Lessons and challenges in land change modeling as revealed by map comparisons

Authors

Robert Gilmore Pontius Jr[1], Jean-Christophe Castella[2], Ton de Nijs[3], Zengqiang Duan[4], Eric Fotsing[5], Noah Goldstein[6], Kasper Kok[7], Eric Koomen[8], Christopher D. Lippitt[1], William McConnell[9], Alias Mohd Sood[10], Bryan Pijanowski[11], A. Tom Veldkamp[7], and Peter H. Verburg[7]

Affiliations

[1] Clark University, Worcester MA, USA

[2] Institut de Recherche pour le Développement Montpellier, France

[3] Netherlands Environmental Assessment Agency, Bilthoven, The Netherlands

[4] China Agricultural University, Beijing, China

[5] University of Dschang, Maroua, Cameroon

[1] University of California, Santa Barbara, CA, USA

[7] Wageningen University, Wageningen, The Netherlands

[8] Vrije Universiteit, Amsterdam, The Netherlands

[9] Michigan State University, East Lansing, MI, USA

[10] Forestry Department Peninsular Malaysia, Kuala Lumpur, Malaysia

[11] Purdue University, West Lafayette, IN, USA

Correspondence

Robert Gilmore Pontius Jr (rpontius@clarku.edu)

Abstract

This paper presents the most important lessons from a multi-year collaboration that compared thirteen cases of spatially-explicit land change modeling. A previous paper reports the statistical results of the validation exercise, while this paper offers the broader implications of those findings for the land use and land cover change modeling community. We express the lessons as nine challenges grouped under three themes: mapping, modeling, and learning. The mapping challenges are: to prepare data appropriately, to select relevant resolutions, and to differentiate types of land change. The modeling challenges are: to separate calibration from validation, to predict small amounts of change, and to interpret the influence of quantity error. The learning challenges are: to use appropriate map comparison measurements, to learn about land change processes, and to collaborate openly. The paper elaborates on why these challenges are especially important for the future research agenda in land change science.

Keywords

accuracy, land cover, land use, model, prediction, scale, validation.

1 INTRODUCTION

1.1 Collaborative exercise

In 2004, the first author of this paper extended an open invitation to the community of land change modelers to participate in a cross-case comparison of spatially explicit land change modeling applications. The focus was on the assessment of the validation of such models, so the invitation requested that participants submit three maps of land categories for any case study site: 1) a reference map of an initial time that a land change model used for calibration, 2) a reference map of a subsequent time that could be used for validation, and 3) a prediction map of the subsequent time that the land change model produced. Ultimately, we compiled thirteen different cases from around the world, which were submitted from seven different laboratories. Pontius et al. (in press) applied metrics to compare those various cases in a manner that takes into consideration the differences in study areas and data formats. This work has been presented at several scientific conferences. The most common initial reaction that audiences have when they hear about this exercise is to ask ‘Which model is best?’ This reaction has been one of the inspirations for this paper, since the popularity of the question indicates that we must be careful to interpret the results properly, because the purpose of the exercise can be easily misinterpreted. We have found that the presentation of the methods and results inspires quite disparate conclusions from different scientists. The fact is we did not intend for the exercise to rank the models. The purpose of the exercise was to gain insight into the scientific process of modeling, so we can learn the most from our modeling efforts. Therefore, this paper shares the most important lessons from the cross-case comparison.

1.2 Mapping, modeling, and learning about land change

We find it helpful to think of these lessons in terms of the flows and feedbacks of information among the various components and procedures for a systematic analysis as shown in figure 1. The figure begins with the real landscape in the upper left corner. The landscape gives rise to data via a mapping procedure in which the modeler usually participates. There is a tremendous amount of information that the scientist can derive from simply analyzing the data using methods of map characterization for a map from a single point in time and methods of map comparison for maps from two points in time. Modelers anticipate that they can learn even more by engaging in a more elaborate modeling procedure that produces a dynamic model of land change. The modeler usually uses a conceptual understanding of landscape dynamics to guide the production of algorithms that attempt to expresses those dynamics. This paper uses the word ‘model’ to refer to such a set of algorithms, whereas the word ‘case’ refers to an application of the model to a one particular study site. One important way to assess a case is to examine the output that the model produces. Ultimately, a major purpose of the analysis is for the modeler to learn valuable information from the measurements of the data and the outputs from the model. The modeler can use this learning to revise the mapping, to revise the modeling and/or to revise the methods for measurement of the data and model output. The components of figure 1 reflect the structure of this paper in that this paper’s

‘Methods’ section summarizes the techniques to measure both the data and the model’s output, while the subsequent ‘Results & Discussion’ section presents the most important lessons, organized under the themes of mapping, modeling, and learning.

[Insert figure 1 here].

2 METHODS

All the modeling techniques have been published in peer-reviewed journals and books, and the maps have been submitted by scientists from the laboratories that actually developed the models. Collectively, the sample of models covers a range of some of the most common modeling techniques such as statistical regression, cellular automata, and machine learning. SAMBA is the single agent-based model in the collection. Table 1 offers some important specific characteristics of the nine models used for the thirteen cases. These characteristics are important for interpreting the model’s output. Geomod, Logistic Regression, and Land Transformation Model (LTM) use maps of two categories to predict a single transition from exactly one losing category to exactly one gaining category. The other six models use maps of multiple categories to predict multiple transitions, where any category can simultaneously gain in some locations and lose in other locations. For seven of the models, the user can exogenously set the quantity of each land cover category in the predicted map, and then the model predicts the location of the land categories. SLEUTH and SAMBA do not have this characteristic. This paper’s cases that derive from LTM, CLUE-S, and CLUE use the quantity of each category in the map of the subsequent time as input to the model, so the model is assured to simulate the correct quantity of each category, thus the purpose of the modeling application is to predict only the location of change. Most of the models are designed to use pixels that are categorized as exactly one category, while Land Use Scanner, Environment Explorer and CLUE can use heterogeneous mixed pixels for both input and output.

Both Land Use Scanner and Environment Explorer are applied to the entire country of The Netherlands, so we can see two different modeling approaches to the same study area. One substantial difference between these two cases is that the number of categories in the output map for the application of Land Use Scanner is eight, while the number of categories is fifteen for the application of Environment Explorer. LTM, CLUE-S, and CLUE are applied to more than one study area, which allows us to see some variation of how a single model can behave in different case studies.

[Insert table 1 here].

Figure 2 shows the mapped results for each of the thirteen cases. Each map in figure 2 derives from an overlay of the three maps that a modeler submitted. The first eleven of the thirteen cases share the same legend, while Costa Rica and Honduras have a different legend because those two cases have large mixed pixels. The maps in figure 2 can be used to see the places where the data indicate change and the places where the model predicts change. The intersection of these two denotes the places where the model predicts change correctly. For the nine cases that predict simultaneous gains and losses among multiple categories, we can have a situation where the data show change and the model predicts change, but the model predicts the change as the wrong gaining category. There is persistence according to both the data and the prediction on more than 40 percent of the study area for all of the cases.

[Insert figure 2 here].

