic land change and ultimately to link pattern to process.

The most pragmatic way to analyze land change is to do the following: obtain maps from time 1 and time 2; examine the changes with a transition matrix to identify the most important transitions; and then research the processes that generate the transitions. The traditional cross-tabulation matrix or transition matrix follows the format of Table 1, wherein the rows display the categories of time 1 and the columns display the categories of time 2. The notation P_{ij} denotes the proportion of the landscape that experiences a transition from category i to category j where the number of categories is J. Entries on the diagonal indicate persistence, thus P_{jj} denotes the proportion of the landscape that shows persistence of category j. Entries off the diagonal indicate a transition from category i to a different category j. In the Total column, the notation P_{i+} denotes the proportion of the landscape in category i in time 1, which is the sum over all j of P_{ij}. In the Total row, the notation P_{+j} denotes the proportion of the landscape in category j in time 2, which is the sum over all i of P_{ij}. Usually the transition matrix ends there. Table 1, however, shows an additional row and an additional column. The additional column on the right indicates the proportion of the landscape that experiences gross loss of category i between time 1 and time 2. The additional row on the bottom indicates the proportion of the landscape that experiences gross gain of category j between time 1 and time 2.

Most statistics courses teach us to analyze the matrix with a chi-square type of test. The chi-square approach compares the matrix of observed values to a matrix of expected values that are generated by random chance. The chi-square approach computes the expected values by assuming that each total, P_{i+} and $P_{+j,}$ is given a priori. The expected proportion of the landscape that experiences a transition from category i to category j due to random chance is P_{i+} times P_{+j}. More specifically, the expected proportion of the landscape that experiences persistence of category j due to chance is P_{j+} times P_{+j}. Eq. (1) gives the formula for the chi-square statistic, where N is the number of grid cells in the map:

$$X^2 = \sum_{i=1}^{J}\sum_{j=1}^{J}\left\{N \times \frac{[P_{ij} - (P_{i+} \times P_{+j})]^2}{(P_{i+} \times P_{+j})}\right\} \qquad (1)$$

For nearly all landscapes, the observed persistence is much greater than the expected persistence due to chance according to the standard chi-square calculation, so the chi-square result appropriately detects a strong signal of persistence. The problem is that scientists usually already know that persistence dominates the landscape. Scientists want to identify the dominant signals of land change. The methods of this

Table 1
General cross-tabulation matrix for comparing two maps from different points in time

	Time 2				Total time 1	Loss
	Category 1	Category 2	Category 3	Category 4		
Time 1						
Category 1	P_{11}	P_{12}	P_{13}	P_{14}	P_{1+}	$P_{1+} - P_{11}$
Category 2	P_{21}	P_{22}	P_{23}	P_{24}	P_{2+}	$P_{2+} - P_{22}$
Category 3	P_{31}	P_{32}	P_{33}	P_{34}	P_{3+}	$P_{3+} - P_{33}$
Category 4	P_{41}	P_{42}	P_{43}	P_{44}	P_{+}	$P_{4+} - P_{44}$
Total time 2	P_{+1}	P_{+2}	P_{+3}	P_{+4}	1	
Gain	$P_{+1} - P_{11}$	$P_{+2} - P_{22}$	$P_{+3} - P_{33}$	$P_{+4} - P_{44}$		

paper allow scientists to identify the signals of change separately from any given level of persistence.

1.2. Approaching the solution

Scientists can analyze Table 1 at several levels of detail. This paper espouses an analytical progression from general to more detailed levels of information.

At the most general level of information, the Total row lists the quantity of each category at time 2 and the Total column lists the quantity of each category at time 1. The difference between the two is termed the "net change". While this information can be useful, a lack of net change does not necessarily indicate a lack of change on the landscape. It is possible that change occurs in such a way that the location of a category changes between time 1 and time 2, while the quantity remains the same. For example, a given quantity of forest loss at one location can be accompanied by the same quantity of forest gain at another location. This type of change in location is termed a "swap". A net change in the quantity of a category indicates a definite change on the landscape; a lack of net change does not necessarily indicate a lack of change on the landscape since the net change fails to capture the swapping component of change.

The concept of swap is particularly important because some of the most common sources of land-cover data are available in a form that gives only the quantity of each land-cover type over time. The United Nations Food and Agriculture Organization publishes data concerning the area of various land-cover types by country by year (FAOSTAT, 2002). Due to its format, the data can be used to compute the annual net change of the area of any land-cover type, but cannot be used to compute the gross gain, gross loss or swap of any category. If the total forest area for a country is constant, it is impossible to know whether the landscape is stable or whether deforestation is accompanied by regrowth. The danger is that the net change can dramatically underestimate the total change on the landscape. For example, Mertens and Lambin (2000) analyzed the entire transition matrix to find that the net deforestation was less than half of the gross deforestation in their study area in Cameroon from 1991 to 1996. In order to quantify the total change, a scientist must have data in the format of Table 1 and examine the table at a greater level of detail than just the row and column totals. Unfortunately, it is common practice to report only net change (Yang and Lo, 2002). This problem of data format is so prevalent that many researchers are left with little alternative but to use data of only net change (Gallopin et al., 1997).

The diagonal entries of Table 1 indicate the total amount of persistence, which dominates most landscapes, including those where authors claim that the change is important and/or large (Wear and Bolstad, 1998; Mertens and Lambin, 2000; Geoghegan et al., 2001; Saczuk, 2001; Schneider and Pontius, 2001; Chen et al., 2002). Even in the Atlanta Metropolitan Area, which is renowned as one of the United States' fastest growing metropolises, there has been 75% persistence over the last three decades (Yang, 2002; Yang and Lo, 2002). Consequently, it is important that statistical methods account for persistence when examining land change. The persistence shown on the diagonal of Table 1 is required to compute two types of change: gains and losses. As mentioned previously, the bottom row shows the quantity gained for each category and the right-hand column shows the quantity lost for each category. The gains are the differences between the column totals and persistence. The losses are the differences between row totals and persistence.

Analysis of persistence, gains and losses is instructive, but it fails to inform whether there are systematic transitions among the categories because this general analysis fails to examine the dynamics among the off-diagonal entries of Table 1. The following methods section shows how to analyze the off-diagonal entries to identify systematic transitions of land change for a given landscape's degree of persistence. A subsequent subsection shows how to perform the analysis at multiple resolutions in order to examine how land change occurs over geographic distance. The paper also formalizes terminology that is useful for discussing common types of land change.

Taken in their entirety, the methods sections of this paper allow land-cover change scientists to answer the following battery of increasingly detailed questions: (1) What is the net change of each category? (2) What are the gain, loss and swap of each category? (3) What are the most systematic transitions among categories? (4) Over what distances does change occur for each category?

2. Methods

2.1. Data and a naïve matrix

Maps for 10 towns in central Massachusetts for the years 1971 and 1999 illustrate the methods (MassGIS, 2002). Figs. 1 and 2 show these maps for four categories: Forest, Open, Residential, and Other. Forest is a land-cover category as defined by the classification system of the maps from MassGIS. The Open category consists of abandoned agriculture, power lines, and undeveloped areas that lack dense vegetation. Residential land consists of residential areas with quarter- to half-acre lots. Other land comprises mostly high-density residential, agricultural, and transportation land. Each map contains 651,488 grid cells, wherein each cell has a resolution of 30 m × 30 m.

