Ann Reg Sci (2008) 42:11–37
DOI 10.1007/s00168-007-0138-2

SPECIAL ISSUE PAPER

# Comparing the input, output, and validation maps for several models of land change

**Robert Gilmore Pontius Jr · Wideke Boersma · Jean-Christophe Castella · Keith Clarke · Ton de Nijs · Charles Dietzel · Zengqiang Duan · Eric Fotsing · Noah Goldstein · Kasper Kok · Eric Koomen · Christopher D. Lippitt · William McConnell · Alias Mohd Sood · Bryan Pijanowski · Snehal Pithadia · Sean Sweeney · Tran Ngoc Trung · A. Tom Veldkamp · Peter H. Verburg**

Received: 6 March 2006 / Revised: 19 July 2006 / Published online: 16 August 2007


**Abstract** This paper applies methods of multiple resolution map comparison to quantify characteristics for 13 applications of 9 different popular peer-reviewed land change models. Each modeling application simulates change of land categories in raster maps from an initial time to a subsequent time. For each modeling application, the statistical methods compare: (1) a reference map of the initial time, (2) a refer-

R. G. Pontius Jr (✉)
Department of International Developement, Community and Environment, School of Geography, Clark University, 950 Main Street, Worcester, MA 01610-1477, USA
e-mail: rpontius@clarku.edu

W. Boersma · T. de Nijs
Netherlands Environmental Assessment Agency, Bilthoven, The Netherlands
e-mail: wideke.boersma@mnp.nl

T. de Nijs
e-mail: Ton.de.Nijs@rivm.nl

J.-C. Castella
Department of Social Sciences, Institut de Recherche pour le Développement, Montpellier, France
e-mail: j.castella@ird.fr

K. Clarke · C. Dietzel · N. Goldstein
Department of Geography, University of California, Santa Barbara, CA, USA
e-mail: kclarke@geog.ucsb.edu

C. Dietzel
e-mail: dietzel@geog.ucsb.edu

N. Goldstein
e-mail: goldstein8@llnl.gov

Z. Duan
Resources and Environment College, China Agricultural University, Beijing, China
e-mail: duanzengqiang@yahoo.com.cn

ence map of the subsequent time, and (3) a prediction map of the subsequent time. The three possible two-map comparisons for each application characterize: (1) the dynamics of the landscape, (2) the behavior of the model, and (3) the accuracy of the prediction. The three-map comparison for each application specifies the amount of the prediction's accuracy that is attributable to land persistence versus land change. Results show that the amount of error is larger than the amount of correctly predicted change for 12 of the 13 applications at the resolution of the raw data. The applications are summarized and compared using two statistics: the null resolution and the figure of merit. According to the figure of merit, the more accurate applications are the

E. Fotsing
Computer Science Department of the University Institute of Technology at Bandjoun, and Center for Environmental Studies and Development in Cameroon, University of Dschang, Maroua, Cameroon
e-mail: fotsing@hotmail.com

K. Kok · A. T. Veldkamp · P. H. Verburg
Department of Environmental Sciences, Wageningen University, Wageningen, The Netherlands
e-mail: kasper.kok@wur.nl

A. T. Veldkamp
e-mail: Tom.Veldkamp@wur.nl

P. H. Verburg
e-mail: peter.verburg@wur.nl

E. Koomen
SPINlab, Department of Spatial Economics, Vrije Universiteit, Amsterdam, The Netherlands
e-mail: ekoomen@feweb.vu.nl

C. D. Lippitt
School of Geography, Clark University, Worcester, MA, USA
e-mail: christopher.lippitt@gmail.com

W. McConnell
Center for Systems Integration & Sustainability, Michigan State University, East Lansing, MI, USA
e-mail: mcconn64@msu.edu

A. Mohd Sood
Forestry Department Peninsular Malaysia, Kuala Lumpur, Malaysia
e-mail: alias@forestry.gov.my

B. Pijanowski · S. Pithadia
Department Forestry and Natural Resources, Purdue University, West Lafayette, IN, USA
e-mail: bpijanow@purdue.edu

S. Pithadia
e-mail: snehal_pithadia@unc.edu

S. Sweeney
Center for the Study of Institutions, Populations, and Environmental Change, Indiana University, Bloomington, IN, USA
e-mail: spsweene@indiana.edu

T. N. Trung
Mountain Agrarian Systems Program, Vietnam Agricultural Sciences Institute, Hanoi, Vietnam
e-mail: trungtnvn@yahoo.com

ones where the amount of observed net change in the reference maps is larger. This paper facilitates communication among land change modelers, because it illustrates the range of results for a variety of models using scientifically rigorous, generally applicable, and intellectually accessible statistical techniques.

**JEL Classification** C52 · C53 · Q15 · Q24 · R14 · R52

## 1 Introduction

Spatially-explicit models of land-use and land-cover change (LUCC) typically begin with a digital map of an initial time and then simulate transitions in order to produce a prediction map for a subsequent time. Upon seeing the resulting prediction map, an obvious first question is, "How well did the model perform?" Whatever the level of performance, a common second question is, "How does the performance compare to the range that is typically found in land change modeling?" These apparently simple questions can quickly become tricky when scientists begin to decide upon the specific techniques of analysis to use in order to address the questions. This paper offers a set of concepts and accompanying analytical procedures to answer these questions in a manner that is useful for many modeling applications and is intellectually accessible for many audiences.

This paper illustrates how to answer these questions by analyzing a collection of modeling applications that have been submitted in response to a call for voluntary participation in an international comparison exercise. The invitation requested each participant to submit three maps: (1) a reference map of an initial time 1, (2) a reference map of a subsequent time 2, and (3) a prediction map for the subsequent time 2. The reference maps are by definition the most accurate maps available for the particular points in time, so they serve as the basis to measure the accuracy of the prediction. All of the techniques and measurements in this paper derive from fairly simple but carefully thought out overlays of various combinations of the three raster maps for each modeling application.

In addition, we asked the contributing scientists to describe three qualitative characteristics of their modeling applications: (1) the processes of land transformation, (2) the accuracy of the reference data, and (3) the structure of the model. The invitation requested that each LUCC model generates its prediction map based on information at or before time 1, meaning that the LUCC model should not use information subsequent to time 1 to help to predict the change between time 1 and time 2. All submissions were welcomed regardless of whether they satisfied this criterion. For applications where the criterion was not satisfied, we asked the participant to describe how the model uses information subsequent to time 1 for calibration. Most importantly, the exercise welcomed any modelers who would be willing to allow us to inspect, analyze, and present their modeling results openly in a level of detail that was not specified a priori. Many of the modelers who participated did so because they think this type of open comparative exercise is crucial to move the field of land change modeling forward.

Seven different laboratories contributed maps from 18 different applications of 9 different land-change models to 12 different sites from around the world. This article includes all scientists who sent maps of land cover categories and presents the most representative application from each model as it applies to a particular landscape. Consequently, this article presents 13 applications that demonstrate various landscape dynamics, data formats, modeling approaches, and common challenges.

The models used a variety of techniques including linear extrapolation, suitability mapping, genetic algorithms, neural networks, scenario analysis, expert opinion, public participation, and agent-based modeling. While each of the models has its unique features, Table 1 summarizes some of the distinguishing features that are commonly used to characterize models. In Table 1, "statistical regression" means that the model uses statistical regression as a major technique of calibration somewhere in its approach. "Cellular automata" means that the model's decision concerning whether to change the state of a pixel takes into consideration explicitly the state of the neighboring pixels. "Machine learning" means that the model's algorithm runs for an indefinite period of time until it learns the patterns in the calibration data, then uses the learned patterns to make a prediction. "Exogenous quantity" means that the model's user specifies the quantity of each category in the prediction map independently from the location of categories. "Pure pixels" means that the model uses pixels that have complete membership to exactly one category, as opposed to mixed pixels that have partial membership to more than one category. For the entries of Table 1, "Yes" means that the model has the characteristic as a fundamental feature. "Optional" means that the model's user has the option to use the feature for any particular application. "No" means that the model does not include the feature.