Figure 3 summarizes some important results, where a segmented bar quantifies each case in terms of the legend of figure 2. Each bar is a type of rectangular Venn diagram that shows the intersection of the observed change and the predicted change, as shown explicitly for the case of Perinet. The 'figure of merit' is a summary statistic that is defined as a ratio, where the numerator is the length of the bar segment for 'correct due to observed change predicted as change' and the denominator is the length of the entire bar. If the model's prediction were perfect, then there would be perfect intersection between the observed change and the predicted change, and so the figure of merit would be 100 percent. If there were no intersection between the observed change and the predicted change, then the figure of merit would be zero. Figure 3 orders the cases in terms of the figure of merit, which is given as a percent at the right of each bar.

[Insert figure 3 here].

It is also helpful to consider a null model for each case. The definition of this null model is a prediction of complete persistence, i.e. no change, between the initial and the subsequent time, therefore the accuracy of the null model is 100% minus the amount of observed change. Figure 3 shows that the accuracy of the land change model exceeds the accuracy of its corresponding null model for seven of the thirteen cases at the resolution of the raw data (Pontius et al. 2004a).

Figure 4 plots the predictive accuracy, as indicated by the figure of merit, versus the amount of observed change, as indicated by reference maps, for each case. Figure 4 reveals two clusters. The particularly tight cluster near the origin shows that all of the cases that have a figure of merit less than 15% also have an observed change less than 10%. We analyzed many factors that we suspected might explain the predictive power for these cases and found that the amount of change in the reference maps had the strongest relationship with predictive accuracy (Pontius et al. in press).

[Insert figure 4 here].

We have been soliciting feedback on these results both formally and informally since 2004 in order to derive the most important lessons that the authors and others can take from this analysis. We have presented this work at four international scientific conferences: the 2004 Workshop on the Integrated Assessment of the Land System in Amsterdam The Netherlands, the 2005 Open Meeting of the Human Dimensions of Global Environmental Change Research Community in Bonn Germany, the 2006 Meeting of the Association of American Geographers in Chicago USA, the 2007 World Congress of the International Association for Landscape Ecology in Wageningen The Netherlands. There were panel discussions in Amsterdam, Chicago and Wageningen, where authors shared their experiences and audience members shared their reactions. We have analyzed the feedback from editors and anonymous reviewers of scientific journals. The co-authors have been reflecting on and communicating about this work since 2004. The next section of this paper synthesizes the most important lessons to date.

3 RESULTS & DISCUSSION

This section offers nine important lessons. Each lesson has implications concerning the agenda for future research; therefore each lesson corresponds to a sub-section heading that articulates an important challenge for future modeling efforts. These are grouped under three themes: mapping, modeling, and learning. These groupings emerged as the authors reflected on the various types of lessons learned. The first theme demonstrates that the selection of the study site and the production of the data have a substantial influence on the final results, so researchers need to pay as much attention to the mapping procedure as they do to the modeling procedure. This message reinforces known fundamental concepts in mapping, which must be kept at the front of the minds of modelers. The second theme concerns the modeling process. The challenges under this theme derive from insights that have emerged specifically as a result of this cross-case exercise. They have important implications for how scientists should design and assess modeling procedures. The third theme focuses on learning, thus it emphasizes the importance of careful reflection on the mapping and modeling procedures. This is important because if mapping and modeling are not interpreted properly, then one can exert a tremendous amount of time and energy without learning efficiently. This third theme contains ideas for how modelers can maximize learning from information produced by the mapping and modeling procedures.

3.1 Mapping challenges

3.1.1 To prepare data appropriately.

The decision concerning how to format the data is one of the most important decisions that a scientist makes. In some cases the scientists adopts the existing format of the available data, while in other cases the scientist purposely formats the data for the particular research project. Scientists should think carefully about the purpose of the modeling exercise when determining the format of the data. Important formatting decisions concern the spatial, temporal and categorical scales in terms of both extent and resolution. This is important because the apparent complexity of landscape change is a function of how a scientist chooses to envision it, which is usually reflected in a mapping procedure. If a scientist chooses a great level of detail, then any landscape can appear to be greatly complex; while if a scientist chooses less detail, then the same landscape can appear much simpler. For example, the Holland landscape is not inherently more complex than the Perinet landscape, however the Perinet data were formatted to show a one-way transition from forest to non-forest while the Holland(15) data were formatted to show multiple transitions among fifteen categories based on the data formatting decisions of the modelers. One could have attempted to analyze Holland as two categories of built versus non-built, and could have attempted to analyze the Perinet data as numerous categories of various types of uses and covers. For example, Laney (2002) chose to analyze land change in Madagascar at a much finer level of detail and higher level of complexity than McConnell et al. (2004). Anyone can choose a great level of detail in the data that will overwhelm any particular model. More detail does not necessarily lead to a better case study, just as less detail does not necessarily lead to a better case study. Scientists face the challenge to select a categorical scale, spatial resolution, spatial extent, temporal resolution, and temporal extent, for which a model can illuminate issues that are relevant for the particular purpose of the inquiry.

Decisions concerning the format and detail of the data are fundamental to understanding and evaluating the performance of the model (Dietzel and Clarke 2004). The Holland(8) case demonstrates this clearly as it relates to the reformatting from maps that describe many heterogeneous categories within each pixel to maps that describe the single dominant category within each pixel. The Land Use Scanner model was run for heterogeneous pixels of 36 categories, and then the output was reformatted to homogenous pixels of eight categories. This reformatting is common in the visualization of such mixed pixel data. A major drawback of this approach is that it can introduce substantial over representation of categories that tend to cover less than the entire pixel but more than any other category within the pixel. Consequently, it can also introduce substantial under representation of the other categories. Such biases substantially influenced the analysis of the performance of the Holland (8) case and caused the large error of quantity.

While the decisions concerning how to format the data are important, scientists lack clear guidelines on how to make such decisions. It makes sense to simplify the data to the level that the calibration procedure and validation procedure detect a meaningful signal of land change. It also makes sense to simplify the data so that the computer algorithms focus on only the important transitions among categories, where importance is related to the practical purpose of the model. Scientists who attempt to analyze all transitions among a large number of categories face substantial challenges. For the Santa Barbara, Holland(8), and Holland(15) cases, each particular transition from one category to another category in the reference maps occurs on less than 1 percent of the study area. Each of these individual transitions would need to have an extremely strong and unique relationship with the independent variables in order for any model to predict them accurately.

These decisions concerning the data are related closely to decisions concerning the level of complexity of the models. Models that simulate only a one-way transition from one category to one other category can be simpler than models that simulate all possible transitions among multiple categories. If scientists choose to analyze very detailed data, then they will be tempted or forced to use very complex models. It is not clear whether it is worthwhile to include great detail in the data and/or in the models, because it is not clear whether more detail leads to better information or to more error.

Modelers should consider the certainty of the data, since much of the apparent land change could be due to map error, especially when the observed difference in the reference maps is small (Pontius and Lippitt 2006). Participating scientists suspect that error accounts for a substantial amount of the observed difference in the reference maps for Maroua, Kuala Lumpur, and Holland(15). Scientists should use data for which there is more variation over time due to the dynamics of the landscape than due to map error. This can be quite a challenge in situations where map producers are satisfied with 85 percent accuracy, which implies up to 15 percent error, while many study areas show less than 15 percent land change.