Fig. 3 focuses on the change in the Forest category from 1971 to 1999. Light gray shows persistence of forest and dark gray shows persistence of non-forest. Black shows deforestation and white shows forest re-growth. There is a net loss of forest since there is more black than white. Fig. 4 shows the change in the Open category. Open accounts for a small proportion of the landscape and there are equal amounts of gain and loss, thus the net change in Open is zero. Also notice that the patches of gain and loss of Open are not clustered particularly close together. Fig. 5 shows the change in the Residential category, which is characterized by a net increase with almost no loss of Residential area, hence almost no swapping.

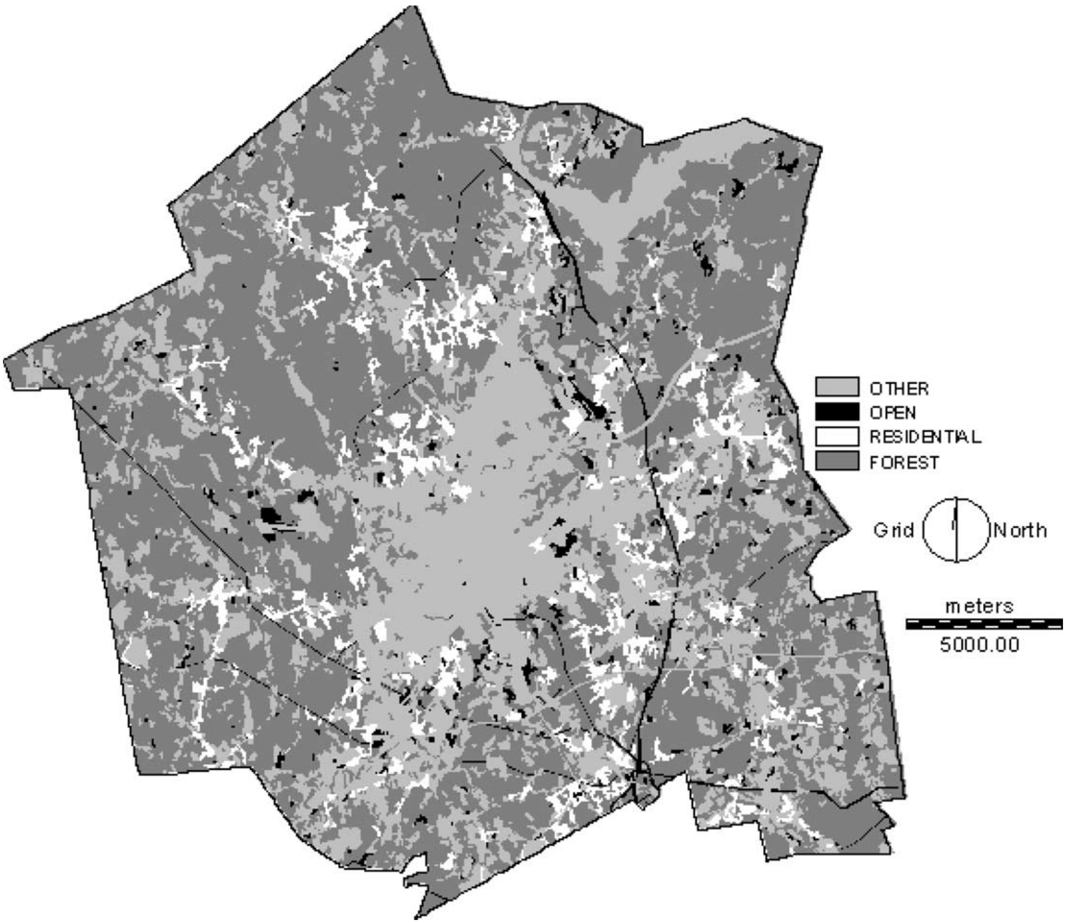

Fig. 1. This map shows the land cover of the study area in 1971 according to a four-category classification scheme.

Fig. 2. This map shows the land cover of the study area in 1999 according to a four-category classification scheme.

The 1971 and the 1999 maps are compared to produce a cross-tabulation matrix that shows the percentage of the landscape within each combination of categories. Tables 2 and 3 earmark these percentage values in bold in the same arrangement of rows and columns as Table 1.

The first step in analyzing the matrices of Tables 2 and 3 is to examine the Total column and the Total row, which demonstrates that the two largest categories are Forest and Other for both 1971 and 1999. The next step is to examine the diagonal entries, which calculate the percentage of the landscape that persists for each category. The diagonal entries are used to compute the gains and losses within each category, according to the formulas of Table 1. This computation exhibits that Forest experiences the largest loss, 7.20% of the landscape, and that Other experiences the largest gain, 5.50% of the landscape. The next subsection focuses on the persistence, gains and losses of Tables 2 and 3; the subsequent subsection analyzes the off-diagonal entries of Tables 2 and 3.

2.2. Net change and swap

This subsection shows how to answer two questions: (1) What is the net change of each category? (2) What are the gain, loss and swap of each category? Table 4 gives the answer to the first question by computing for each category j the absolute value of net change, which is $|P_{+j} - P_{j+}|$.