Table 2 describes important characteristics of the reference maps and the specific modeling applications. The 13 contributions include applications to 12 different locations, since two models apply to The Netherlands, albeit with the data formatted differently. Among the applications, the number of categories ranges from 2 to 15, the time interval of the prediction ranges from 4 to 43 years, the spatial resolution of the pixels ranges from 26 m to 15 km, and the spatial extent ranges from 123 to 96,975 km$^2$. Consequently, the number of pixels ranges from 216 to 13 million, which indicates the range for the level of detail in the maps and resulting demands for computational resources. For some models, the sample includes applications to more

**Table 1** Characteristics of nine models

| LUCC Model | Statistical Regression | Cellular Automata | Machine Learning | Exogenous Quantity | Pure Pixels |
|---|---|---|---|---|---|
| Geomod | Optional | Optional | No | Yes | Yes |
| SLEUTH | Yes | Yes | Yes | No | Yes |
| Land Use Scanner | Optional | No | No | Yes | No |
| Environment Explorer | Optional | Yes | No | Optional | Optional |
| Logistic Regression | Yes | No | No | Yes | Yes |
| SAMBA | No | Optional | No | No | Yes |
| LTM | No | Optional | Yes | Yes | Yes |
| CLUE-S | Optional | Optional | No | Yes | Yes |
| CLUE | Yes | No | No | Yes | No |

**Table 2** Characteristics of reference and prediction maps for thirteen modeling applications

| Site name | Spatial extent ($km^2$) | Spatial resolution (m) | # of pixels | # of classes | Year 1 | Year 2 | Year interval | Uses year 2 quantity | Null resolution (km)[d] | LUCC Model |
|---|---|---|---|---|---|---|---|---|---|---|
| Worcester, USA | 586 | 30 | 651, 591 | 2 | 1971 | 1999 | 28 | No | 4 | Geomod |
| Santa Barbara, USA | 123 | 50 | 49, 210 | 7 | 1986 | 1998 | 12 | No | 3 | SLEUTH |
| Holland(8) | 37, 280 | 500 | 149, 119 | 8[a] | 1996 | 2000 | 4 | No | Worse | Land Use Scanner |
| Holland(15) | 37, 280 | 500 | 149, 119 | 15 | 1996 | 2000 | 4 | No | 16 | Environment Explorer |
| Perinet, Madagascar | 715 | 30 | 794, 955 | 2[b] | 1957 | 2000 | 43 | No | Better | Logistic Regression |
| Cho Don, Vietnam | 892 | 32 | 892, 136 | 6 | 1990 | 2001 | 11 | No | Better | SAMBA |
| Detroit, USA | 9, 175 | 26 | 13, 209, 072 | 2[b] | 1978 | 1998 | 20 | Yes | 27 | LTM |
| Twin Cities, USA | 6, 347 | 30 | 7, 052, 459 | 2[b] | 1991 | 1998 | 7 | Yes | 2 | LTM |
| Maroua, Cameroon | 3, 572 | 250 | 57, 144 | 6 | 1987 | 1999 | 12 | Yes | Better | CLUE-S |
| Kuala Lumpur, Malaysia | 3, 810 | 150 | 169, 333 | 6 | 1990 | 1999 | 9 | Yes | Better | CLUE-S |
| Haidian, China | 431 | 100 | 43, 077 | 8 | 1991 | 2001 | 9 | Yes | Better | CLUE-S |
| Honduras | 96, 975 | 15, 000 | 431 | 6[c] | 1974 | 1993 | 19 | Yes | Better | CLUE |
| Costa Rica | 48, 600 | 15, 000 | 216 | 6[c] | 1973 | 1984 | 11 | Yes | Better | CLUE |

[a] The original pixels contain partial membership to 36 categories, which are reassigned to one of eight categories for this exercise. This reformatting introduces considerable additional error in the predicted quantity of change

[b] The reference and the prediction maps are designed to show exclusively a one-way transition

[c] The pixels contain simultaneous partial membership to multiple categories

[d] "Better" means that the LUCC model is more accurate than its corresponding null model at all resolutions. "Worse" means that the LUCC model is less accurate than its corresponding null model at all resolutions

than one site, which reveals how a single model can behave differently on different landscapes.

This is a voluntary sample, so it is not assured to be representative of all land-change modeling. However, this sample covers a wide range of analytical approaches and landscape types, which serves the purpose of the exercise. All of the models in the sample have passed peer-review in scientific literature. The authors include many prominent leaders in the field who have been developing their models for decades.

The authors realize from the beginning that this exercise has tremendous potential for misinterpretation, so we are careful to state the characteristics that some readers might initially assume this exercise has, but in fact lacks. This exercise is not a competition, meaning that we are not looking to crown the best LUCC model and we are not intending to rank the models. It is not the goal of this exercise to congratulate LUCC modelers for our successes or to condemn LUCC modelers for our failures. A major purpose of the exercise is to allow us to communicate in ways that are not possible by reading each others publications or by focusing on a single model at a time.

In order to compare both quantitative and qualitative aspects of the modeling applications, the methods are described in two parts. The first part immediately below describes the techniques to analyze the applications with a unified statistical approach. The second part describes each individual modeling application in the Appendix. The results section highlights the most important findings. The discussion section illuminates the lessons learned.

## 2 Methods

### 2.1 Three possible two-map comparisons

This subsection describes how we summarize the modeling applications by comparing pairs of maps for each application. There are three possible two-map comparisons, given the three maps submitted for each application. Comparison between the reference map of time 1 and the reference map of time 2 characterizes the observed change in the maps, which reflects the dynamics of the landscape. Comparison between the reference map of time 1 and the prediction map of time 2 characterizes the model's predicted change, which reflects the behavior of the model. Comparison between the reference map of time 2 and the prediction map of time 2 characterizes the accuracy of the prediction, which is frequently a primary interest. In order to interpret this third two-map comparison properly, it is necessary to consider the preceding two two-map comparisons. It is important to consider all three of these two-map comparisons for each application in order to compare across modeling applications.

Figures 1 and 2 use one of the applications to illustrate the analytic procedure that we apply to all 13 applications. The example application considers the maps for Worcester, MA, USA, which has two categories, built and non-built. Figure 1 shows the three possible two-map overlays, while Fig. 2 quantifies the differences in each of the two-map overlays.

Figure 1a examines the difference between the reference map of time 1 and the reference map of time 2, which Fig. 2 quantifies by the length of the top bar labeled

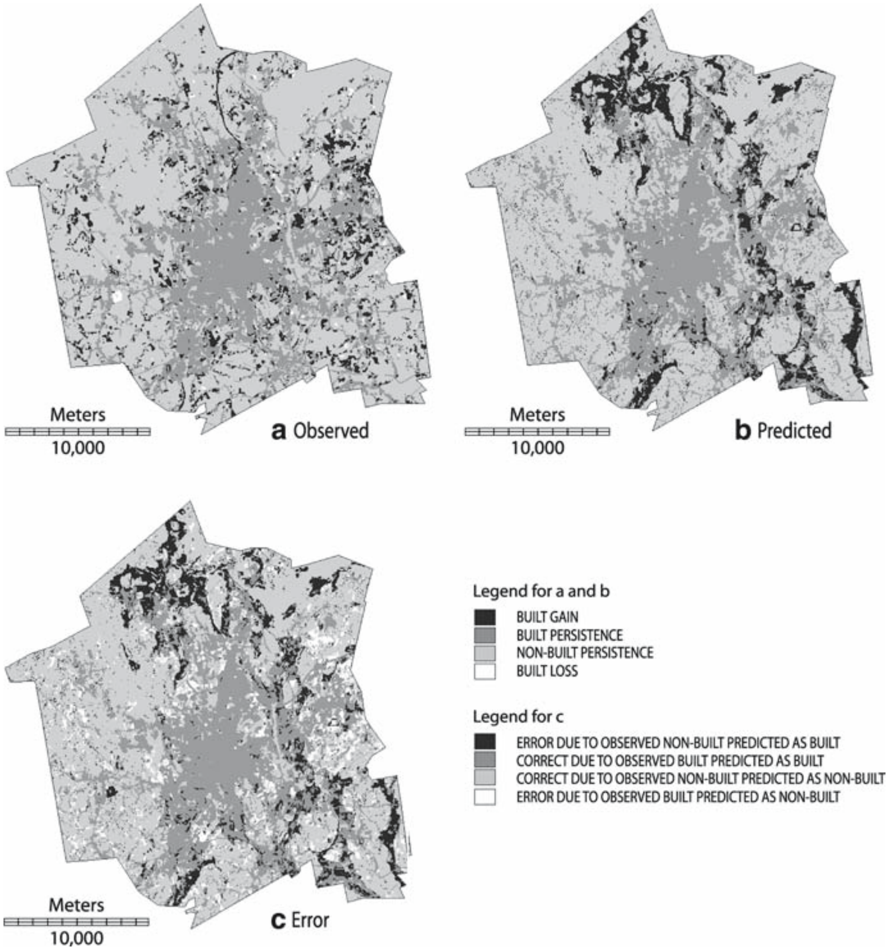


**Fig. 1** The Worcester maps of: **a** observed change 1971–1999, **b** predicted change 1971–1999, and **c** prediction error 1999

Worcester-Observed. A series of papers describe how to budget the total disagreement between any two maps that share a categorical variable in terms of separable components (Pontius 2000, 2002; Pontius et al. 2004b). The two most important components are quantity disagreement (i.e., net change) and location disagreement (i.e., swap change), which sum to the total disagreement. Quantity disagreement derives from differences between the maps in terms of the number of pixels for each category. Location disagreement is the disagreement that could be resolved by rearranging the pixels spatially within one map so that its agreement with the other map is as large as possible. For the Worcester application, most of the observed change is quantity disagreement since the gain of built is larger than the loss of built, while there is some location disagreement since there exists simultaneous gain and loss of built.