3.1.2 To select relevant resolutions.

Spatial resolution is a component of data format that warrants special attention because: 1) it has a particularly important influence, 2) it is something that modelers usually can determine directly, and 3) it is not obvious how to select an appropriate resolution. The spatial resolution at which landscapes are modeled is often determined by data availability and computational capacity. For example, if the resolution and extent of a satellite image is selected as it was in the Maroua case, then the boundaries of the study area and the apparent unit of analysis are determined in part by the satellite imaging system, not necessarily by the theoretical or policy imperatives of the modeling exercise. Kok et al. (2001) argue that the selection of resolution should take into consideration the purpose of the modeling application and the scale of the LUCC process. For example, the Worcester case uses 30-meter resolution data, but few stakeholders in Worcester need a prediction of land change to be accurate to within 30 meters. Some stakeholders would like to know generally what an extrapolation of present trends would imply over the next decade to within a few kilometers, which is a resolution at which Geomod predicts better than a null model. Therefore, it is helpful from the standpoint of model performance to measure the accuracy of the prediction at resolutions coarser than the resolution of the raw data. Pontius et al. (in press) give results from a multiple-resolution analysis for each case. Multiple-resolution analysis shows that most of the errors are due to inaccurate spatial allocation over small distances for seven of the thirteen cases, and most of the error vanishes when the results are assessed at a resolution of 64 times coarser than the resolution of the original data. Errors of location vanish as resolution becomes coarser, but errors of quantity are effectively independent of resolution when assessed using an appropriate multiple-resolution method of map comparison (Pontius et al. 2004a).

3.1.3 To differentiate types of land change.

Modelers should select the types of land change that are of interest before deciding which model to use, because some types of land change present particular challenges for models. We find it useful to think of two major types of change: net and swap. Net change refers to a difference in the quantity of the categories in the reference maps between the initial and subsequent times, while swap change refers to a difference in only the location of the categories (Pontius et al. 2004b). The reference maps for Holland(15), Cho Don, Haidian, Honduras and Costa Rica demonstrate more swap than net change. In particular, nearly all the observed change in Costa Rica is attributable to differences in location, which means that quantity of each category at the initial time is about the same as it is at the subsequent time. When there is substantial swap in the observed data, the model must be able to predict simultaneous gains and losses for a category in order to predict the change accurately. This can be much more challenging than to predict a one-way transition from one category to one other category. The computer algorithm needs to be more elaborate in order to accommodate simultaneous gains and losses among multiple categories, compared to a single specific transition. For example, the Worcester, Perinet, Detroit, and Twin Cities cases use models that are designed to simulate net change of one category, while all the other cases use models that are designed to allow for simultaneous transitions among several categories.

3.2 Modeling challenges

3.2.1 To separate calibration from validation.

Nearly all of the cases used some information subsequent to the initial time 1 in order to predict the change between time 1 and time 2. In seven of the thirteen cases, the model's calibration procedure uses correct information directly from the reference map of time 2 concerning the quantity of each category. Other cases used influential variables, such as excluded areas, that derive from contemporary points in time subsequent to time 1. In these situations, it is impossible to determine whether the model's apparent accuracy indicates its predictive power. If a model uses information from both time 1 and time 2 for calibration, then we would think that the model's so called prediction map of time 2 could be a good match with the reference map of time 2 simply because the model parameters might be over fit to the data. The apparent accuracy would reflect a level of agreement higher than the level of agreement attributable to the model's actual predictive power.

There are some practical reasons why modelers use information subsequent to time 1 to predict the change between time 1 and time 2. Some reasons relate to the purpose of the model; other reasons relate to data availability.

This paper's cases that use LTM, CLUE-S and CLUE are cases that use information directly from the reference map of time 2 concerning the quantity of each category, because priority is accorded in these models to predicting the location of land change. The user can specify the quantity of each category independently from the location for these models, which can be an advantage allowing them to be used with tabular data and other types of models that generate non-spatial information concerning only the quantity of each land type (Pontius et al. 2003). For example, CLUE-S and CLUE can set the quantity of each category by using case-study and scale-specific methods ranging from trend extrapolations to complex sectoral models of world trade.

Some models such as SAMBA require information that is available for only post-time 1 periods. SAMBA is an agent-based modeling framework that used information from interviews with farmers concerning their land practices. For the Cho Don case, these interviews were conducted subsequent to both time 1 and time 2. Furthermore, the purpose of the SAMBA model is to explore scenarios with local stakeholders, not to predict exact land conversions. The SAMBA team has been developing other methods for validation of various aspects of their model (Castella et al. 2005b, Castella and Verburg 2007).

There are costs associated with separating calibration from validation information. The Worcester case accomplished separation between calibration information and validation information by restricting severely the use of independent variables. For example, maps of contemporary zoning and roads are available in digital form, but those maps contain some post-1971 information, therefore Geomod refrained from using those variables that are commonly associated with land change. Consequently, the Worcester case uses only slope and surficial geology as independent variables. Nevertheless, Pontius and Malanson (2005) show that there would not have been much measurable benefit in using the zoning map, since a zoning map shows the places where land change is unlikely to occur, not the few places where it is likely to occur.

3.2.2 To predict small amounts of change.

All thirteen of the cases have less than 50 percent observed change, seven of the cases show less than 10 percent observed change, while the Holland(8), Santa Barbara, and Twin Cities cases demonstrate less than 4 percent observed change on the landscape. Land change over a small time interval is usually a rare event, and rare events tend to be difficult to predict accurately. Figure 4 shows that a small amount of change on the landscape is associated with a low level of predictive accuracy.

The challenge to detect and to predict change is made even more difficult by insisting upon rigorous separation of calibration data from validation data in situations where data are scarce. For example, many models such as Environment Explorer are designed to examine the change during a calibration interval from time 0 to time 1, and then to predict the change during a validation interval from time 1 to time 2. The Holland(15) case separates calibration information from validation information using this technique, consequently the calibration interval and the validation interval both show a small amount of change since the calibration interval is only seven years and the validation interval is only four years. In such situations, statistical models may have difficulty in detecting a strong relationship between land change and the independent variables during the calibration phase, and the validation measurements may have difficulty in finding a strong relationship between the predicted land change and the observed land change. One solution would be for scientists to invest the necessary effort to digitize maps of historic land cover, so that they can have a larger time interval over which to examine land change.

3.2.3 To interpret the influence of quantity error.

Models that do not use the correct quantity of each category for time 2 must somehow predict the quantity for each category for time 2. Modelers need to be aware of how error in the prediction of quantity influences other parts of the validation process. Models typically fail to predict correct locations precisely; so models that predict too much change are likely to produce more errors than models that predict too little change, when assessed at fine spatial resolutions. For example, the Worcester case predicts more than the observed amount of change, which leads to substantial error. If the model were to predict less than the observed amount of change, then its output would probably have less error. In contrast, SLEUTH predicts substantially less than the amount of observed change for the Santa Barbara case, thus the accuracy is very close to that of a null model.