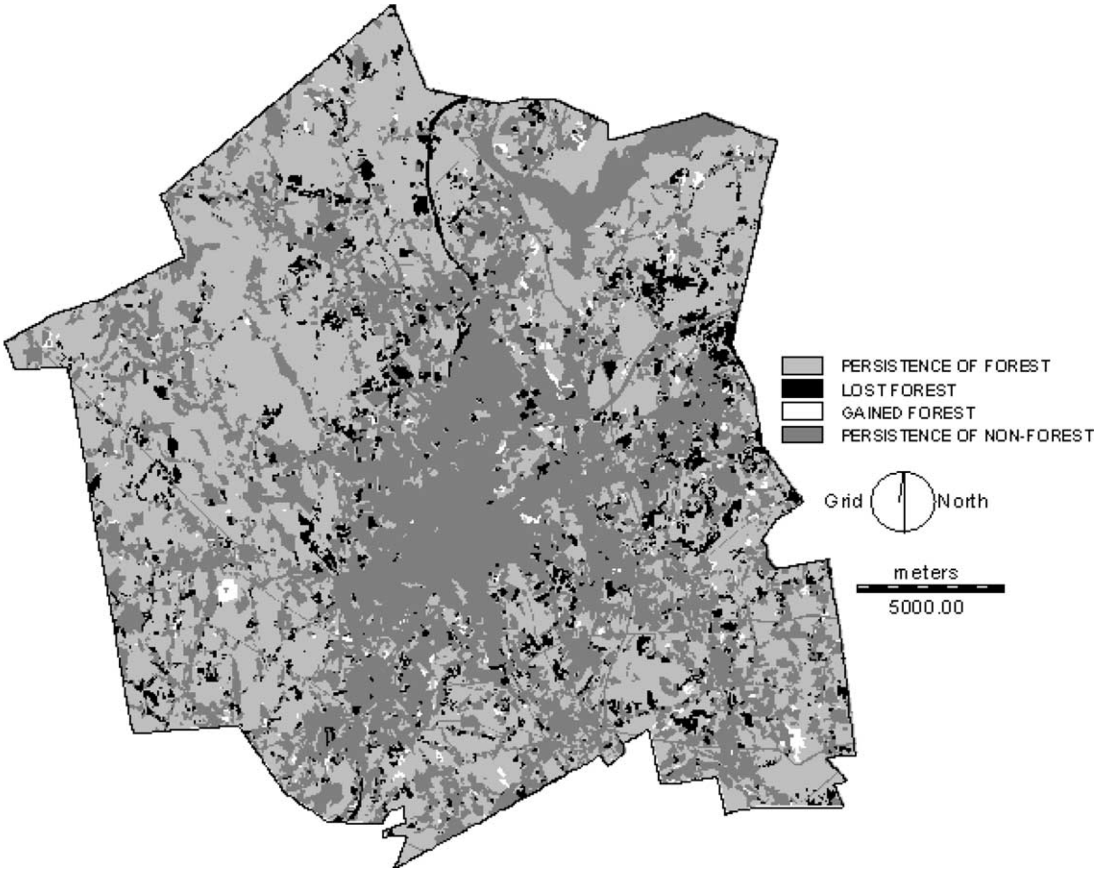

Fig. 3. This map shows the change in the Forest category between 1971 and 1999.

To answer the second question, Table 2 gives additional information concerning gain, loss, persistence, and swap. Table 2 illustrates the concept of swap by showing that the quantity of the Open category in both 1971 and 1999 is 2.78%. If one knew only the total quantity at each time, a naïve interpretation would conclude that the Open category does not change. However, the persistence in the Open category is 1.72% of the landscape, so the amount of gain between time 1 and time 2 is 1.06%, which is equal to the loss between time 1 and time 2. Therefore, all change in the Open category is a swapping-change dynamic. Fig. 4 demonstrates this swapping.

In contrast, the Residential category indicates almost no swapping. For the Residential category, the total in 1971 is 6.50% of the landscape and the persistence is 6.48%, so only 0.02% of the landscape shows loss of Residential. In 1999, 8.92% of the landscape is Residential, so 2.45% of the landscape shows a gain in Residential. Most of the change in the Residential category is net change. Fig. 5 shows this net change.

Eqs. (2)–(4) formalize the language of these fundamental types of change for any particular category. Eq. (2) defines the amount of swap, denoted S_j, for each category j, as two times the minimum of the gain and loss. Each grid cell that gains is paired with a grid cell that loses to create a pair of grid cells that swap. For the Open category, it is possible to pair each gain with a loss because the amount of gain is equal to the amount of loss. For the Residential category, however, it is not possible to pair each gain with a loss

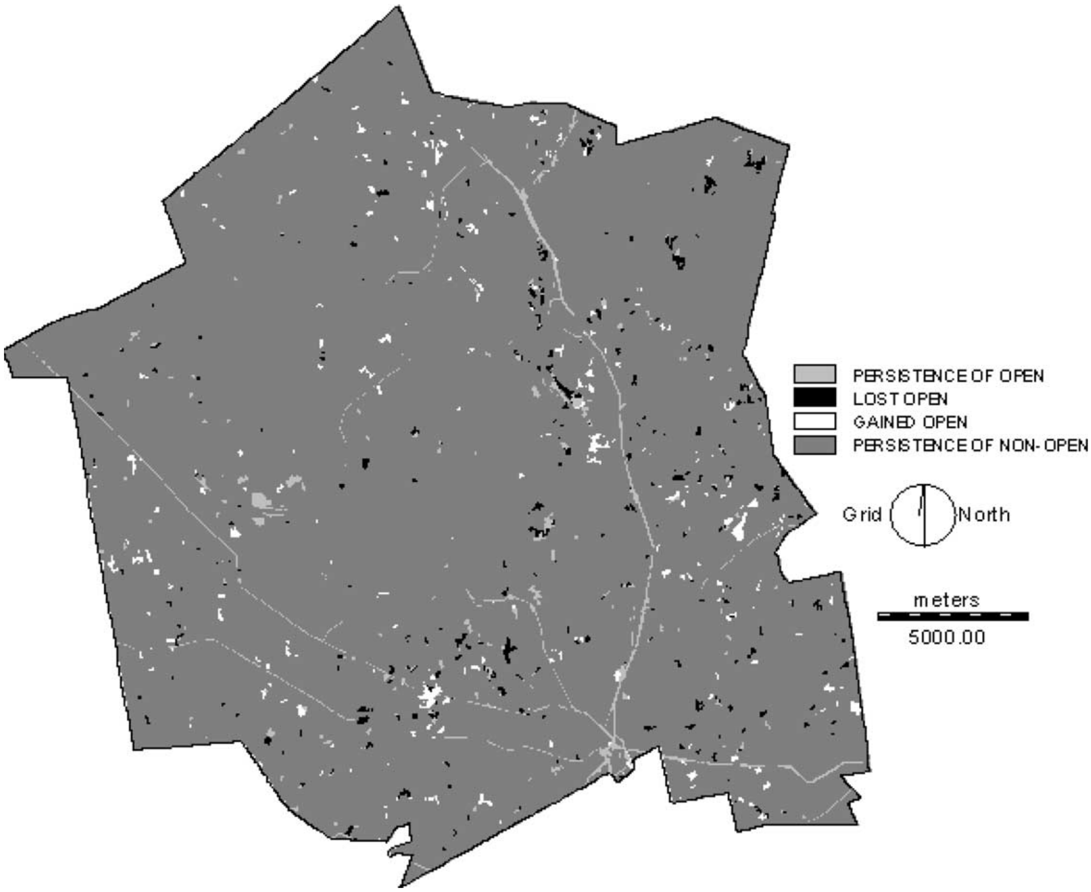

Fig. 4. This map shows the change in the Open category between 1971 and 1999.

because the amount of gain is larger than the amount of loss:

$$S_j = 2 \times \mathrm{MIN}(P_{j+} - P_{jj}, P_{+j} - P_{jj}) \quad (2)$$

Eq. (3) defines the absolute value of the net change, denoted D_j, for category j as the maximum of the gain and loss minus the minimum of the gain and loss. This net change is the remaining unpaired gain or loss after all gains and losses have been paired to compute the amount of swap. Therefore, the net change for the Open category is zero, and the net change for the Residential category is 2.43% of the landscape. Eq. (3) is a helpful way to think about net change. Eq. (3) yields the same results as the previously mentioned simple formula, $|P_{+j} - P_{j+}|$:

$$D_j = \mathrm{MAX}(P_{j+} - P_{jj}, P_{+j} - P_{jj}) - \mathrm{MIN}(P_{j+} - P_{jj}, P_{+j} - P_{jj}) = |P_{+j} - P_{j+}| \quad (3)$$

Eq. (4) shows that one can express total change for each category as either the sum of the net change and swap or the sum of the gains and losses. Notice that if $\mathrm{MAX}(P_{j+} - P_{jj}, P_{+j} - P_{jj})$ is the gain, then $MIN(P_{j+} - P_{jj}, P_{+j} - P_{jj})$ is the loss; and if $\mathrm{MAX}(P_{j+} - P_{jj}, P_{+j} - P_{jj})$ is the loss, then $\mathrm{MIN}(P_{j+} - P_{jj}, P_{+j} - P_{jj})$ is the gain. Table 4 shows how Eq. (4) applies to the example for each category:

$$C_j = D_j + S_j = \mathrm{MAX}(P_{j+} - P_{jj}, P_{+j} - P_{jj}) + \mathrm{MIN}(P_{j+} - P_{jj}, P_{+j} - P_{jj}) \quad (4)$$

Fig. 5. This map shows the change in the Residential category between 1971 and 1999.