Figure 1b examines the difference between the reference map of time 1 and the prediction map for time 2, which Fig. 2 quantifies by the length of the bar labeled

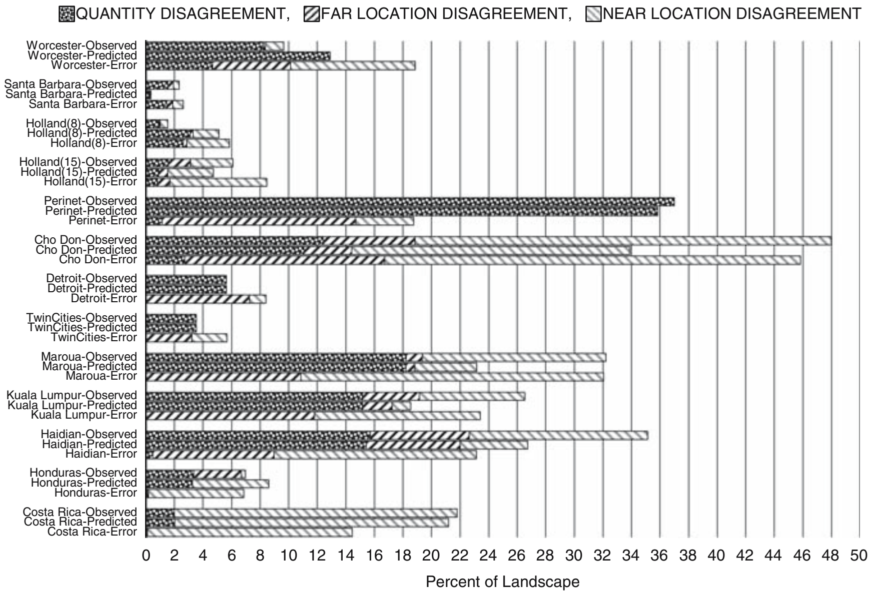


**Fig. 2** Observed change, predicted change, and prediction error for 13 applications. Near location disagreement becomes resolved at a resolution of 64 times the original fine-resolution pixels

Worcester-Predicted. If the model were to predict the observed change perfectly, then Fig. 1a would be identical to Fig. 1b, and the Worcester-Observed bar would be identical to the Worcester-Predicted bar in Fig. 2. However, the model predicts gain of built and no loss of built. Consequently, the Worcester-Predicted bar in Fig. 2 shows only quantity disagreement and zero location disagreement. Ultimately, we want to know whether the model predicts time 2 accurately, which is why we must compare the reference map of time 2 to the prediction map.

Figure 1c examines the difference between the reference map of time 2 and the prediction map for time 2, which Fig. 2 quantifies by the length of the bar labeled Worcester-Error. Most of the error is location disagreement, which occurs primarily because the model predicts land change at the wrong locations. It would be possible to fix two pixels of location error within the prediction map by swapping the location of a pixel of incorrectly predicted built with the location of a pixel of incorrectly predicted non-built. If the location disagreement can be resolved by swapping the pixels over small distances, then Fig. 2 budgets the error as “near” location disagreement. If the location disagreement cannot be resolved by swapping over small distances, then Fig. 2 budgets the error as “far” location disagreement. In this paper, near location disagreement is defined specifically as the location disagreement that can be resolved by swapping within 64-row × 64-column clusters of pixels of the raw data. In order to distinguish near location disagreement from far location disagreement, we convert each application’s maps to a coarser resolution where the side of each coarse pixel is 64 times larger than the spatial resolution in Table 2. The coarsening procedure uses an averaging rule that maintains the quantity of each category in the map, so the coarsening procedure does not affect quantity disagreement. The coarsening procedure can cause the location disagreement to shrink, where the amount of shrinkage is equal to the near location disagreement and the remaining location disagreement is equal to the far location disagreement. Pontius (2002) and Pontius and Cheuk (2006) describe in greater depth the method to compute the difference between near location disagreement and far location disagreement.

The bars of Fig. 2 are helpful to compare the 13 applications because they use a single technique to show important characteristics for each application. Furthermore, it is essential to consider the observed and predicted bars in order to interpret the error bar properly. In particular, the observed bar is the error of a null model that predicts pure persistence, i.e., no change between time 1 and time 2; so if the observed bar is smaller than the error bar, then the null model is more accurate than the LUCC model, as the Worcester application illustrates. It is also important to consider the components of the predicted bar, because the error bar is a function of the LUCC model's ability to predict the correct: amount of quantity change, amount of location change, amount of each particular transition from one category to another category, and location of each particular transition.

### 2.2 One possible three-map comparison

An additional validation technique considers the overlay of all three maps: the reference map of time 1, the reference map of time 2, and the prediction map for time 2 (Fig. 3). This three-map comparison allows one to distinguish the pixels that are correct due to persistence versus the pixels that are correct due to change. The black pixels in Fig. 3 show where the LUCC model predicts change correctly. Dark gray pixels show where change is observed and the LUCC model predicts change, however, the model predicts a transition to the wrong category, which is a type of error that can occur in multi-category models. Medium gray pixels show error where change is observed at locations where the model predicts persistence. Light gray pixels show error where persistence is observed at locations where the model predicts change. White pixels show locations where the LUCC model predicts persistence correctly or locations that are excluded from the results. The exclusion applies to some of the pixels in the applications of Logistic Regression, i.e., Perinet, and Land Transformation Model, i.e. Detroit and Twin Cities, because those two models simulate a one-way transition from non-disturbed to disturbed. The validation results exclude pixels are that are already disturbed at the initial time, because those pixels are not candidates for change according to the structure of Logistic Regression and Land Transformation Model approaches. The reader can obtain color versions of the maps by contacting the first author or visiting www.clarku.edu/∼rpontius.

A null model that predicts complete persistence would predict correctly the white pixels but not the black pixels. Furthermore, a null model would predict correctly the light gray pixels, but would predict incorrectly the medium and dark gray pixels. Thus, a LUCC model is more accurate than its corresponding null model for any application where there are more black pixels than light gray pixels.

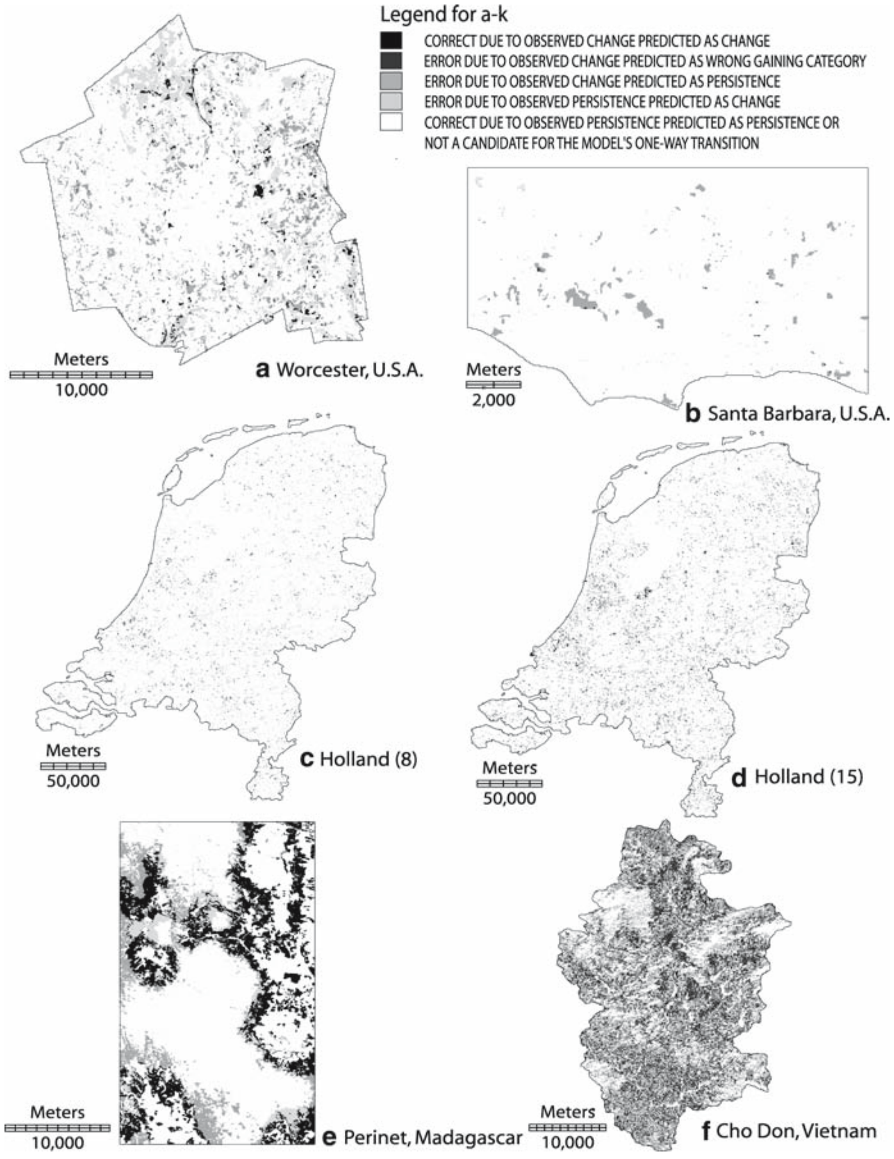


**Fig. 3** Validation maps for thirteen applications obtained by overlaying the reference map of time 1, reference map of time 2, and prediction map for time 2

Figure 3 allows the reader to assess visually the nature of the prediction errors, which are various shades of gray. For example, there are more light gray pixels than medium gray pixels in Fig. 3a, which indicates the presence of quantity disagreement in the error for the Worcester case. This type of disagreement occurs when the LUCC model predicts more change than is observed, as the Worcester bars of Fig. 2 indicate. The LUCC model would need to predict a different quantity of change in order to resolve quantity disagreement in the error. Figure 3a also shows location