It does not make sense to use criteria that reward systematic underestimates or overestimates of the quantity of each category. This is a weakness of using the null model exclusively as a benchmark for predictive accuracy.

It is difficult to evaluate the model's prediction of location when there is large error in quantity, especially when the model predicts too little change. We can assess the model's ability to predict location somewhat when the model predicts the correct quantity, which is one reason some modelers use the correct quantity for simulation. Nevertheless, if we use only one potential realization of the model's output map, then the model's specification of location is confounded with its single specification of quantity. The Relative Operating Characteristic (ROC) is a statistic that can be used in some instances to measure a model's ability to specify location of land change in a manner that does not force the modeler to commit to any one particular specification of quantity (Swets 1988; Pontius and Schneider 2001; Pontius and Batchu 2003). This is possible when the model generates a map of relative priority of land change, which many models do in their intermediate steps.

3.3 Learning challenges

3.3.1 To use appropriate map comparison measurements.

Scientists have invested a tremendous amount of effort to create elaborate algorithms to model landscape change. We are now at a point in our development as a scientific community to begin to answer the next type of question, specifically, 'How well do these models perform and how do we communicate model performance to peers and others?' Therefore, we need useful measurements of map comparison and model performance. Pontius et al. (in press) derived a set of metrics to compare maps in a manner that we hope is both intellectually accessible and scientifically revealing because analysis of rigorous, clear measurements is an effective way to learn. The initial invitation to participants asked them to submit their recommended criteria for map comparison. Few participants submitted any criteria, and those that did, typically recommended the percent correct of pixels in agreement between the reference map of time 2 and the prediction map of time 2.

This percent correct criterion is one that many modelers are likely to consider. However, it can be extremely misleading, especially for cross-case comparisons, because it fails to consider the landscape dynamics, since it fails to include the reference map of the initial time. For example, the Santa Barbara case has by far the largest percent correct, 97%, simply because there is very little observed change on the landscape and the model predicts less than the amount of observed change. On the other hand, the Cho Don case has the by far smallest percent correct, 54%, primarily because the Cho Don case has far more observed change than any other case. The Perinet case has the largest figure of merit, while its percent correct of 81% ranks just below the median. Producer's Accuracy, User's Accuracy, and Kappa are other indices of agreement that are extremely common in GIS and can be extremely misleading to assess the accuracy of land change models (Pontius 2000). One of the major findings of this research is that the figure of merit has many desirable properties to evaluate the maps from land change models.

We need to continue to invest effort to improve methods of map comparison. Software is now becoming available specifically for map comparison. The Map

Comparison Kit includes a variety of new tools (Visser and de Nijs 2006; Hagen 2003). Selected modules in the GIS software Idrisi allow scientists to compare maps where the pixels have simultaneous partial membership to several categories (Pontius 2002; Pontius and Cheuk 2006).

3.3.2 To learn about land change processes.

During the panel discussions, participants agreed that a main reason for modeling land use and cover change is to increase our understanding of land use and land cover change (LUCC) processes, and that scientists should design a research agenda in order to maximize the opportunity for learning about such processes. Therefore, scientists should strive to glean from this validation exercise important lessons about the processes of land change and about the next most important steps in the research agenda.

Some attendees at the panel discussion expressed concern that this validation exercise focuses too much on prediction to the exclusion of increasing our understanding of the underlying processes of LUCC. Many modelers claim that they seek explanation, not necessarily prediction. Some scientists think that a model can predict accurately for the wrong reasons; in addition these scientists think a model can capture the general LUCC processes, but not necessarily predict accurately due to inherent unpredictability of the processes.

Other scientists view the situation differently, and see validation as an important means to distinguish better explanations from poorer explanations concerning the LUCC processes. For these other scientists, validation is a crucial step because a modeler needs to measure the degree to which the behavior of the model is similar to the behavior of the land system. Furthermore, it is important to test the extent to which the past is useful to predict the future because this allows us to measure the scale at which LUCC processes are stationary, i.e. stable over time. A model's failure to predict accurately may indicate that the process of land change is non-stationary, in which case the validation exercise would reveal information that would indeed be helpful to learn about the LUCC processes. If scientists interpret the validation procedure in an intelligent manner, then they can perhaps learn more from inaccurate predictions than from accurate ones. Consequently, inaccurate predictions do not mean that the modeling exercise is fruitless, since the validation procedure can be fruitful whatever the revealed level of accuracy.

This difference in views might explain the variation in the LUCC modeling community concerning how best to proceed. One group thinks that models are too simple so that future work should consider more variables and develop more complex algorithms so the models can generate a multitude of possible outcomes. A second group insists that such an approach would only exacerbate an existing problem that models are already too complicated to allow for clear communication of conclusions, even to other experienced experts. From this second perspective, contemporary models lack important aspects of scientific rigor that would not be corrected by making the models more complex. For example, many existing models fail to separate calibration information from validation information, fail to apply useful methods of map comparison, and fail to measure how scale influences the analysis. For this second group of scientists, it would be folly to make more complicated algorithms and to include more variables before we tackle these important basic issues, because we will not be able to measure whether more complex models actually facilitate learning about LUCC processes until we develop and use helpful measures of model performance. This apparent tension could be resolved if the scientist who develop more complex models collaborate with the scientists who develop clearer methods of model assessment.

3.3.3 To collaborate openly.

Participants at the panel sessions found the discussions session particularly helpful because they facilitated open and frank cross-laboratory communication. Many conference participants expressed gratitude to the modelers who submitted their maps in a spirit of openness for the rest of the community to analyze in ways that were not specified a priori. The design of the exercise encouraged participation and open collaboration because it was clear to the participants that the analysis was not attempting to answer the question ‘Which model is best?’

Some participants in the conference discussions reported that they have felt professional pressure to claim that their models performed well in order for their manuscripts to be accepted for publication in peer-reviewed journals. We hope that this paper opens the door for honest and helpful reporting about modeling results. In particular, we hope that editors and reviewers will learn as much from this paper as the conference participants did, so that future literature includes useful information about model assessment.

There is clearly a desire to somehow continue this productive collaboration because it greatly increases learning. One particularly constructive suggestion is to build a LUCC data digital library so that we would have access to each others data, models, and modeling results. The data would be peer-reviewed and have metadata sufficient so that anyone could perform cross-model comparison with any of the entries in the library.

In order for this to be successful, scientists would need sufficient motivation to participate, which would require both funding and professional recognition for participation.

4 CONCLUSIONS

This collective experience clearly supports the statement by Box (1979) that ‘all models are wrong, some models are useful’. It is therefore very important to document how the data and the models are developed and for what purpose, in order to judge them fairly. This paper has found that it can be quite challenging to establish procedures and measurements for fair judgment. This paper illuminates some common pitfalls, explains them, and offers some guidance for ways to avoid and to address them. If scientists meet the challenges specified in this paper, then we are likely to make progress in using the validation step to facilitate efficient learning. Hopefully, articulation of these challenges will help scientists prioritize a research agenda for land change science.

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 U.S.A. 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®.

LITERATURE

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. *Simulations & Gaming*, *34(3)*, 342-347.