Table 4 shows also how to aggregate the change of the individual categories to compute the change for the entire landscape. The change for the landscape is equal to the total gains of the individual categories, which is equal to the total losses of the individual categories. The Total change column of Table 4 shows that the sum of the changes in the individual categories double-counts the change on the landscape because change in one grid cell counts as a gain in one category and a loss in another category, such that the total change on the landscape is one-half the sum of the changes in the individual categories. Similarly, the total swap on the landscape is one-half the sum of the swaps in the individual categories; and the total net change on the landscape is one-half the sum of the net changes in the individual categories.

2.3. Inter-category transitions

The next step is to examine the off-diagonal entries of the cross-tabulation matrix, which show that the most prominent transition is conversion from Forest to Other, accounting for 5.00% of the landscape. Therefore, a naïve interpretation of Table 2 would indicate that the most systematic process on the landscape is the transition from Forest to Other. However, such an interpretation fails to consider that Forest and Other are the two largest categories; even a random process of land change would cause a large transition from Forest to Other. The fact that the largest transition is from Forest to Other is insufficient evidence to conclude that the Other category is systematically targeting the Forest category for replacement. In order to identify systematic transitions within the transition

Table 2
This matrix analyzes percent of land change in terms of gains[a]

	1999				Total 1971	Loss
	Forest	Open	Residential	Other		
1971						
Forest	**48.74**	**0.33**	**1.86**	**5.00**	**55.94**	**7.20**
	48.74	*0.61*	*1.46*	*4.71*	*55.53*	*6.79*
	(0.00)	(−0.28)	(0.40)	(0.29)	(0.41)	(0.41)
	[0.00]	[−0.45]	[0.27]	[0.06]	[0.01]	[0.06]
Open	**0.48**	**1.72**	**0.11**	**0.47**	**2.78**	**1.06**
	0.07	*1.72*	*0.07*	*0.23*	*2.09*	*0.37*
	(0.42)	(0.00)	(0.03)	(0.24)	(0.68)	(0.68)
	[6.14]	[0.00]	[0.45]	[1.01]	[0.33]	[1.83]
Residential	**0.00**	**0.00**	**6.48**	**0.02**	**6.50**	**0.02**
	0.16	*0.07*	*6.48*	*0.55*	*7.25*	*0.77*
	(−0.16)	(−0.07)	(0.00)	(−0.52)	(−0.75)	(−0.75)
	[−0.99]	[−0.99]	[0.00]	[−0.96]	[−0.10]	[−0.97]
Other	**0.59**	**0.73**	**0.48**	**32.99**	**34.78**	**1.80**
	0.85	*0.38*	*0.91*	*32.99*	*35.13*	*2.14*
	(−0.26)	(0.35)	(−0.43)	(0.00)	(−0.34)	(−0.34)
	[−0.30]	[0.91]	[−0.47]	[0.00]	[−0.01]	[−0.16]
Total 1999	**49.81**	**2.78**	**8.92**	**38.48**	**100.00**	**10.08**
	49.81	*2.78*	*8.92*	*38.48*	*100.00*	*10.08*
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]
Gain	**1.07**	**1.06**	**2.45**	**5.50**	**10.08**	
	1.07	*1.06*	*2.45*	*5.50*	*10.08*	
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	
	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	

[a] The number in bold is the actual percent of the landscape. The number in italics is the percent of the landscape that would be expected if the process of change were random. The number in round parentheses is the actual minus expected percent. The number in square brackets is the number in round parentheses divided by the number in italics.

matrix, one must interpret the transitions relative to the sizes of the categories. The following sections explain how to do this using the other numbers in Tables 2 and 3.

Table 2 gives four numbers for each combination of categories in time 1 and time 2. The top number in bold is the combination's percent observed on the landscape. Below the bold number, the second number, in italics, is the combination's percent that would be expected if the gain in each category were to occur randomly as calculated by

$$G_{ij} = (P_{+j} - P_{jj})\left(\frac{P_{i+}}{\sum_{i=1,i\neq j}^{J} P_{i+}}\right) \quad (5)$$

Eq. (5) assumes that the gain of each category and the proportion of each category at time 2 is fixed, and then distributes the gain across the other categories according to the relative proportion of the other categories in time 1. These expected values represent a random process of gain because a category that gains will replace other categories in proportion to how those other categories populate the landscape at time 1, if it replaces the other categories at random. That is, Eq. (5) distributes the gain in each column among the off-diagonal entries within the column.

For the diagonal entries, the expected number is equal to the observed number so that the matrix resulting from random transitions has the same amount of persistence as the observed landscape. This is necessary in order to examine the off-diagonal transitions given the amount of observed persistence, i.e. to hold the persistence constant and thus account for it in the calculations.

Table 3
This matrix analyzes percent of land change in terms of losses[a]

	1999				Total 1971	Loss
	Forest	Open	Residential	Other		
1971						
Forest	**48.74**	**0.33**	**1.86**	**5.00**	**55.94**	**7.20**
	48.74	*0.40*	*1.28*	*5.52*	*55.94*	*7.20*
	(0.00)	(−0.06)	(0.58)	(−0.52)	(0.00)	(0.00)
	[0.00]	[−0.16]	[0.46]	[−0.09]	[0.00]	
Open	**0.48**	**1.72**	**0.11**	**0.47**	**2.78**	**1.06**
	0.54	*1.72*	*0.10*	*0.42*	*2.78*	*1.06*
	(−0.06)	(0.00)	(0.01)	(0.05)	(0.00)	(0.00)
	[−0.11]	[0.00]	[0.08]	[0.12]	[0.00]	
Residential	**0.00**	**0.00**	**6.48**	**0.02**	**6.50**	**0.02**
	0.01	*0.00*	*6.48*	*0.01*	*6.50*	*0.02*
	(−0.01)	(−0.00)	(0.00)	(0.01)	(0.00)	(0.00)
	[−0.91]	[−0.39]	[0.00]	[1.21]	[0.00]	
Other	**0.59**	**0.73**	**0.48**	**32.99**	**34.78**	**1.80**
	1.45	*0.08*	*0.26*	*32.99*	*34.78*	*1.80*
	(−0.86)	(0.65)	(0.22)	(0.00)	(0.00)	(0.00)
	[−0.59]	[7.97]	[0.83]	[0.00]	[0.00]	
Total 1999	**49.81**	**2.78**	**8.92**	**38.48**	**100.00**	**10.08**
	50.75	*2.20*	*8.11*	*38.94*	*100.00*	*10.08*
	(−0.94)	(0.58)	(0.81)	(−0.45)	(0.00)	(0.00)
	[−0.02]	[0.26]	[0.10]	[−0.01]	[0.00]	[0.00]
Gain	**1.07**	**1.06**	**2.45**	**5.50**	**10.08**	
	2.01	*0.48*	*1.63*	*5.95*	*10.08*	
	(−0.94)	(0.58)	(0.81)	(−0.45)	(0.00)	
	[−0.47]	[1.21]	[0.49]	[−0.08]	[0.00]	

[a]The number in bold is the actual percent of the landscape. The number in italics is the percent of the landscape that would be expected if the process of change were random. The number in round parentheses is the actual minus expected percent. The number in square brackets is the number in round parentheses divided by the number in italics.