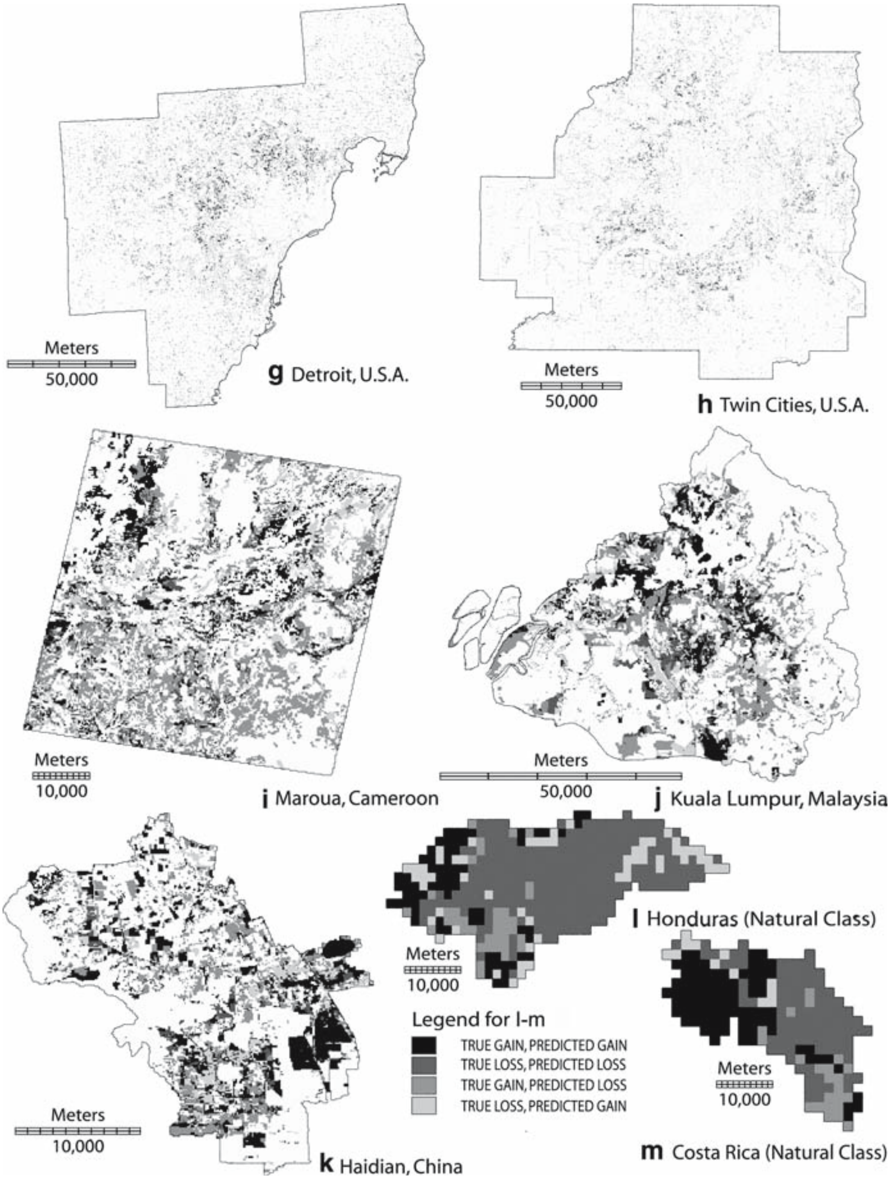


**Fig. 3** continued

disagreement in the error. The LUCC model would need to move the predicted change of the light gray pixels to the observed change of the medium gray pixels in order to resolve location disagreement. If the light gray pixels are close to the medium gray pixels, then the location error is considered "near" in Fig. 2, otherwise the location error is "far".

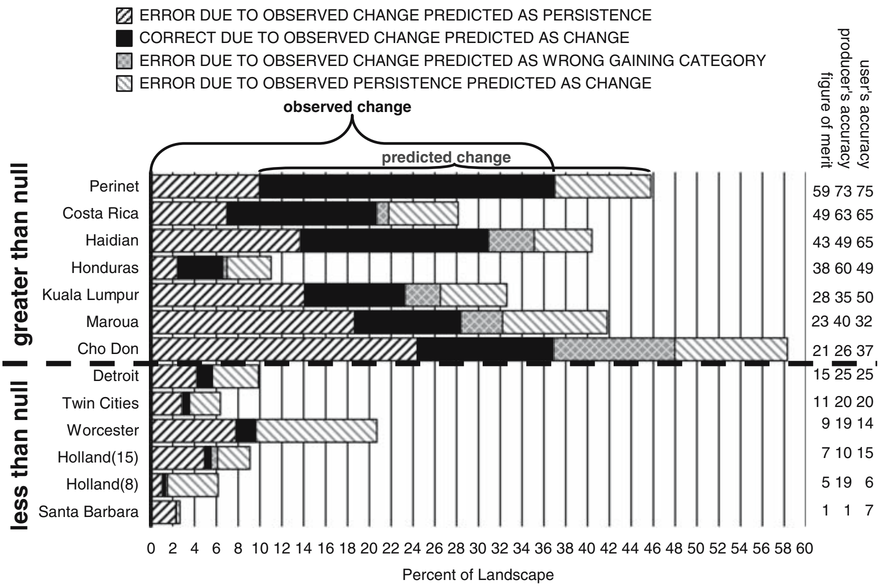


**Fig. 4** Sources of percent correct and percent error in the validation for 13 modeling applications. Each bar is a Venn diagram where the solid and cross-hatched segments show the intersection of the observed change and the predicted change

The applications for Honduras and Costa Rica have heterogeneous pixels, so their maps in Fig. 3 have a different legend than the other 11 applications. Figure 3l–m shows the dynamics of only the nature category, which is the single most prominent category.

Figure 4 presents a summary of the applications according to the logic of the legend for Fig. 3a–k. Each bar is a rectangular Venn diagram where the solid and cross-hatched central segments represent the intersection of the observed change and the predicted change; the central solid black segment is change that the model predicts correctly. The union of the segments on the left and center portions of each bar represents the area of change according to the reference maps, and the union of the segments on the center and right portions of each bar represents the area of change according to the prediction map. If a prediction were perfect, then its entire bar would have exactly one solid segment, which would have a length equal to both the observed change and the predicted change.

The "figure of merit" is a statistical measurement that derives directly from the information in the segments of the bars in Fig. 4. The figure of merit is the ratio of the intersection of the observed change and predicted change to the union of the observed change and predicted change (Klug et al. 1992; Perica and Foufoula-Georgiou 1996). This translates in Fig. 4 as the ratio of the length of the segment of correctly predicted change to the length of the entire bar, which Eq. 1 defines mathematically. The figure of merit can range from 0%, meaning no overlap between observed and predicted change, to 100%, meaning perfect overlap between observed and predicted change, i.e., a perfectly accurate prediction.

$$\text{Figure of merit} = B\big/(A + B + C + D) \tag{1}$$

where $A$ is the area of error due to observed change predicted as persistence, $B$ area of correct due to observed change predicted as change, $C$ area of error due to observed change predicted as wrong gaining category, and $D$ area of error due to observed persistence predicted as change.

Figure 4 can also be used to show two types of conditional accuracy, which some scientists call producer's accuracy and user's accuracy. Equation 2 gives the producer's accuracy, which is the proportion of pixels that the model predicts accurately as change, given that the reference maps indicate observed change. Equation 3 gives the user's accuracy, which is the proportion of pixels that the model predicts accurately as change, given that the model predicts change. Figure 4 expresses these statistics as percents.

$$\text{Producer's Accuracy} = B\big/(A + B + C) \tag{2}$$

$$\text{User's Accuracy} = B\big/(B + C + D) \tag{3}$$

Pontius et al. (2004b) describe an additional statistical method of validation that considers all three maps simultaneously. The technique compares the accuracy of the LUCC model to the accuracy of its null model at multiple resolutions. The accuracy of the LUCC model is the percent of pixels in agreement for the comparison between the reference map of time 2 and the prediction map for time 2, shown by the solid circles in Fig. 5. The accuracy of the null model is the percent of pixels in agreement for the comparison between the reference map of time 1 and the reference map of time 2, shown by the solid triangles in Fig. 5. The horizontal axis shows the fine resolution of the raw data on the left and coarser resolutions to the right.

Overall agreement increases as resolution becomes coarser for both the LUCC model and its null model, when location disagreement becomes resolved as the resolution becomes coarser, as explained above in the description of near and far location disagreement (Pontius 2002; Pontius and Cheuk 2006). Overall agreement increases to the level at which the only remaining error is quantity disagreement, indicated by the horizontal dotted lines in Fig. 5. It is common for a LUCC model to have accuracy less than its null model at the fine resolution of the raw data. If a LUCC model predicts the quantity of the categories more accurately than its null model, then the LUCC model must be more accurate than its null model at the coarsest resolution, which is the resolution where the entire study area is in one large pixel. If the LUCC model is less accurate than its null model at a fine resolution and more accurate than its null model at a coarse resolution, then there must be a resolution at which the accuracy of the LUCC model is equal to the accuracy of its null model. Pontius et al. (2004b) define this resolution as the null resolution. For each application, Table 2 gives the null resolution in terms of kilometers of the side of a coarse pixel, which is computed as the length of the side of a fine resolution pixel of the raw data times the multiple of the pixel for the null resolution shown in Fig. 5. Smaller null resolutions indicate that location errors occur over smaller distances.