Borsboom-van Beurden, J., Bakema, A. & Tijbosch, H. (2007). A land-use modelling system for environmental impact assessment. In E. Koomen, J. Stillwell, A. Bakema, & H. Scholten (Eds.), *Modelling land-use change; progress and applications*, (pp. 281-296). Dordrecht: Springer Press.

Box, G. E. P. (1979). Robustness in the strategy of scientific model building. In R. L. Launer & G. N. Wilkinson (Eds.), *Robustness in Statistics*, (pp. 201-236). New York: Academic Press.

Castella, J.-C. & Verburg, P. H. (2007). Combination of process-oriented and pattern-oriented models of land-use change in a mountain area of Vietnam. *Ecological Modelling, 10(1)*, 410-420.

Castella, J.-C., Boissau, S., Trung, T. N. & Quang, D. D. (2005a). Agrarian transition and lowland-upland interactions in mountain areas in northern Vietnam: Application of a multi-agent simulation model. *Agricultural Systems, 86(3)*, 312-332.

Castella, J.-C., Trung, T. N. & 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. *Ecology and Society, 10(1)*, 27.

de Koning, G. H. J., Verburg, P. H., Veldkamp, T. A., & Fresco, L. O. (1999). Multi-scale modelling of land use change dynamics in Ecuador. *Agricultural Systems, 61*, 77-93.

de Nijs, T. C. M., de Niet, R. & Crommentuijn, L. (2004). Constructing land-use maps of the Netherlands in 2030. *Journal of Environmental Management, 72(1-2)*, 35-42.

Dietzel, C. K. & Clarke, K. C. (2004). Spatial differences in multi-resolution urban automata modeling. *Transactions in GIS, 8*, 479-492.

Duan, Z., Verburg, P. H., Fengrong, Z., & Zhengrong, Y. (2004). Construction of a land-use change simulation model and its application in Haidian District, Beijing. *Acta Geographica 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. *Integrated Assessment, 4(2)*, 97-105.

Goldstein, N. C. (2004). Brains vs. Brawn – Comparative strategies for the calibration of a cellular automata-based urban growth model. In P. Atkinson, G. Foody, S. Darby, & F. Wu (Eds.), *GeoDynamics* (pp. 249-272). Boca Raton USA: CRC Press.

Hagen, A. (2003). Fuzzy Set Approach to Assessing Similarity of Categorical Maps. *International Journal of Geographical Information Science, 17(3)*, 235-249.

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. *Journal of Geographical Systems, 1(2)*, 155-177.

Kok, K. & Veldkamp, T. A. (2001). Evaluating impact of spatial scales on land use pattern analysis in Central America. *Agriculture, Ecosystems and Environment, 85(1-3)*, 205-221.

Kok, K., Farrow, A., Veldkamp, T. A. & Verburg, P. H. 2001. A method and application of multi-scale validation in spatial land use models. *Agriculture, Ecosystems, & Environment, 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.

Laney, R. M. (2002). Disaggregating induced intensification for land-change analysis: A case study from Madagascar. *Annals of the Association of American Geographers, 92(4)*, 702-726.

McConnell, W., Sweeney, S. P. & Mulley, B. (2004). Physical and social access to land: spatio-temporal patterns of agricultural expansion in Madagascar. *Agriculture, Ecosystems, & Environment, 101(2-3)*, 171-184.

Pijanowski, B. C., Brown, D. G., Manik, G., & Shellito, B. (2002). Using Neural Nets and GIS to Forecast Land Use Changes: A Land Transformation Model. *Computers, Environment and Urban Systems, 26(6)*, 553-575.

Pijanowski, B. C., Gage, S. H. & Long, D. T. (2000). A Land Transformation Model: Integrating Policy, Socioeconomics and Environmental Drivers using a Geographic Information System. In L. Harris & J. Sanderson (Eds.), *Landscape Ecology: A Top Down Approach*, (pp. 183-198). Boca Raton USA: CRC Press.

Pijanowski, B. C., Pithadia, S., Sellito, B. A. & Alexandridis, K.. (2005). Calibrating a neural network-based urban change model for two metropolitan areas of the Upper Midwest of the United States. *International Journal of Geographical Information Science, 19(2)*, 197-215.

Pontius Jr, R. G. (2000). Quantification error versus location error in comparison of categorical maps. *Photogrammetric Engineering & Remote Sensing, 66(8)*, 1011-1016.

Pontius Jr, R. G. (2002). Statistical methods to partition effects of quantity and location during comparison of categorical maps at multiple resolutions. *Photogrammetric Engineering & Remote Sensing, 68(10)*, 1041-1049.

Pontius Jr, R. G. & Batchu, K. (2003). Using the relative operating characteristic to quantify certainty in prediction of location of land cover change in India. *Transactions in GIS, 7(4)*, 467-484.

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

Pontius Jr, R. G. & Lippitt, C. D. (2006). Can error explain map differences over time? *Cartography and Geographic Information Science 33(2)*, 159-171.

Pontius Jr, R. G. & Malanson, J. (2005). Comparison of the structure and accuracy of two land change models. *International Journal of Geographical Information Science, 19(2)*, 243-265.

Pontius Jr, R. G. & Schneider, L. (2001). Land-use change model validation by a ROC method for the Ipswich watershed, Massachusetts, USA. *Agriculture, Ecosystems & Environment, 85(1-3)*, 239-248.

Pontius Jr, R. G. & Spencer, J. (2005). Uncertainty in extrapolations of predictive land change models. *Environment and Planning B, 32*, 211-230.

Pontius Jr, R. G., Agrawal, A., & Huffaker, D. (2003). Estimating the uncertainty of land-cover extrapolations while constructing a raster map from tabular data. *Journal of Geographical Systems, 5(3)*, 253-273.

Pontius Jr, R. G., Boersma, W., Castella, J.-C., Clarke, K., de Nijs, T., Dietzel, C., Duan, Z., Fotsing, E., Goldstein, N., Kok, K., Koomen, E., Lippitt, C. D., McConnell, W., Mohd Sood, A., Pijanowski, B., Pithadia, S., Sweeney, S., Trung, T. N., Veldkamp, A. T. & Verburg, P. H. (in press). Comparing the input, output, and validation maps for several models of land change. *Annals of Regional Science*.

Pontius Jr, R. G., Cornell, J. & Hall, C. (2001). Modeling the spatial pattern of land-use change with GEOMOD2: application and validation for Costa Rica. *Agriculture, Ecosystems, & Environment, 85(1-3)*, 191-203.

Pontius Jr, R. G., Huffaker, D., & Denman, K. (2004a). Useful techniques of validation for spatially explicit land-change models. *Ecological Modelling, 179(4)*, 445-461.

Pontius Jr, R. G., Shusas, E., & McEachern, M. (2004b). Detecting important categorical land changes while accounting for persistence. *Agriculture, Ecosystems & Environment, 101(2-3)*, 251-268.

Pontius Jr, R. G., Versluis, A. J. & Malizia, N. R. (2006). Visualizing certainty of extrapolations from models of land change. *Landscape Ecology, 21(7)*, 1151-

1166.

Silva, E. A. & Clarke, K. C. (2002). Calibration of the SLEUTH urban growth model for Lisbon and Porto, Portugal. *Computers, Environment, and Urban Systems, 26*, 525-552.