In Table 2, the third number in round parentheses is the combination's observed proportion minus the proportion expected under a random process. In other words, the number in parentheses is the number in bold minus the number in italics, $P_{ij} - G_{ij}$. If the difference in parentheses is positive, then the category in that row lost more to the category in the column than would be expected by any random process of gain in that category of the column. If the difference in parentheses is negative, then the category in the row lost less to the category in the column than would be expected due to a random process of gain in that category of the column. The magnitude of the number in parentheses indicates the percent of the landscape.

Table 4
The values in the table express change in terms of percent of the landscape

	Gain	Loss	Total change	Swap	Absolute value of net change
Forest	1.07	7.20	8.27	2.14	6.13
Open	1.06	1.06	2.12	2.12	0.00
Residential	2.45	0.02	2.47	0.04	2.43
Other	5.50	1.80	7.30	3.60	3.70
Total	10.08	10.08	10.08	3.95	6.13

In Table 2, the fourth number in square brackets is the combination's relative difference between the observed number and the expected number. In other words, the number in brackets is the number in parentheses divided by the number in italics, $(P_{ij} - G_{ij})/G_{ij}$. These ratios are analogous to the ratios that form the basis of chi-square tests, which are (observed value − expected value)/expected value. The magnitude of the number in parentheses indicates the difference between the observed value and the expected value, relative to the magnitude of the expected value.

Each observed number in bold equals the expected number in italics for the entries in the diagonal, the Total row and the Gains row. This follows from the logic of Eq. (5), which holds the persistence and gains fixed for each category. Therefore, the differences in the Total row and the Gains row are zero. The differences in the Total column and the Losses column are non-zero, since they derive from the expected numbers after Eq. (5) is applied. In other words, the time 2 information is the fixed frame of reference for the analysis of Gains.

Table 3 shows the analysis of losses, which is analogous to the analysis of gains shown in Table 2. The logic of Table 3 is the same as Table 2, but the role of the rows and columns is switched. Table 3 shows four numbers for each combination of categories of time 1 and time 2. The top number in bold is the combination's percent observed on the landscape, which is the same as in Table 2. Below the bold number, the number in italics is the combination's percent that would be expected if the loss in each category were to occur randomly, as given by

$$L_{ij} = (P_{i+} - P_{ii}) \left(\frac{P_{+j}}{\sum_{j=1,\, j \neq i}^{J} P_{+j}} \right) \qquad (6)$$

Eq. (6) assumes that the loss of each category is fixed, and then distributes the loss across the other categories according to the relative proportion of the other categories in time 2. These expected values represent a random process of loss because when a category loses, if it is replaced by other categories at random, then it will be replaced by other categories in proportion to how those categories populate the landscape at time 2. In order to portray this concept, Eq. (6) distributes the loss in each row among the off-diagonal entries within the row. Again, the expected number is equal to the observed number for the diagonal entries in order to hold constant the level of persistence on the observed landscape.

In Table 3, the third number in round parentheses is the combination's observed proportion minus the proportion expected under a random process, which is computed as $P_{ij} - L_{ij}$. If the difference in parentheses is positive, then the category in the column gained more from the category in the row than would be expected due to a random process of loss in the category of the row. If the difference in parentheses is negative, then the category in the column gained less from the category in the row than would be expected due to a random process of loss in the category of the row.

In Table 3, the fourth number in square brackets is computed as $(P_{ij} - L_{ij})/L_{ij}$. Just as in Table 2, the magnitude of the number in parentheses indicates the difference between the observed value and the expected value, relative to the magnitude of the expected value.

Each observed number in bold equals the expected number in italics for the diagonal, the Total column and the Gains column. The differences between the observed and expected percentages are non-zero in the Total row and Gains row. This follows from the logic of Eq. (6), which holds the time 1 information as the fixed frame of reference for the analysis of losses. Section 3 interprets Tables 2 and 3.

2.4. Swapping distances

Up to this point, the bold numbers in Table 2 have been the basis of all of the calculations. However, Table 2 shows information for only one resolution, which is the fine 30 m grid cell resolution shown in Figs. 1 and 2. Therefore, Table 2 fails to show any information concerning the geographic distances over which transitions occur. This section describes a multiple-resolution procedure that detects distances over which land change occurs.

The multiple-resolution procedure compares the maps of Figs. 1 and 2 at various resolutions by aggregating contiguous blocks of fine cells into coarser grid cells, then performing calculations on the coarser grid cells. For example, a resolution of 2 aggregates each block of 2 × 2 fine cells into a coarser cell that is four times larger than a fine cell. Similarly, a resolution of 100 aggregates each contiguous block of 100 × 100 fine grid cells into a very coarse grid cell that is 10,000 times larger than a fine cell. Since the fine grid cells are 30 m × 30 m, the resolution of 100 depicts cells that are 3 km × 3 km.

The aggregation procedure produces coarse grid cells that have partial (i.e. fuzzy) membership in each of the categories. The coarse cell's membership in each category is the proportion of fine resolution cells

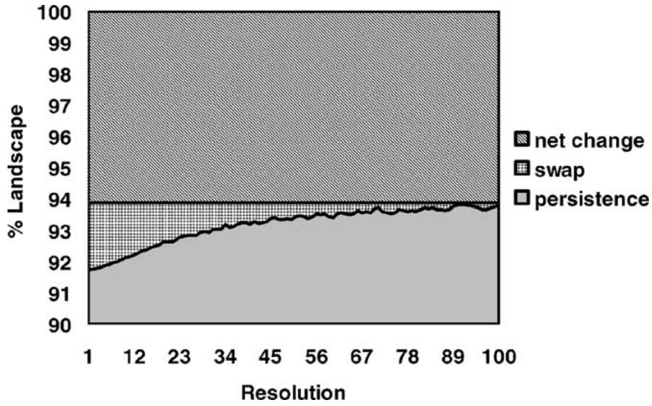

Fig. 6. The Forest change profile shows the amount of persistence, swap, and net change at multiple resolutions, where the resolution is a multiple of the length of the side of a pixel of the raw data.

of each category that constitute the coarse cell. The specification of location becomes less precise as resolution becomes coarser, but the proportion of each category remains constant in the landscape; thus, there is no aggregation bias.