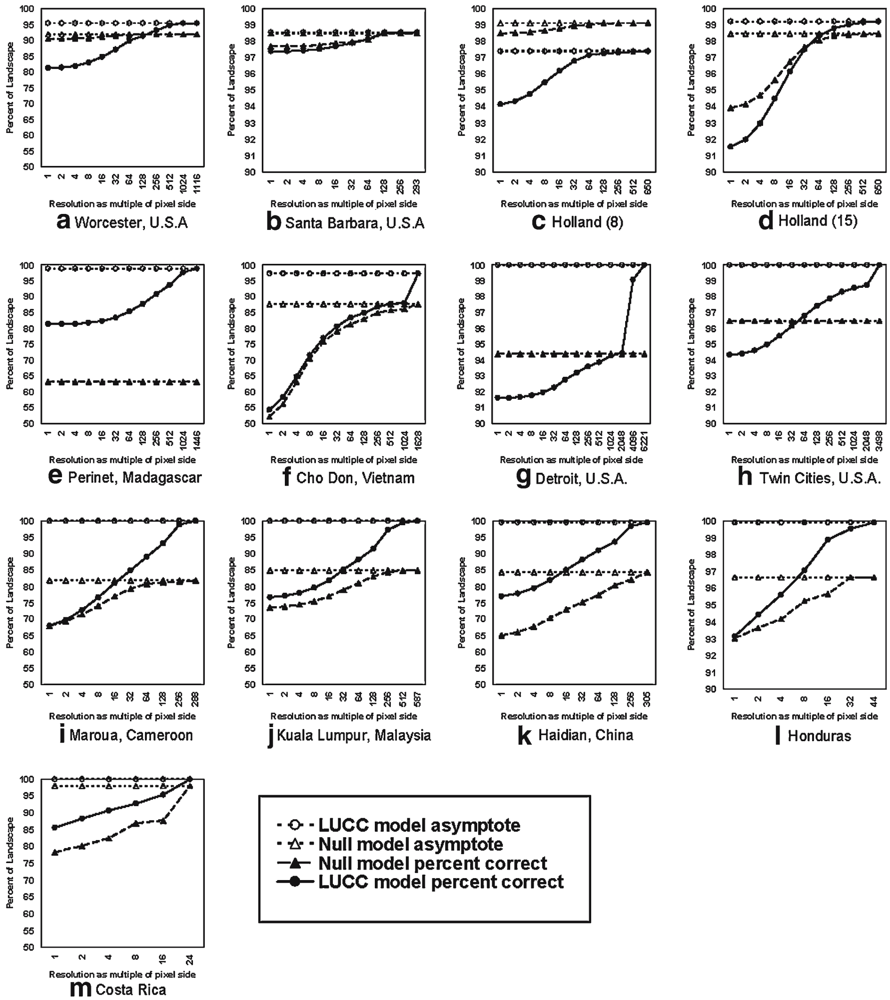


**Fig. 5** Percent correct at multiple resolutions for the 13 LUCC models and their respective null models. The null resolution is the resolution at which the percent correct for the LUCC model equals the percent correct for its null model

## 3 Results

Figure 4 summarizes the most important results in a manner that facilitates cross application comparison at the resolution of the raw data. The applications are ordered with respect to the figure of merit. Figure 4 shows that Perinet is the only application where the amount of correctly predicted change is larger than the sum of the various types of error, i.e., figure of merit is greater than 50%. Producer's accuracy is greater than 50% for Perinet, Honduras, and Costa Rica. User's accuracy is greater than 50% for Perinet, Haidian, and Costa Rica. The seven applications at the top of Fig. 4 are the ones that are more accurate than the null model at the resolution of the raw data. Table 2

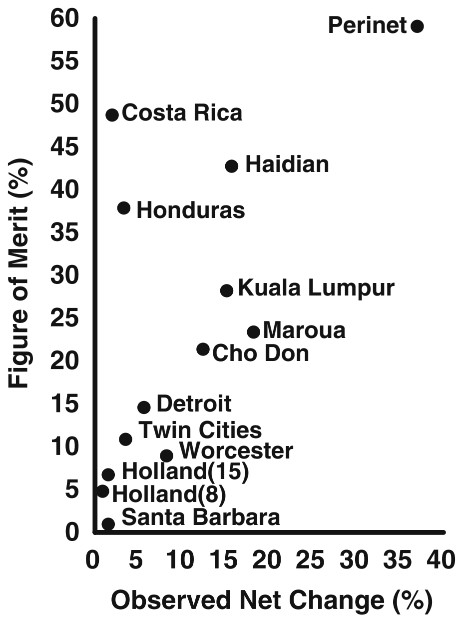


**Fig. 6** Positive relationship between the figure of merit (i.e., prediction accuracy) versus observed net change (i.e., landscape dynamics)

denotes these seven applications with the word "Better" in the column labeled "Null Resolution". Figure 5 summarizes the results at multiple resolutions, which reveals the null resolution.

Applications that have larger amounts of observed net change in the reference maps tend to have larger predictive accuracies as measured by the figure of merit; $R$-squared is 40% for the increasing linear relationship (Fig. 6). $R$-squared is 88%, if we ignore the two CLUE applications to Honduras and Costa Rica, which are fundamentally different than the other applications, because the two CLUE applications have heterogeneous pixels that are very few and very coarse compared to the other applications (Table 2). All six of the applications that have a figure of merit less than 15% have an observed net change of less than 10%. The applications that have a large figure of merit are the applications that use the correct or nearly correct net quantities for the categories in the prediction map. A similar type of relationship exists for the figure of merit versus the observed total change; although with less fit than for the observed net change. We could not find other strong relationships with prediction accuracy when we considered many possible explanatory factors including those in Tables 1 and 2.

The Appendix describes how the calibration procedures for LTM, CLUE-S, and CLUE use the correct net change for each category based on the reference map of time 2, so assessment of these applications should focus on location disagreement only. We compared the predictions for the applications that involve LTM and CLUE-S to a prediction where the correct quantity of net change is distributed at random locations. The LUCC model's accuracy is greater than the accuracy of a random spatial-allocation model for all five such applications, which are Detroit, Twin Cities, Maroua, Kuala Lumpur, and Haidian.

## 4 Discussion

This paper's approach is a good place to begin the evaluation of a model's results for a variety of reasons. This paper uses generally applicable quantitative measurements, so it can facilitate cross case comparison. It requires only three maps that are always available for any application that predicts change between points in time. It encourages scientific rigor because it asks the investigators to expose the degree to which calibration information is separated from validation information. It examines both the behavior of the model and the dynamics of the landscape, so it gives a baseline of a null model that is specific to each landscape. It produces statistics that allows for the extrapolation of the level of certainty into the future (Pontius and Spencer 2005; Pontius et al. 2006). The validation method budgets the reason for model errors as either quantity disagreement or location disagreement at multiple resolutions (Figs. 2, 5), so modelers can consider how to address each type of error when revising the models.

One of the most important general lessons is that the selection of the place, time, and format of the data must be taken into consideration when interpreting the model's performance, because these characteristics can have profound influence on the modeling results. The same model can behave differently in different settings as demonstrated by the applications for LTM, CLUE-S, and CLUE. Even for the applications where two models were used to predict change in The Netherlands from 1996 to 2000, the underlying data were formatted differently so it is not obvious whether the differences in results between the Land Use Scanner and Environment Explorer applications are due to the differences in the models or in the data. Consequently, model assessment must focus primarily on the performance of each model relative to its own data and its own null model, and then secondarily in relation to other data and other models. Even if the goal of this exercise was to rank the models according to predictive power, it would be impossible given the information in this article, because each model is applied to different data, and the data have a large influence on the results.

Figure 6 illustrates this point. We hypothesize that reference maps that show larger amounts of net change offer a model's calibration procedure a stronger statistical signal of change to detect and to predict, whereas location changes of simultaneous gains and losses of land categories are more challenging to predict.

Most LUCC models performed more accurately than their null models of persistence at coarse resolutions. This is also true at the fine resolution of the raw data in 7 of the 13 applications. Models that performed most accurately with respect to their null models either used the correct quantities of the categories in time 2 and/or predicted less than the amount of observed change. Only two applications predicted more net change than the observed net change, and these applications were least accurate with respect to their null models. This shows how if the model predicts change, then it risks predicting it incorrectly; while if the model predicts very little change, then it cannot make very much of that type of error.

Most applications used some information subsequent to time 2 to simulate the change between time 1 and time 2. Therefore many of the results reflect the goodness-of-fit of a mix of both calibration and validation.

Before revising the models, modelers should consider the size of modeling errors with respect to the accuracy of the reference maps. Some co-authors suspect substantial error in their reference maps due to a variety of reasons including errors in georeferencing and classification. Scientists must be cognizant that the differences between the reference maps of the initial and subsequent times can be due to both land change and map error (Pontius and Lippitt 2006). It would be folly to revise the model in order to make it conform to erroneous data.