Swets, J. A. (1988). Measuring the accuracy of diagnostic systems. *Science 240*, 1285-1293.

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, A. T. & Fresco, L. (1996). CLUE-CR: an integrated multi-scale model to simulate land use change scenarios in Costa Rica. *Ecological Modeling, 91*, 231-248.

Verburg, P. H. & Veldkamp, T. A. (2004). Projecting land use transitions at forest fringes in the Philippines at two spatial scales. *Landscape Ecology, 19,* 77-98.

Verburg, P. H., de Koning, F G. H. J., Kok, K. Veldkamp, T. K., Louise O. F. & Bouma, J. 1999. A spatial explicit allocation procedure for modelling the pattern of land use change based upon actual land use. *Ecological Modelling, 116*, 45-61.

Verburg, P. H., de Nijs, T. C. M., Visser, H., and de Jong, K.. 2004 Jan Ritsema van Eck,. A method to analyse neighbourhood characteristics of land use patterns. Computers, Environment and Urban Systems 28(6): 667-690

Verburg, P. H., Soepboer, S., Veldkamp, T. A., Limpiada, R., Espaldon, V. & Sharifah Mastura, S.A. (2002). Modeling the Spatial Dynamics of Regional Land Use: the CLUE-S Model. *Environmental Management, 30(3)*, 391-405.

Visser, H. & de Nijs, T. (2006). The Map Comparison Kit. *Environmental Modelling & Software, 21(3)*, 346-358.

1 # TABLES

2 **Table 1. Some characteristics of nine models and their thirteen cases.**

Model	Predicted Transitions	Exogenous Quantity	Uses Year 2 Quantity	Pure Pixels	Case	Literature
Geomod	Single	Yes	No	Yes	Worcester	Pontius et al. 2001, Pontius and Spencer 2005, Pontius et al. 2006
SLEUTH	Multiple	No	No	Yes	Santa Barbara	Dietzel and Clarke 2004, Goldstein 2004, Silva and Clarke 2002
Land Use Scanner	Multiple	Yes	No	No	Holland(8)	Hilferink and Rietveld 1999, Koomen et al. 2005, Borsboom-van Beurden et al. 2007
Environment Explorer	Multiple	Optional	No	Optional	Holland(15)	de Nijs et al. 2004, Engelen et al. 2003, Verburg et al. 2004
Logistic Regression	Single	Yes	No	Yes	Perinet	McConnell et al. 2004
SAMBA	Multiple	No	No	Yes	Cho Don	Boissau and Castella 2003, Castella et al. 2005a, Castella et al. 2005b
LTM	Single	Yes	Yes	Yes	Twin Cities, Detroit	Pijanowski et al. 2000, Pijanowski et al. 2002, Pijanowski et al. 2005
CLUE-S	Multiple	Yes	Yes	Yes	Kuala Lumpur, Haidian, Maroua	Duan et al. 2004, Tan et al. 2005, Verburg and Veldkamp 2004,Verburg et al. 2002,
CLUE	Multiple	Yes	Yes	No	Costa Rica, Honduras	de Koning et al. 1999, Kok and Veldkamp 2001, Kok et al. 2001, Veldkamp and Fresco 1996, Verburg et al. 1999

FIGURES

page

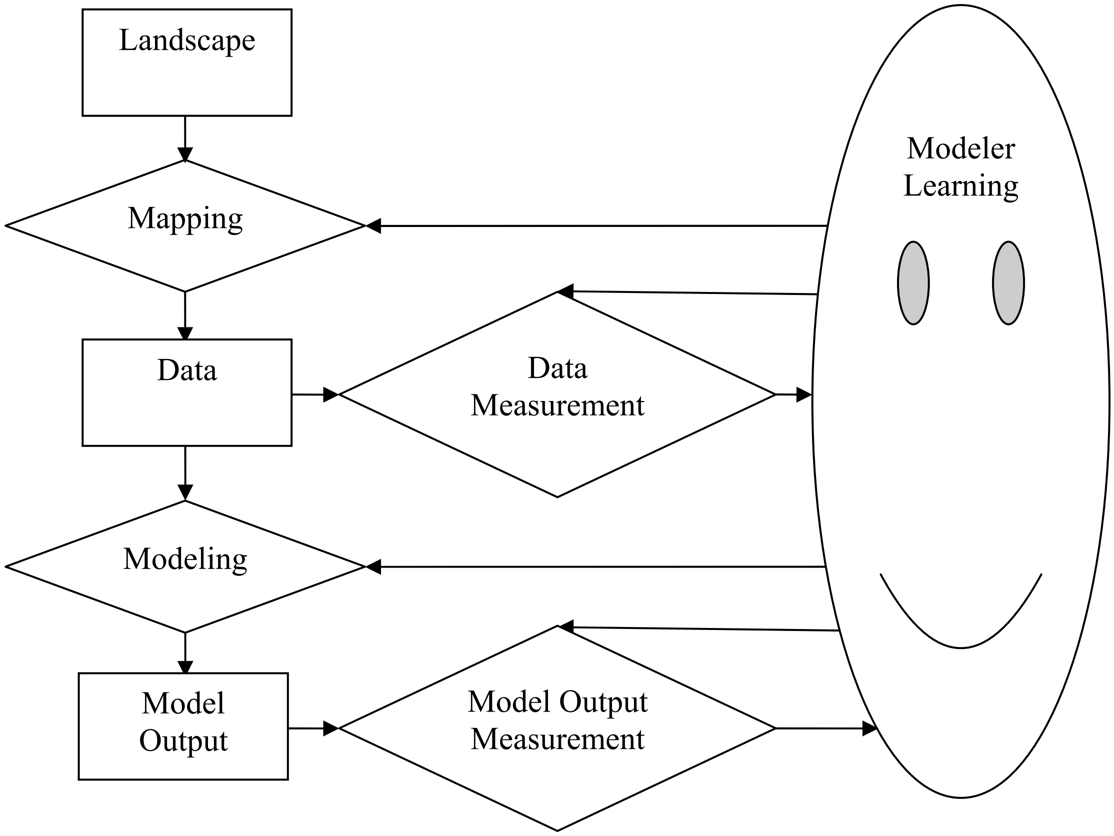


Figure 1. Conceptual diagram to illustrate flows and feedbacks of information among components and procedures for a systematic analysis. Rectangles are components of the research system; diamonds are procedures; the oval is the modeler whose learning can inform methods of mapping, modeling and measuring.