Pontius (2002) shows how to compute the disagreement due to location and the disagreement due to quantity between any two maps at any and all resolutions. Pontius' methodology is used here to compute the swap and net change at multiple resolutions. The multiple-resolution aggregation procedure has no influence on the quantity of each category because the concept of quantity is independent of resolution. However, a change in resolution can have a dramatic influence on swap because the concept of swap depends on location. For example, at a fine resolution, if deforestation occurs in one cell and reforestation occurs in the neighboring cell, then the pair of cells constitutes a swap in the Forest category. However, at a slightly coarser resolution, the deforested cell and the reforested cell are aggregated into the same coarse cell; therefore, the swap disappears. As a result, the amount of swap can diminish at coarser resolutions.

Fig. 6 illustrates the concept of diminishing swap at coarser resolutions. The vertical axis reflects percent of landscape while the horizontal axis reflects resolution. The resolution becomes coarser as one moves from left to right on the horizontal axis. A resolution of 1 corresponds to 30 m grid cells and a resolution of 100 corresponds to 3 km grid cells. The component of change at the top of the figure is the net change, which does not vary with resolution. Below the net change is swap, which diminishes at coarser resolutions. If the resolution were to increase in a geometric sequence, such as 1, 2, 4, 8, 16, ..., then the swap would diminish monotonically. The non-monotonic character of Fig. 6 is attributable to a technical reason related to the fact that the sequence of resolutions is arithmetic, i.e. 1, 2, 3, ... The commensurate Section 3 highlights and interprets the most important findings derived from this Section 2.

3. Results

3.1. Net change and swap

The most obvious result is that one must look carefully to find any change based on a visual comparison of Fig. 1 with Fig. 2. This is because 90% of the landscape persists between 1971 and 1999. Therefore, it is important to use the methods outlined in this paper to analyze the signals of change in light of the overwhelming signal of persistence on the landscape.

Table 4 answers the first two of the most basic questions raised in the introduction, "What is the net change of each category?" and "What are the gain, loss and swap of each category?" Change in Forest consists of both swap and net change. Change in Open is a pure swap-type of change. Change in Residential is nearly pure net change. Change in Other consists of both swap and net change. On the landscape as a

Table 5
This table interprets the most systematic transitions shown in columns of Table 2

Transition	Observed minus expected	Difference divided by expected	Interpretation of systematic transition
Open in 1971 and Forest in 1999	0.42	6.14	When Forest gains, it replaces Open.
Residential in 1971 and Forest in 1999	−0.16	−0.99	When Forest gains, it does not replace Residential.
Forest in 1971 and Residential in 1999	0.40	0.27	When Residential gains, it replaces Forest.
Other in 1971 and Residential in 1999	−0.43	−0.47	When Residential gains, it does not replace Other.
Forest in 1971 and Other in 1999	0.29	0.06	When Other gains, it replaces Forest. This signal is weak.
Residential in 1971 and Other in 1999	−0.52	−0.96	When Other gains, it does not replace Residential.
Open in 1971 and non-Open in 1999	0.68	1.83	When non-Open categories gain, they replace Open. Open loses.
Residential in 1971 and non-Residential in 1999	−0.75	−0.97	When non-Residential categories gain, they no not replace Residential. Residential does not lose.

whole, the change attributable to quantity is larger than the change attributable to swap. Most of the change is associated with the Forest and Other categories, due in part to the fact that these two are the largest categories at both time 1 and time 2.

3.2. Inter-category transitions

Tables 5 and 6 answer another common question concerning a finer level of detail, "What are the most systematic transitions among categories?" If the gains had occurred at random categories, then the differences in round parentheses of Table 2 would all be zero, but many of the differences are not near zero. Table 5 interprets the most important results of Table 2. The first two rows of Table 5 indicate a systematic pattern in which Forest replaces Open but does not replace Residential. Specifically, when Forest gains, it replaces Open at a rate over six times the rate that would be expected if Forest were to gain randomly. Furthermore, if Forest were to gain randomly, then one would expect 0.16% of the landscape to undergo conversion from Residential to Forest, but instead, only scant conversion exists from Residential to Forest between 1971 and 1999. The next two rows of Table 5 indicate that when Residential gains, it is inclined to gain from Forest systematically and is disinclined to gain from Other systematically. The last two rows of Table 5 substantiate that when categories gain they are inclined to replace Open and disinclined to replace Residential in a manner different from what would be expected due to random processes. The analysis shows that Residential tends to persist and Open tends to lose.

If the processes of loss had occurred at random categories, then the differences shown in Table 3 would all be zero, but many are not. Table 6 interprets the most prominent results of Table 3. The first two rows of Table 6 show that when Forest loses, Residential

Table 6
This table interprets the most systematic transitions shown in the rows of Table 3

Transition	Observed minus expected	Difference divided by expected	Interpretation of systematic transition
Forest in 1971 and Residential in 1999	0.58	0.46	When Forest loses, Residential replaces it.
Forest in 1971 and Other in 1999	−0.52	−0.09	When Forest loses, Other does not replace it.
Other in 1971 and Forest in 1999	−0.86	−0.59	When Other loses, Forest does not replace it.
Other in 1971 and Open in 1999	0.65	7.97	When Other loses, Open replaces it.
Non-Forest in 1971 and Forest in 1999	−0.94	−0.47	When non-Forest categories lose, non-Forest categories replace them. Forest does not gain.
Non-Open in 1971 and Open in 1999	0.58	1.21	When non-Open categories lose, Open replaces them. Open gains.
Non-Residential in 1971 and Residential in 1999	0.81	0.49	When non-Residential categories lose, Residential replaces them. Residential gains.

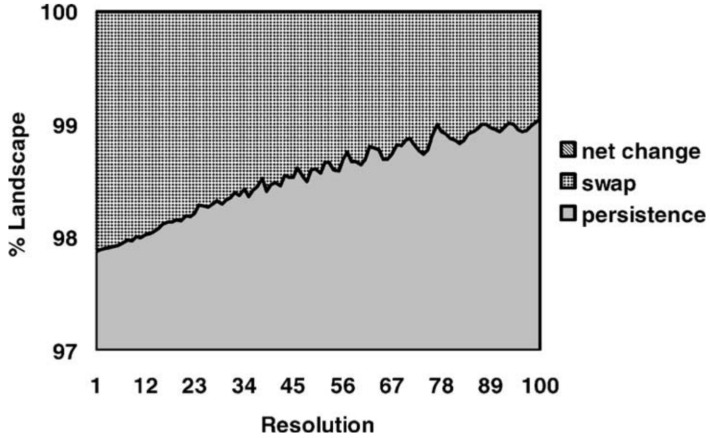

Fig. 7. The Open change profile shows the amount of persistence, swap, and net change at multiple resolutions, where the resolution is a multiple of the length of the side of a pixel of the raw data.

rather than Other tends to replace it. The last three rows of Table 6 concur that when categories lose, they tend to be replaced by Open and Residential but not by Forest.