There are an infinite number of other concepts and techniques that one could consider for the evaluation of a model (Batty and Torrens 2005; Brown et al. 2005). The methods of this paper constitute obvious first steps that are helpful to set the context to interpret more elaborate techniques of model assessment, if those more complex methods are desired.

Many scientists would like for models to be able to simulate accurately various possible dynamics according to numerous alternative scenarios, in which case the model's underlying mechanisms would need to be valid under a wide range of circumstances. This paper does not compare those underlying mechanisms in a quantitative manner. However, even when analysis of alternative scenarios is the goal, scientists should still be interested the model's ability to simulate the single historical scenario that actually occurred according to empirical data, because if the model cannot simulate the observed scenario correctly, then at least some of the model's underlying mechanisms must be wrong. Furthermore, it is reasonable to think that models would be more accurate in simulating the historical observed scenario with which humans have had direct experience, than in simulating alternative scenarios with which humans have not yet had direct experience. The results of this paper show that if we are to have trust in such models, then land change modelers have much work ahead. Our intention is that the results of this paper be used to forge a research agenda that illuminates a productive path forward.

## 5 Conclusions

Twelve of the 13 LUCC modeling applications in this paper's comparison contain more erroneous pixels than pixels of correctly predicted land change at the fine resolution of the raw data. Multiple resolution analysis reveals that these errors vanish at coarser resolutions, since near errors of location over small distances become resolved as resolution becomes slightly coarser. The most synthetic result is that LUCC models that are applied to landscapes that have larger amounts of observed net change tend to have higher rates of predictive accuracy as indicated by the figure of merit, for the voluntary sample of applications that we analyzed. This underscores: (1) the necessity of considering both the observed change and the predicted change in order to interpret the model error, and (2) the importance of characterizing the map differences in terms of quantity disagreement and location disagreement. As scientists continue to develop this rapidly growing field of LUCC modeling, it is essential that we communicate in ways that facilitate cross laboratory comparison. Therefore, we encourage scientists to use the concepts and techniques of this paper in order to communicate with a common language that is scientifically rigorous, generally applicable, and intellectually accessible.

## Appendix

This appendix describes each modeling application in order of conceptual similarity, based in part on the characteristics in Tables 1 and 2. Each of the following 13 subsections contains three paragraphs for each application. The first paragraph describes the process of land transformation, which the observed bar in Fig. 2 characterizes. The second paragraph describes the behavior of the model, which produces the predicted bar of Fig. 2. The third paragraph interprets the error bar in Fig. 2 by considering the observed and predicted changes.

### A.1 Worcester, USA with Geomod

The City of Worcester, located in Central Massachusetts, USA, is the third largest city in New England, and has been experiencing substantial land transition from forest to residential since the 1950s. Proximity to Boston combined with construction of roads has caused tremendous growth in a sprawling housing pattern.

Geomod is a LUCC model designed to simulate a one-way transition from one category to one other category (Pontius et al. 2001; Pontius and Malanson 2005; Pontius and Spencer 2005). Geomod uses linear interpolation of the quantity of built area between 1951 and 1971 in order to extrapolate linearly the net increase in quantity of built area between 1971 and 1999. It then distributes that net change spatially among the pixels that are non-built in 1971 according to the largest relative suitability as specified in a suitability map. Geomod generates the suitability map empirically by computing the relationship between the reference map of 1971 and independent variables that predate 1971, hence Geomod places additional built areas at locations that are generally flat and relatively sandy.

For this application of Geomod to Worcester, most of the error is location disagreement that derives from the model's inability to specify the location of the gain in built area. The quantity error derives from the model's prediction of a larger net increase in built area than was observed.

### A.2 Santa Barbara, USA with SLEUTH

The city of Santa Barbara and the town of Goleta near the Pacific Ocean have experienced radical change over the last 10–15 years, producing a landscape that is essentially built out. Transitions among rangeland, agriculture, and urban account for 97% of the observed difference between the reference maps of 1986 and 1998.

SLEUTH (2005) is a shareware cellular automata model of urban growth and land use change. For model calibration and extrapolation, SLEUTH uses data for the variables denoted in the letters of name of the model, which are: Slope; Land use of 1975 and 1986; Excluded areas of 1998; Urban extent of 1954, 1965, 1975, and 1986; Transportation; and Hillshade. The SLEUTH model was calibrated using four different methods: the traditional brute force method (Silva and Clarke 2002), a full resolution brute force method (Dietzel and Clarke 2004), a genetic algorithm (Goldstein 2004), and a randomized parameter search. There are substantial differences in the calibration algorithms, while there are not glaring differences in the resulting prediction maps for this application to Santa Barbara, so this article presents the results for only the genetic algorithm. Ongoing research demonstrates that the model can over-fit the data, leading to a prediction of less change than observed.

For this application of SLEUTH to Santa Barbara, the error is nearly equal to the observed change because the predicted change is very small. Most of the error is quantity disagreement because most of the observed change is net change, which is associated with gain in urban.

### A.3 Holland of eight categories with Land Use Scanner

The Netherlands (i.e., Holland, for short) has experienced increased population and prosperity in recent decades, which have caused a steady increase in area dedicated to residence, business, recreation, and infrastructure. Between 1996 and 2000, the largest observed transitions have involved the loss of agricultural land.

Land Use Scanner (2005) is a GIS-based model that uses a logit model and expert opinion to simulate future land use patterns (Koomen et al. 2005; Hilferink and Rietveld 1999; Schotten et al. 2001). The expected quantities of changes are based on a linear extrapolation of the national trend in land use statistics from 1981 to 1996. The regional demand for each land use is allocated to individual pixels based on suitability. Suitability maps are generated for all different land uses based on physical properties, operative policies, relations to nearby land-use functions, and expert judgment. The model uses data in which each pixel possesses a specific proportion of 36 possible categories. For this paper's map comparisons, the data for Holland(8) have been aggregated and simplified such that each pixel portrays exactly one of eight major categories.

For this application of Land Use Scanner to Holland, the error is equally distributed between location disagreement and quantity disagreement. There is more quantity disagreement in the prediction error than in the observed change, so the LUCC model is less accurate than its null model at all resolutions, as noted by the word "Worse" for the null resolution in Table 2. A large portion of this apparent error is attributable to the reformatting of each originally heterogeneous pixel into its single dominant category.

### A.4 Holland of 15 categories with Environment Explorer

The Holland(15) and Holland(8) applications both analyze The Netherlands from 1996 to 2000 on the same 500 m grid. In spite of this, the data for each application are different. Whereas the data for Holland(8) show 8 categories, the data for Holland(15) show 15 categories. For the 15-category data, much of the observed change is attributable to simultaneous loss of agriculture in some locations and gain of agriculture in other locations.

Environment Explorer (2005) is a dynamic cellular automata model, which consists of three spatial levels (de Nijs et al. 2004; Engelen et al. 2003; Verburg et al. 2004).

At the national level, the model combines countrywide economic and demographic scenarios, and distributes them at the regional level. The regional level uses a dynamic spatial interaction model to calculate the number of inhabitants and number of jobs over 40 regions, and then proceeds to model the land-use demands. Allocation of the land-use demands on the 500 m grid is determined by a weighted sum of the maps of zoning, suitability, accessibility, and neighborhood potential. Semi-automatic routines use the observed land use of 1996 for calibration.

For this application of Environment Explorer to Holland, most of the errors are location disagreement over small distances. There is more total error than total observed change, so the null model is more accurate than the LUCC model at the resolution of the raw data. If the near location disagreement is ignored, then the LUCC model is more accurate than its null model.

### A.5 Perinet, Madagascar with Logistic Regression

Perinet is a station on the railway line that links Madagascar's highland capital, Antananarivo, with the island's main seaport, Tamatave. The initial land cover is presumed to have been continuous forest, and the overwhelming proximate cause of deforestation is hypothesized to be conversion to agriculture via the Betsimisaraka production system, which does not often lead to abandonment and forest regrowth in this region (McConnell et al. 2004).

The deforestation process was modeled using binary logistic regression. The model is calibrated using land cover of 1957 as the dependent variable; independent variables are elevation and distance from settlements of 1957. The regression equation associates larger fitted probabilities of non-forest with lower elevations and nearness to villages. The map of fitted probabilities is reclassified into a Boolean prediction map by applying a threshold that selects the forested pixels that have the highest probability for deforestation. In order to determine the threshold, the quantity of predicted deforestation was computed based on published deforestation estimates for the first half of the twentieth century (Jarosz 1993). The model predicts a one-way transition from forest to non-forest and does not attempt to predict forest regrowth, so the non-forest of 1957 is eliminated from the assessment of the observed change, predicted change, and prediction error.

For the application of logistic regression to Perinet, a small portion of the error is quantity disagreement because the model predicted fairly accurately the observed net loss of forest. The model error is less than the observed change at all resolutions, so the LUCC model is more accurate than its null model at all resolutions (Fig. 5).

### A.6 Cho Don, Vietnam with SAMBA

Cho Don District is in a mountainous area of northern Vietnam. This region, like the rest of Vietnam, underwent major economic reforms in the 1980s that marked the shift from socialist centrally-planned agriculture to market family-based agriculture. Forest and shrub categories account for 96% of the difference between the maps of 1995 and

2001. The largest transitions are the exchanges between forest and shrub; in addition, both forest and shrub gain from upland cropland.