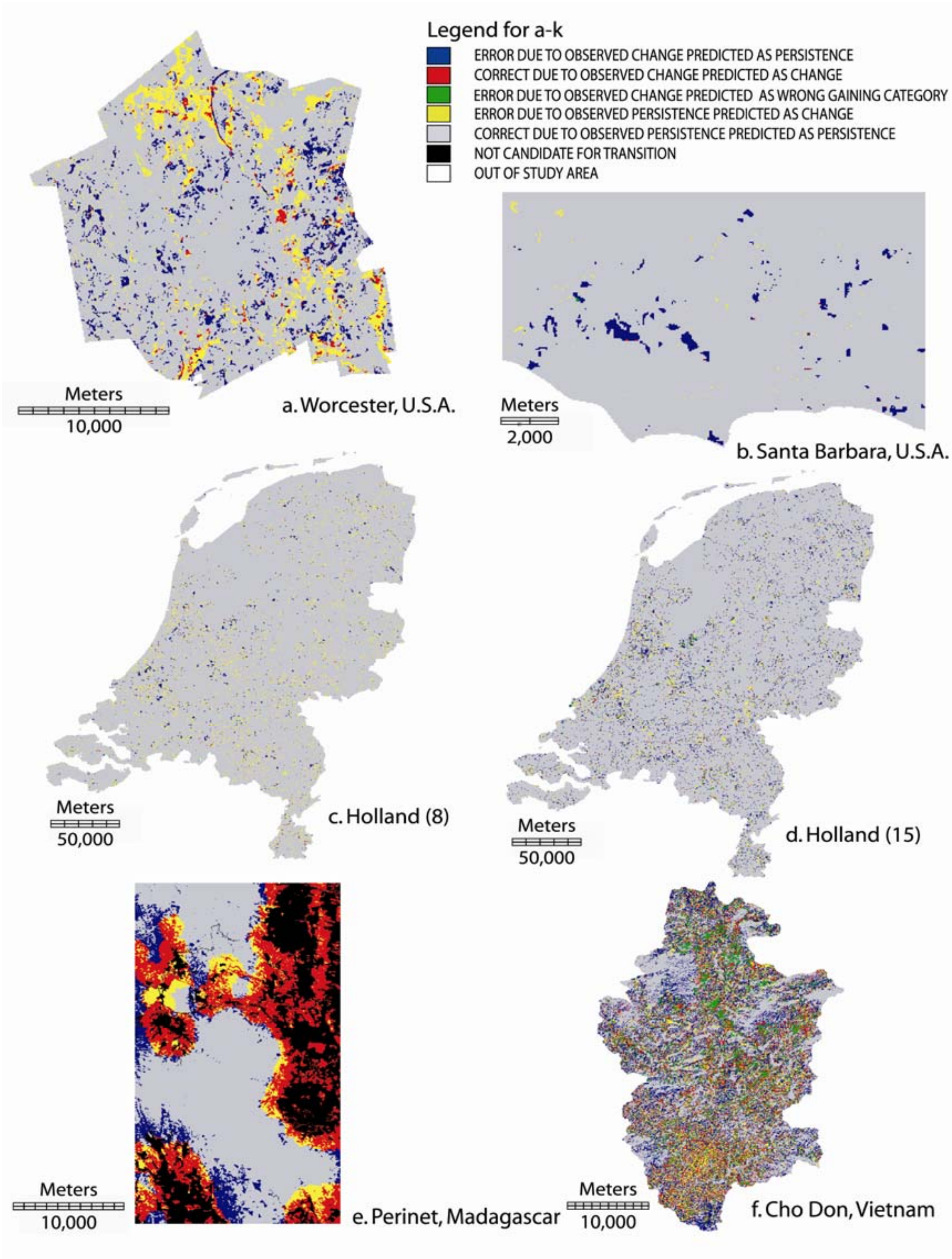
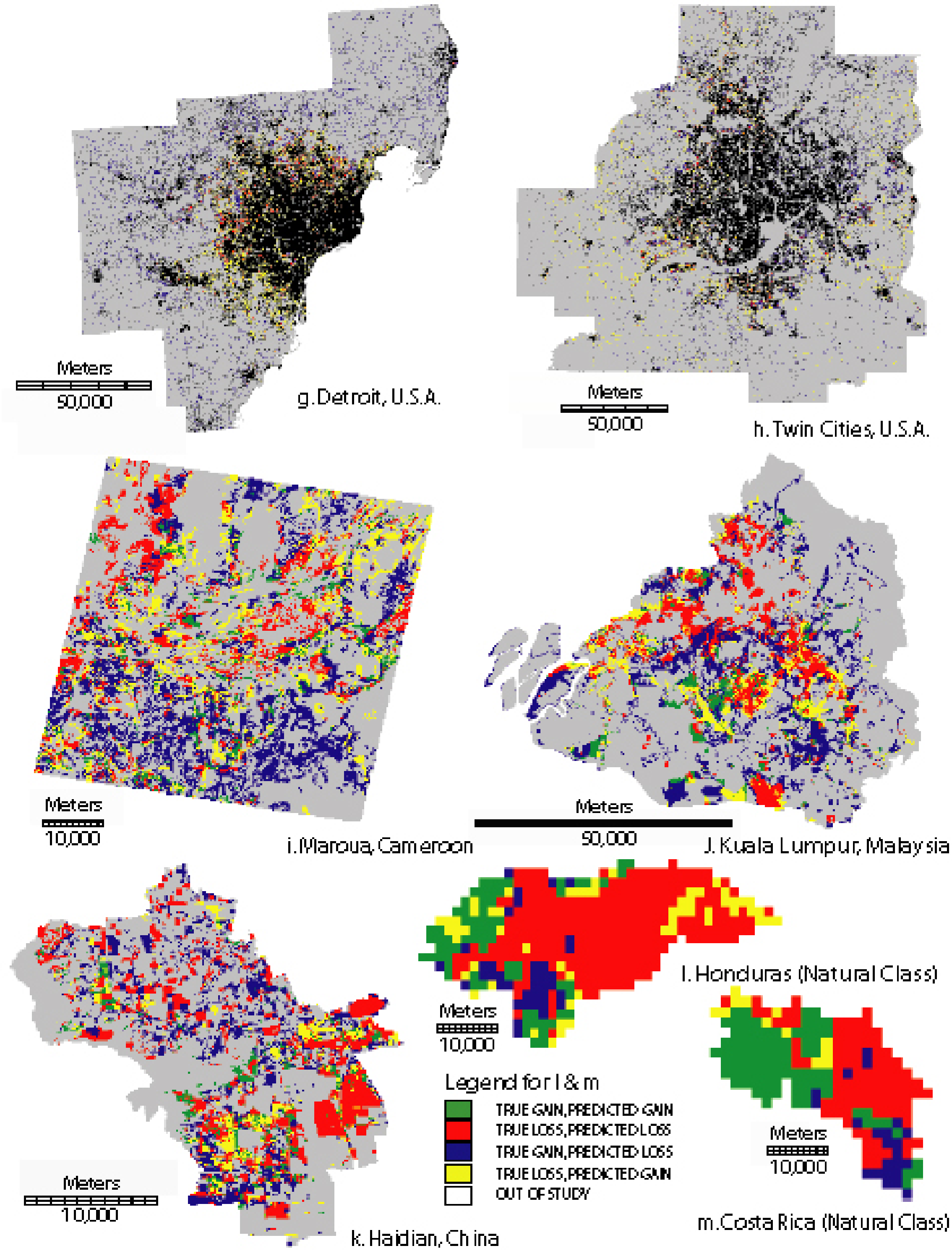
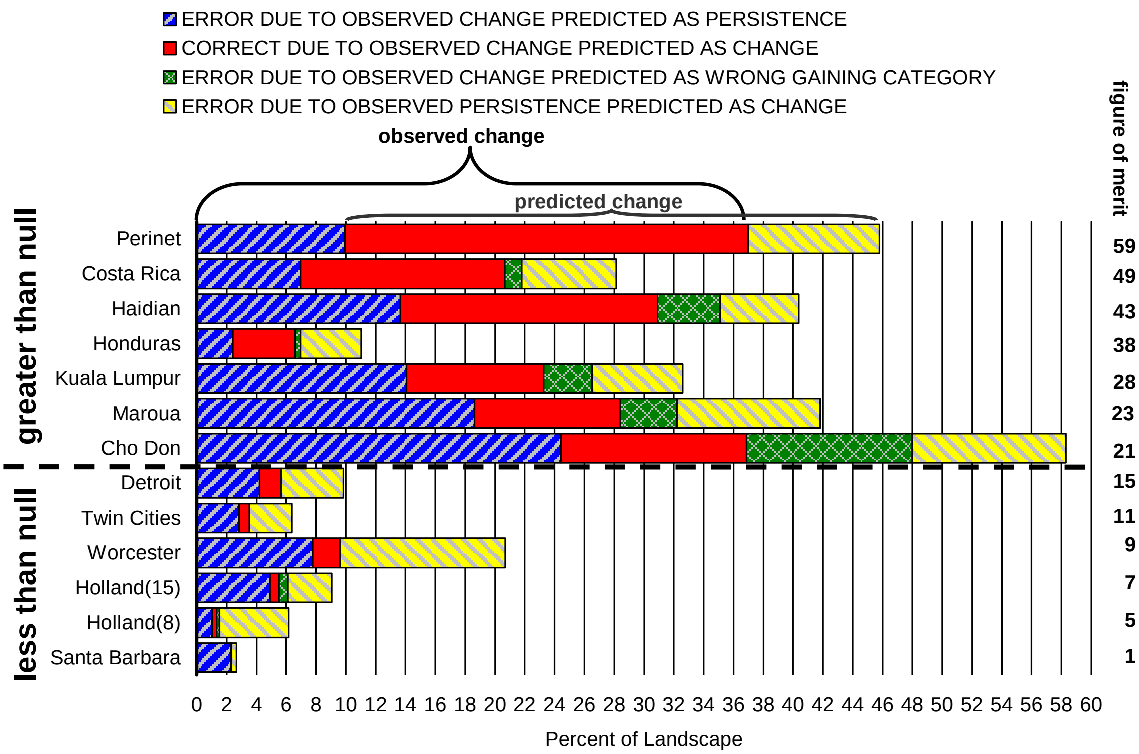