3.3. Swapping distances

Figs. 6–8 answer the most detailed question, "Over what distances does change occur for each category?" Fig. 6 shows that nearly all the swapping in the Forest category occurs over distances of less than 3 km. This is evident because the swap is negligible at the coarsest resolution of three kilometers. This indicates that gain of Forest happens within 3 km of the nearest loss in the Forest category.

Fig. 4 shows that many of the gains in the Open category are not near the losses in the Open category. Fig. 7 illustrates this by showing that the swap is 2% of the landscape at the finest resolution and is 1% at the coarsest resolution. Thus, about half of the swap in the Open category transpires over distances greater than 3 km.

Fig. 5 shows that the Residential category gains far more than it loses. Fig. 8 shows that the small

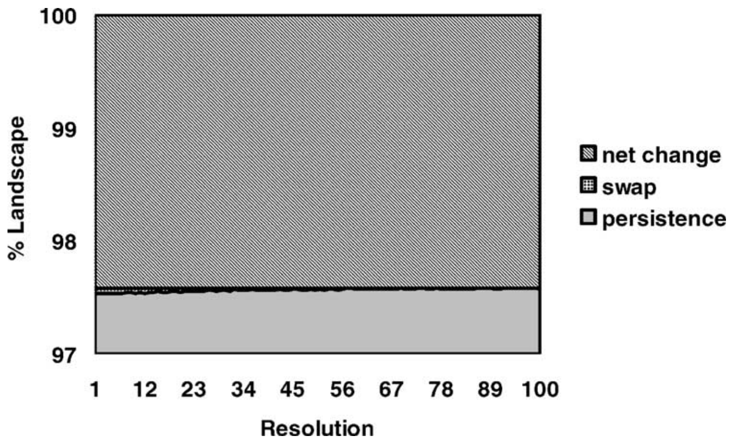

Fig. 8. The Residential change profile shows the amount of persistence, swap, and net change at multiple resolutions, where the resolution is a multiple of the length of the side of a pixel of the raw data.

degree of swap disappears by resolution 30. Hence, the nominal level of Residential loss unfolds within 900 m of Residential gain.

4. Discussion

4.1. Importance of methodology

This paper describes statistical methods for examining non-experimental observational data in order to detect signals of possible cause-and-effect processes. The methods are not designed to detect the mechanisms of land transformation. Landscape ecologists and geographers must perform additional research to identify the processes that create extant landscape patterns (Gardner et al., 1987; Turner, 1990; Gustafson and Parker, 1992). This paper's methods are first steps toward helping scientists focus such research on the most prevalent systematic processes of land change.

For the example, the methods elucidate what a less detailed conventional analysis might overlook. First, one of the most important systematic transitions is the conversion from Forest to Residential land. This transition is particularly important because it is largely permanent; Residential land tends toward persistence and Forest land tends away from swapping. The large conversion of Forest to Other may be attributable to the fact that Forest and Other are the two largest categories, since the quantity of the conversion is nearly equivalent to what would be expected from a random process. Relative to its size, the Open category tends to gain the most and to lose the most, and the patches of gain and loss are far from each other. When Forest regrows, it tends to regrow within three kilometers of where deforestation occurs. Such information is instructive when one begins to investigate the processes that explain these transitions.

A simplistic interpretation of the transition matrix in Table 3 may induce scientists to focus only on the largest transition from Forest to Other since it accounts for half of the landscape change. On the other hand, this large transitional dynamic does not necessarily indicate that Forest is losing systematically to Other, nor does it follow that Other is systematically gaining from Forest. To find the requisite evidence for a systematic process one must examine Tables 2 and 3. Table 2 shows that when Other gains it indeed tends to replace Forest, but not by much more than the expectations from random processes. Table 3 shows that when Forest loses, it tends to be replaced systematically by Residential and not by Other. So, if scientists are inclined to research transitions from Forest, then they should focus attention on the systematic transition from Forest to Residential, not exclusively on the transition from Forest to Other. If scientists focus research only on the predominating transitions, they are likely to omit the most systematic transition processes.

Certainly, scientists should consider any major transition, such as the transition from Forest to Other, but the types of questions scientists ask about a particular transition should depend in part on the insights of this paper's prescribed methods. For example, the transition from Forest to Other might be due largely to the fact that Forest tends toward loss for reasons independent of the other categories. In fact, the Loss column of Table 2 confirms that the observed loss of Forest is larger than the loss that would be expected from random gains in the other categories. On the other hand, scientists could hypothesize that the transition from Forest to Other is large due exclusively to the fact that Other tends to gain for reasons independent of the other categories. Table 3 gives evidence against this hypothesis because the Gain row shows that the observed gain of Other is less than the gain that would be expected from random losses in the other categories.

4.2. Significance of differences

Tables 2 and 3 express the difference between the observed values and the expected values in two distinct ways. The first is by subtraction, which yields the values in parentheses. The second way is by division, which yields the values in brackets. An obvious question is, "How large a difference is needed in order for it to be considered important?" There are several issues scientists should bear in mind when answering this question.

The subtraction method gives values that indicate a percent of the landscape; the magnitude of the difference indicates the size of the fingerprint left on the landscape due to a systematic transition. The magnitude of the transition is important relative to the magnitude of the total change on the landscape. In the example, all of the differences are less than 1% of the landscape, but the total amount of change is only

10% of the landscape. It is easier for larger categories to leave larger fingerprints on the landscape, even when a transition is not particularly systematic. It is subsequently critical to consider the division method as well.

The division method gives values that indicate a systematic process relative to the size of the category involved. The magnitude of the ratio indicates the strength of the systematic transition. In the example, Table 6 shows that largest ratio occurs when Open replaces Other at a rate nearly eight times beyond what would be expected due to chance. However, the subtraction method shows that this systematic transition leaves a fingerprint that accounts for a difference of only 0.65% of the landscape. It is common for small categories to have large ratios of transition because small categories can have small denominators in the ratio. A large ratio means that the transition is systematic. It is also common for small categories to have minute differences that explain only a small fraction of the landscape.

For several reasons, scientists should resist the temptation to perform hypothesis testing to detect differences that are statistically significant. First, statistical significance does not necessarily indicate practical importance. Second, the units of observation are grid cells that are determined by the technology of a GIS system that manipulates the maps. Hence, the number of grid cells does not indicate an appropriate number of degrees of freedom. Third, any statistical hypothesis testing would be fraught with complications due to spatial auto-correlation. In conclusion, the simple subtraction method and division method offer scientists sufficient tools to think critically about which transitions have practical importance for a particular application.

To illustrate these points about statistical hypothesis testing, note that when Eq. (1) is applied to Table 2, it yields a chi-square statistic greater than one million, with nine degrees of freedom, so that the P-value is much smaller than 0.0001. This chi-square value is statistically significant because persistence accounts for the strong agreement between the maps and there are a large number of grid cells, as denoted in Eq. (1) for which $N = 651{,}488$. Eq. (1) shows clearly that as N increases, the value of the chi-square statistic increases. The size of N, however, is not necessarily a function of the landscape; it is a function of the decision of the scientist concerning the format of the digital map. In this case, conventional statistical hypothesis testing yields information that is not particularly important to understanding the pattern of change on the landscape.