SAMBA (2005) is an agent-based modeling framework. The SAMBA team developed a number of scenarios that were discussed by scientists and local stakeholders as part of a negotiation platform on natural resources management through a participatory process combining role-play gaming and agent-based modeling (Boissau and Castella 2003; Castella et al. 2005a,b). The model is parameterized according to local specificities, e.g., soil, climate, livestock, population, ethnicity, and gender. Interviews during 2000 and 2001 serve as the basis for information concerning a variety of influential factors. The model uses information from post-1990 to simulate land change, so the assessment of the results for the modeling run from 1990 to 2001 should be interpreted as an analysis of the goodness-of-fit for a combination of calibration and validation.

For this application of SAMBA to Cho Don, most of the error is location disagreement over small distances due in part to the fact that near location disagreement characterizes most of the observed and predicted changes. Quantity disagreement in the error is small because the model predicts the nearly correct amounts of net changes in the categories. The error of the LUCC model is less than the observed change, so the LUCC model is more accurate than its null model at all resolutions (Fig. 5).

### A.7 Detroit, USA with Land Transformation Model

The Detroit Metropolitan Area (DMA) application and the Twin Cities Metropolitan Area (TCMA) application, in the next subsection, share a variety of characteristics. For example, both are analyzed with the Land Transformation Model (LTM), both are in the Upper Midwest United States, and both are composed of seven counties. These multi-county regional governmental organizations coordinate planning, transportation, education, environment, community, and economy. The DMA had over 4.7 million residents in 1980 and nearly 4.9 million in 2000. The reference maps show that 6% of the area available for new urbanization in 1978 became urban by 1995 (Fig. 2).

Land Transformation Model (2005) uses artificial neural networks to simulate land change (Pijanowski et al. 2000, 2002, 2005). The neural net trains on an input–output relationship until it obtains a satisfactory fit between the data concerning urban growth and the independent variables. The independent variables for the applications to both the DMA and the TCMA are elevation and distance to highways, streets, lakes, rivers, and the urban center. The DMA application and the TCMA application separate calibration data from validation data spatially by exchanging calibration parameters between the study areas. Specifically, the neural net obtains parameters by fitting a relationship between the independent variables and the urbanization in TCMA from 1991 to 1997. This relationship is then used to predict the urbanization in DMA from 1978 to 1995. The procedure generates a map of real numbers ranging between 0 and 1 that indicate relative propensity for urbanization in DMA. This map is then reclassified into a Boolean prediction map that shows urban growth versus no urban growth, such that the number of pixels of predicted urban growth matches the number of pixels of observed urban growth based on the reference map of time 2, as the ninth column of Table 2 shows. Consequently, the prediction has no error due to quantity by design, and thus the prediction accuracy indicates the fit in terms of location only. The applications for DMA and TCMA focus on the one-way transition from non-urban to urban, so urban pixels in the reference map of time 1 are eliminated from the bars in Fig. 2, similar to the Perinet application, as the fifth column of Table 2 shows.

For this application of the LTM to Detroit, the error has no quantity disagreement because the model uses the correct net change based on the time 2 reference map. The modeling application is less accurate than its corresponding null model, as the far location disagreement in the error is greater than the amount of observed change.

### A.8 Twin Cities, USA with Land Transformation Model

The Twin Cities Metropolitan Area (TCMA) is a region that contains the neighboring cities of Minneapolis and Saint Paul in the state of Minnesota. The TCMA contained over 2.2 million residents in 1990 and over 2.6 million in 2000. The reference maps show that 4% of the area available for new urbanization in 1991 became urban by 1997.

As mentioned in the previous subsection, the LTM obtains a relationship between the independent variables and urban growth by presenting the data for the DMA to the neural net, then the DMA relationship is used to predict urban growth for the TCMA application. The fitted relationship generates a map of real numbers that show the relative propensity for urban growth in TCMA. This propensity map is then reclassified to create a Boolean prediction map of urban growth versus no urban growth, such that the number of pixels of predicted urban growth matches the number of pixels of observed urban growth according to the two reference maps of TCMA.

For this application of the LTM to Detroit, the error has no quantity disagreement by design. The total error is greater than the observed change, while the far location disagreement in the error is less than the observed change. So, if we ignore the near location disagreement, then this application is more accurate than its null model.

### A.9 Maroua, Cameroon with CLUE-S

The Maroua study area is in northern Cameroon, and is representative of the Sudano-Sahelian savannah zone. The center of the study area is the urban centre of Maroua, which has an important influence in the region as increasing population has induced changes in land use. Two particular transitions, from bush to rain crops and from bush to sorghum, account for about half of the observed difference in the reference maps between 1987 and 1999.

CLUE-S (2005) is a fundamentally revised version of the model called Conversion of Land Use and its Effects (CLUE). CLUE-S is designed to work with fine resolution data where each pixel represents a single dominant land use, rather than a heterogeneous mix of various categories as in the original CLUE model (Verburg et al. 2002; Verburg and Veldkamp 2004). CLUE-S consists of two main components. The first component supports a multi-scale spatially-explicit methodology to quantify empirical relationships between land-use patterns and their driving forces. The second component uses the results from the first component in a dynamic simulation technique to explore changes in land use under various scenarios. A combination of expert knowledge and empirical analysis usually serve for calibration. A user of CLUE-S can specify any quantity of land change based on various sectoral models. For the three CLUE-S applications described in this paper, the calibration is based on the single reference map of time 1 due to lack of time series data on land cover, so it is impossible to use historic information to predict the quantity of each category for time 2. Therefore, CLUE-S sets the simulated quantity of each category for time 2 to be equal to the correct quantity as observed in the reference map for time 2. While CLUE-S uses the correct total quantity of each category for time 2, it must predict exactly how the correct net change from time 1 derives from a wide variety of possible combinations of gross gains and gross losses for numerous categories.

For the application of CLUE-S to Maroua, there is no quantity disagreement in the error by design. The total error is less than the observed change, so the LUCC model is more accurate than its null model at all resolutions (Fig. 5). Most of the error is near location disagreement. If near location disagreement is ignored, then the LUCC model is much more accurate than its null model. CLUE-S produces similar results for its other two applications, so we do not elaborate on the error in the next two subsections.

### A.10 Kuala Lumpur, Maylasia with CLUE-S

The Klang-Langat Watershed is located in the mid-western part of Peninsular Malaysia and is the most highly urbanized region of the country. The northern part of the watershed contains Kuala Lumpur, the capital city of about 1.5 million people, whereas the entire region has about 4.2 million inhabitants. The largest observed changes involve a net gain in urban and a location change in agriculture. The single transition from agriculture to urban accounts for 42% of the observed difference in the reference maps.

CLUE-S uses 13 independent variables for the Kuala Lumpur application. These are: elevation, slope, geology, soils, erosion sensitivity, forest protection zones, other protected areas, distance to the coast, and travel time to highways, roads, sawmills, important towns, and other towns.

### A.11 Haidian, China with CLUE-S

Haidian is a district of Beijing, China, where urbanization is sprawling on the best quality agricultural land. These changes are part of a larger process of urban sprawl in the periphery of Beijing where the area of urban land has doubled between 1990 and 2000 (Tan et al. 2005). Change that involves the industrial category accounts for 47% of the observed difference, as industrial land simultaneously loses to urban and gains from arable. Changes that involve arable land are nearly equally prominent, as arable loses to both industrial and forest.

Independent variables include: elevation, slope, soil texture, soil thickness, agricultural income, agricultural population, travel time to central city, travel time to nearest village, distance to village, distance to various of types of roads, and the government's land allocation plans (Duan et al. 2004). The inclusion of the government's land allocation plans is apparently the factor that enables accurate prediction of nearly every pixel of the irregularly shaped patches of forest gain and forest loss.

### A.12 Honduras with CLUE

The process of land change in Honduras has been influenced by high pressures from human population, large gradients in both topography and climate, and economic instability after the second oil crisis of the 1970s, which led to the initiation of extensive land redistribution programs. These factors have caused substantial deforestation. The land change is distributed equally between changes in quantity and changes in location of categories. The net changes in quantity are attributable primarily to a transition from nature to pasture. The changes in location are attributable primarily to the gain of pasture in some locations and the loss of pasture in other locations. The pixels for the CLUE applications to Honduras and Costa Rica show proportions for multiple land categories; thus they have a format distinctive from the other eleven applications.

CLUE (2005) is a spatially-explicit, multi-scale model that projects land-use change (Kok and Veldkamp 2001; Veldkamp and Fresco 1996; Verburg et al. 1999). CLUE is the predecessor of CLUE-S, so the two models share many philosophical approaches and computational features. Yearly changes are allocated in a spatially explicit manner in the grid-based allocation module, which consists of a two-step top-down iteration procedure with bottom-up feedbacks. The two CLUE applications set the predicted quantity of each category to be equal to the correct quantity for each category as shown in the reference map for time 2, similar to the CLUE-S and LTM applications (Table 2). Given this information, CLUE must predict how the correct net quantity for each category derives from possible combinations of gross gains and gross losses. In addition, CLUE predicts the location of various land-use transitions. CLUE predicts some of the dynamics by extrapolating linearly the pre-1974 trends in the population census data. The model separates calibration information from validation information for Honduras by using parameters derived from analyses of Costa Rica.