Figure 2. Summary maps of the thirteen cases.

1 

2 **Figure 2 (continued). Summary maps of the thirteen cases.**



1

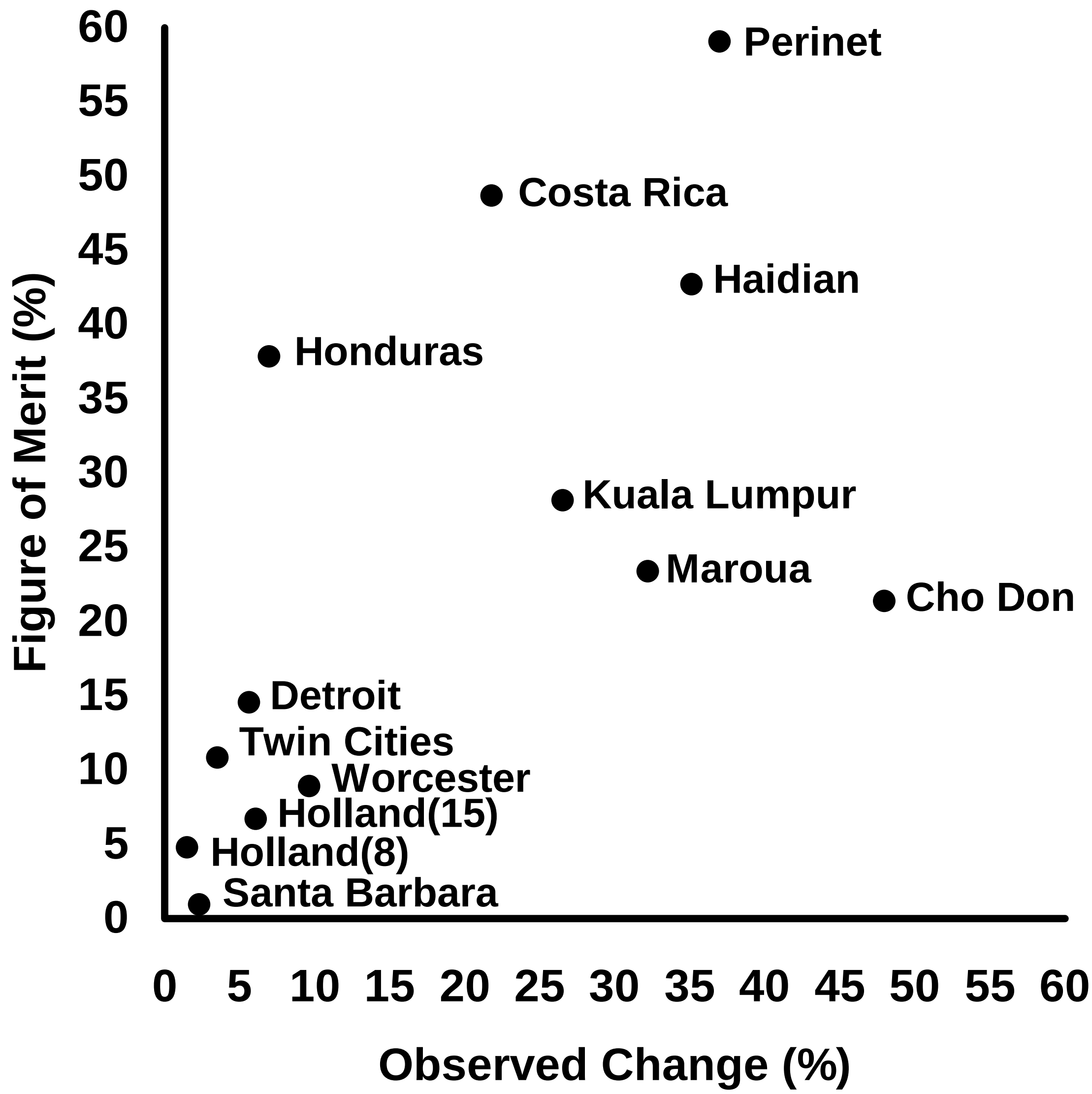
2 **Figure 3. Sources of percent correct and percent error in the validation for thirteen**

3 **modeling applications. Each bar is a Venn diagram where solid and cross hatched**

4 **areas show the intersection of observed change and predicted change. The amount**

5 **of 'Correct due to observed persistence predicted as persistence' is 100% minus the**

6 **length of the segmented bar.**



1

2 **Figure 4. Relationship between the figure of merit (i.e. prediction accuracy) versus**

3 **observed change (e.g. landscape dynamics).**