4.3. Future work

This paper's multiple-resolution analysis of swapping scratches the surface of potentially fruitful research concerning scale. In particular, swapping distance that would be expected from a random process of swap is not yet considered. One should be aware of the random swapping distance when interpreting the observed swapping distance. In the example, a random pattern of swapping in the Open category would likely produce a large swapping distance because there is a dispersed but scant amount of Open category on the landscape. In other words, there is plenty of space for patches of gain to exist far from patches of loss for the Open category. The situation for Forest is the reverse. A random pattern of swapping in the Forest category would likely produce a small swapping distance because Forest is spread over half of the landscape.

It would be even more interesting to examine the distances at which transitions among categories occur. A prerequisite for such analysis is the ability to create a cross-tabulation matrix of Table 1 format for a map in which each coarse grid cell has partial membership in several categories. Lewis and Brown (2001) have devised one approach to generate such a matrix; however, their derivation is not useful for multiple-resolution analysis for technical reasons. We are in the process of creating a method by which scientists can compute a cross-tabulation matrix for all possible resolutions. This will ultimately allow examination of how category transitions change with scale, which is important because factors that determine land change at fine local scales are not the same factors which determine land change at coarser global scales (Turner et al., 1989; Li and Reynolds, 1997).

5. Conclusions

This paper establishes terminology and methodology for analyzing land change. Scientists in the fields of geography and landscape ecology should adopt these methods because existing popular methods fail to segregate land change according to its different components and thus fail to gain maximum insight into the processes driving the change. This paper endorses an approach that moves from broad to more detailed levels of observation. Persistence usually accounts for the close association between maps of two points in time. After persistence is accounted for, it is useful to view total change in terms of two pairs of components: net change and swap, as well as gross gains and gross losses. Scientists can then detect systematic transitions of land change by comparing the observed change to the expected change arising from chance for any given degree of persistence. Multiple-resolution analysis is useful in examining the distances over which swap occurs. These new methods should help scientists to focus research on the most important land transitions and ultimately to facilitate linking pattern to process.

Acknowledgements

National Science Foundation supported this work through the grant "Infrastructure to Develop a Human-Environmental Regional Observatory Network" (Award ID 9978052), and a Research Experience for Undergraduates supplementary fund under a National Science Foundation Prime Award No. BCS-9978052 (Penn State University Log No. 56301). The authors also acknowledge the Center for Integrated Study of the Human Dimensions of Global Change of Carnegie Mellon University, which was created through a cooperative agreement between the National Science Foundation (NSF, SRB-9521914) and Carnegie Mellon University, with additional support from the Electric Power Research Institute, the Exxon Mobil Foundation, and the American Petroleum Institute. Additional support comes from the General Fund at Clark University, the Howard and Leah Green fund at the George Perkins Marsh Institute.

References

Brown, D., Goovaerts, P., Burnicki, A., Li, M.-Y., 2002. Stochastic simulation of land-cover change using geostatistics and generalized additive models. Photogramm. Eng. Remote Sensing 68 (10), 1051–1061.

Chen, J., Gong, P., He, C., Luo, W., Tamura, M., Shi, P., 2002. Assessment of the urban development plan of Beijing by using a CA-based urban growth model. Photogramm. Eng. Remote Sensing 68 (10), 1063–1071.

FAOSTAT, 2002. FAOSTAT Agriculture Data, Land Domain. http://apps.fao.org/page/collections.

Gallopin, G., Hammond, A., Raskin, P., Swart, R., 1997. Branch Points: Global Scenarios and Human Choice, PoleStar Series Report 7. Stockholm Environment Institute, Stockholm.

Gardner, R.H., Milne, B.T., Turner, M.G., O'Neill, R.V., 1987. Neutral models for the analysis of broad-scale landscape pattern. Landscape Ecol. 1 (1), 19–28.

Geoghegan, J., Villar, S.C., Klepeis, P., Mendoza, P.M., Ogneva-Himmelberger, Y., Roy Chowdhury, R., Turner, B.L., Vance, C., 2001. Modeling tropical deforestation in the southern Yucatan peninsular region: comparing survey and satellite data. Agric. Ecosyst. Environ. 85 (1–3), 25–46.

Gustafson, E.J., Parker, G.R., 1992. Relationships between landcover proportion and indices of landscape spatial pattern. Landscape Ecol. 7 (2), 101–110.

Lewis, H.G., Brown, M., 2001. A generalized confusion matrix for assessing area estimates from remotely sensed data. Int. J. Remote Sensing 22 (16), 3223–3235.

Li, H., Reynolds, J.F., 1997. Modeling effects of spatial pattern, drought, and grazing on rates of rangeland degradation: a combined Markov and cellular automaton approach. In: Quattrochi, D.A., Goodchild, M.F. (Eds.), Scale in Remote Sensing and GIS. Lewis, New York, pp. 211–230.

Lo, C.P., Yang, X., 2002. Drivers of land-use/land-cover changes and dynamic modeling for the Atlanta, Georgia metropolitan area. Photogramm. Eng. Remote Sensing 68 (10), 1073–1082.

Manson, S.M., 2002. Integrated assessment and projection of land-use and land-cover change in the Southern Yucatán Peninsular Region of Mexico. Doctoral Dissertation. Graduate School of Geography. Clark University, Worcester Massachusetts.

MassGIS, 2002. http://www.state.ma.us/mgis/.

Mertens, B., Lambin, E., 2000. Land-cover-change trajectories in southern Cameroon. Ann. Assoc. Am. Geogr. 90 (3), 467–494.

Pontius Jr., R.G., 2002. Statistical methods to partition effects of quantity and location during comparison of categorical maps at multiple resolutions. Photogramm. Eng. Remote Sensing 68 (10), 1041–1049.

Saczuk, E.A., 2001. Has the Kullu district experienced an increase in natural hazard activity over the past 27 years?—a case study in risk and land use/cover change. GIS@Development. http://www.gisdevelopment.net/magazine/gisdev/2001/oct/kullu.shtml.

Schneider, L., Pontius, R., 2001. Modeling land-use change in the Ipswich watershed, Massachusetts, USA. Agric. Ecosyst. Environ. 85 (1–3), 83–94.

Turner, M.G., 1990. Spatial and temporal analysis of landscape patterns. Landscape Ecol. 4 (1), 21–30.

Turner, M.G., Dale, V.H., Gardner, R.H., 1989. Predicting across scales: theory development and testing. Landscape Ecol. 3 (3/4), 245–252.

Wear, D., Bolstad, P., 1998. Land-use changes in southern Appalachian landscapes: spatial analysis and forecast evaluation. Ecosystems 1, 575–594.

Yang, X., 2002. Satellite monitoring of urban spatial growth in the Atlanta metropolitan area. Photogramm. Eng. Remote Sensing 68 (7), 725–734.

Yang, X., Lo, C.P., 2002. Using a time series of satellite imagery to detect land use and land cover changes in the Atlanta, Georgia metropolitan area. Int. J. Remote Sensing 23 (9), 1775–1798.