For the application of CLUE to Honduras, there is no quantity disagreement in the error by design, just as in the applications of CLUE-S and LTM. Nearly all of the error is near location disagreement. The total error is less than the observed change so the LUCC model is more accurate than its null model. CLUE produces similar results for its application to Costa Rica.

### A.13 Costa Rica with CLUE

Costa Rica and Honduras share many processes of land change due to similarities with respect to population pressures, oil crises, and geophysical characteristics. However, Costa Rica has some distinctive aspects. In particular, the Costa Rican government bought large tracts of land between 1960 and 1990 with the objective to stimulate smallholder development. This caused large demographic movements from west to east. Ninety-one percent of the difference in the reference maps between 1973 and 1984 is location change as the pasture category shifted location from west to east, resulting in a loss of the nature category in the east and a gain of the nature category in the west.

The CLUE application to Costa Rica calibrates its parameters with some data that reflect an application to Ecuador (de Koning et al. 1999). Other aspects of the calibration information for Costa Rica reflect the influence of the post-1973 land reforms, which would not have been predicted by an extrapolation of pre-1973 trends. If the Costa Rica application were to have assumed less influence by the land reforms, then the prediction map of 1984 would probably agree less with the reference map of 1984, because the historical process of land change before 1973 was fundamentally different than the process during the prediction interval from 1973 to 1984.

**Acknowledgments** The C. T. DeWit Graduate School for Production Ecology& Resource Conservation of Wageningen University sponsored the first author's sabbatical, during which he led the collaborative exercise that is the basis for this article. The National Science Foundation of the USA supported this work via the grant "Infrastructure to Develop a Human-Environment Regional Observatory (HERO) Network" (Award ID 9978052). Clark Labs has made the building blocks of this analysis available in the GIS software, Idrisi® .

## References

Batty M, Torrens PM (2005) Modeling and prediction in a complex world. Futures 37(7):745–766

Boissau S, Castella J-C (2003) Constructing a common representation of local institutions and land use systems through simulation-gaming and multi-agent modeling in rural areas of Northern Vietnam: the SAMBA-Week methodology. Simulat Gaming 34(3):342–347

Brown DG, Page S, Riolo R, Zellner M, Rand W (2005) Path dependence and the validation of agent-based spatial models of land use. Int J Geogr Informati Sci 19(2):153–174

Castella J-C, Boissau S, Trung TN, Quang DD (2005a) Agrarian transition and lowland-upland interactions in mountain areas in northern Vietnam: application of a multi-agent simulation model. Agric Syst 86(3):312–332

Castella J-C, Trung TN, Boissau S (2005b) Participatory simulation of land use changes in the Northern Mountains of Vietnam: the combined use of an agent-based model, a role-playing game, and a geographic information system. Ecol Soc 10(1):27

de Koning GHJ, Verburg PH, Veldkamp TA, Fresco LO (1999) Multi-scale modelling of land use change dynamics in Ecuador. Agric Syst 61:77–93

de Nijs TCM, de Niet R, Crommentuijn L (2004) Constructing land-use maps of the Netherlands in 2030. J Environ Manag 72(1–2):35–42

Dietzel CK, Clarke KC (2004) Spatial differences in multi-resolution urban automata modeling. Trans GIS 8:479–492

Duan Z, Verburg PH, Fengrong Z, Zhengrong Y (2004) Construction of a land-use change simulation model and its application in Haidian District, Beijing. Acta Geogra Sinica 59(6):1037–1046 (in Chinese)

Engelen G, White R, de Nijs T (2003) The Environment Explorer: spatial support system for integrated assessment of socio-economic and environmental policies in the Netherlands. Integr Assess 4(2):97–105

Goldstein NC (2004) Brains vs. Brawn — comparative strategies for the calibration of a cellular automata-based urban growth model. In: Atkinson P, Foody G, Darby S, Wu F (eds) GeoDynamics, pp 249–272 CRC, Boca Raton

Hilferink M, Rietveld P (1999) Land Use Scanner: an integrated GIS based model for long term projections of land use in urban and rural areas. J Geogr Syst 1(2):155–177
Jarosz L (1993) Defining and explaining tropical deforestation: shifting cultivation and population growth in colonial Madagascar (1896–1940). Econ Geogr 69(4):366–379
Klug W, Graziani G, Grippa G, Pierce D, Tassone C (eds) (1992) Evaluation of long range atmospheric transport models using environmental radioactivity data from the Chernobyl accident: the ATMES Report. Elsevier, London, 366 pages
Kok K, Veldkamp TA (2001) Evaluating impact of spatial scales on land use pattern analysis in Central America. Agric Ecosyst Environ 85(1–3):205–221
Kok K, Farrow A, Veldkamp TA, Verburg PH (2001) A method and application of multi-scale validation in spatial land use models. Agric Ecosyst Environ 85(1–3):223–238
Koomen E, Kuhlman T, Groen J, Bouwman A (2005) Simulating the future of agricultural land use in the Netherlands. Tijdschrift voor economische en sociale geografie 96(2):218–224 (in Dutch)
McConnell W, Sweeney SP, Mulley B (2004) Physical and social access to land: spatio-temporal patterns of agricultural expansion in Madagascar. Agric Ecosyst Environ 101(2–3):171–184
Perica S, Foufoula-Georgiou E (1996) Model for multiscale disaggregation of spatial rainfall based on coupling meteorological and scaling descriptions. J Geophys Res 101(D21) 26347–26361
Pijanowski BC, Gage SH, Long DT (2000) A Land Transformation Model: integrating policy, socioeconomics and environmental drivers using a geographic information system. In: Harris L, Sanderson J (eds) *Landscape ecology: a top down approach,* pp 183–198 CRC Press, Boca Raton
Pijanowski BC, Brown DG, Manik G, Shellito B (2002) Using neural nets and GIS to forecast land use changes: a land transformation model. Computers Environ Urban Syst 26(6):553–575
Pijanowski BC, Pithadia S, Shellito BA, Alexandridis K (2005) Calibrating a neural network-based urban change model for two metropolitan areas of the Upper Midwest of the United States. Int J Geogr Informat Sci 19(2):197–215
Pontius Jr RG (2000) Quantification error versus location error in comparison of categorical maps. Photogramm Eng Remote Sensing 66(8):1011–1016
Pontius Jr RG (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
Pontius Jr RG, Lippitt CD (2006) Can error explain map differences over time? Cartogr Geogr Informat Sci 33(2):159–171
Pontius Jr RG, Malanson J (2005) Comparison of the structure and accuracy of two land change models. Int J Geogr Informat Sci 19(2):243–265
Pontius Jr RG, Spencer J (2005) Uncertainty in extrapolations of predictive land change models. Environ Plan B 32:211–230
Pontius Jr RG, Cornell J, Hall C (2001) Modeling the spatial pattern of land-use change with GEOMOD2: application and validation for Costa Rica. Agric Ecosyst Environ 85(1–3):191–203
Pontius Jr RG, Huffaker D, Denman K (2004a) Useful techniques of validation for spatially explicit land-change models. Ecol Model 179(4):445–461
Pontius Jr RG, Shusas E, McEachern M (2004b) Detecting important categorical land changes while accounting for persistence. Agric Ecosyst Environ 101(2–3):251–268
Pontius Jr RG, Cheuk ML (2006) A generalized cross-tabulation matrix for comparing soft-classified maps at multiple resolutions. Int J Geogr Informat Sci 20(1):1–30
Pontius Jr RG, Versluis AJ, Malizia NR (2006) Visualizing certainty of extrapolations from models of land change. Landscape Ecol 21(7):1151–1166
Schotten K, Goetgeluk R, Hilferink M, Rietveld P, Scholten H (2001) Residential construction, land use, and the environment: simulations for the Netherlands using a GIS-based land use model. Environ Model Assess 6:133–143
Silva EA, Clarke KC (2002) Calibration of the SLEUTH urban growth model for Lisbon and Porto, Portugal. Computers Environ Urban Syst 26:525–552
Tan M, Li X, Xie H, Lu C (2005) Urban land expansion and arable land loss in China — a case study of Beijing-Tianjin-Hebei region. Land Use Policy 22(3):187–196
Veldkamp AT, Fresco L (1996) CLUE-CR: an integrated multi-scale model to simulate land use change scenarios in Costa Rica. Ecol Model 91:231–248
Verburg PH, Veldkamp TA (2004) Projecting land use transitions at forest fringes in the Philippines at two spatial scales. Landscape Ecol 19:77–98

Verburg PH, de Koning FGHJ, Kok K, Veldkamp TA, Fresco LO, Bouma J (1999) A spatial explicit allocation procedure for modelling the pattern of land use change based upon actual land use. Ecol Model 116:45–61

Verburg PH, Soepboer W, Veldkamp TA, Limpiada R, Espaldon V, Sharifah Mastura SA (2002) Modeling the Spatial Dynamics of Regional Land Use: the CLUE-S Model. Environ Manage 30(3):391–405

Verburg PH, de Nijs TCM, van Eck JR, Visser H, de Jong K (2004) A method to analyse neighbourhood characteristics of land use patterns. Computers Environ Urban Syst 28(6):667–690