Transactions in GIS

Research Article

Transactions in GIS, 2013, ••(••): ••–••

Temporal Accuracy in Urban Growth Forecasting: A Study Using the SLEUTH Model

Gargi Chaudhuri* and Keith C. Clarke†

*Department of Geography and Earth Science, University of Wisconsin-La Crosse
†Department of Geography, University of California Santa Barbara

Abstract

This study attempts to establish multi-temporal accuracy of the predicted maps produced by a land use change simulation model over time. Validation of the forecasted results is an essential part of predictive modeling and it becomes even more important when the models are used for decision making purposes. The present study uses a popular land use change model called SLEUTH to investigate the temporal trend of accuracy of the predicted maps. The study first investigates the trend of accuracy of the predicted maps from the immediate future to the distant future. Secondly, it investigates the impact of the prediction date range on the accuracy of the predicted maps. The objectives are tested for the city of Gorizia (Italy) using three sets of map comparison techniques, Kappa coefficients, Kappa Simulation and quantity disagreement and allocation disagreement. Results show that, in addition to the model's performance, the decrease in the accuracy of the predicted maps is dependent on factors such as urban history, uncertainty of input data and accuracy of reference maps.

1 Introduction

Land use change studies usually compare the landscape at two points in time and model the transition quantities and proportions of change both across the landscape and among land use and cover classes (Lambin and Geist 2006). Developments in computation and improved availability of multi-temporal geospatial data, especially from remote sensing, have revolutionized the study of land use change, and present new opportunities for effective modeling. Land use change simulation models involve the creation of abstractions of geographical space and algorithmic representation of the interactions among the different land use types. The use of these models helps to capture the dynamic processes that take place within the space (Clarke 2004; Verburg et al. 2006). Thus these models are necessarily simplifications of reality, and their predictions are at best estimates.

In practice, applying a model consists of four stages (Clarke 2004): input of the initial conditions; calibration; prediction; and validation. Each stage of modeling has its associated error and uncertainty incurred while transforming and representing the real world and during simulation. By prediction, we mean the creation of possible future states based on models, data and theoretical assumptions. An accuracy assessment of the predicted maps outside the model is necessary to establish confidence in the modeled results. When models are used for long-term planning and/or critical decision-making, then they must produce sufficiently accurate results in order to undertake meaningful and informed decision support and planning.

Address for correspondence: Gargi Chaudhuri, Department of Geography and Earth Science, University of Wisconsin – La Crosse, 1707 Pine St., 2031, Cowley Hall, La Crosse, WI – 54601.Telephone: (001) 608 785 8338.E-mail: gchaudhuri@uwlax.edu
Acknowledgments: This study was supported by a UCTC Dissertation Research grant (2009). We thank Dr. Andrea Favretto of the University of Trieste and Dr. Federico Martellozzo of McGill University for data, and Dr. Robert Gilmore Pontius of Clark University for inputs regarding map comparison with quantity disagreement and allocation disagreement. We would also like to thank the anonymous reviewers whose comments have helped to improve the article.

doi: 10.1111/tgis.12047

The present study used a popular land use change model SLEUTH (Clarke et al. 2007) to investigate the temporal uncertainty of spatially distributed forecasts. In this article, uncertainty analysis is defined as the measure of the accuracy of the predicted maps, and temporal uncertainty as the amount of inaccuracy in the multi-temporal predicted maps and its trend over the predicted years. At the final stage of modeling, validation of the predicted maps determines the accuracy of the maps when compared with the observed maps for the corresponding years. In this study, the temporal uncertainty of the predicted maps was calculated by comparing multiple predicted and observed maps of the area under focus from the immediate future to the distant future. The decrease in the accuracy of the predicted maps over time was quantified to explore the temporal uncertainty when using the SLEUTH model.

It is well known that mechanical and information systems inherit the property of entropy, that is, the tendency of a system over time is to become more disordered. An urban model's use is no different, over time its predictive power should decrease and its uncertainties magnify and dissipate. Thus the objectives of this study are to test two null hypotheses: (1) the prediction date range should not impact the accuracy of the predicted maps; and (2) the accuracy of the predicted maps should be equal both in the immediate future and the distant future. It is possible that as longer time periods are modeled, the accuracy of the predicted maps decreases in quantity and errors become more spatially distributed. To test these hypotheses, it is important both to quantify and to spatially locate the inaccuracies in order to understand the magnitude and distribution of variation.

Chaudhuri and Clarke (2012) applied SLEUTH in Gorizia (Italy) and Nova Gorica (Slovenia) to analyze the impact of policy on land use change and the predicted maps from that study provide a good case study to test accuracy over time. To evaluate the first hypothesis, the SLEUTH land use change model was used to make forecasts over multiple date ranges both in the past and the future (prediction from: (1) 1970–2040; (2) 1985–2040; (3) 2000–2040; (4) 2004–2040) to measure the effect of temporal extent on the accuracy of the predicted maps. For the second, the study used the Gorizia-Nova Gorica results from the prediction range that held the maximum accuracy (2004–2040), and explored the accuracy of each of the predicted maps from the start date of the prediction range (2005) further into the future (2010), where data were available for validation.

2 The Sensitivity of Predicted Maps

According to Saltelli (2000, p. 4) 'Sensitivity analysis studies the relationships between information flowing in and out of the model.' In other words, sensitivity analysis consists of the identification and quantification of the different sources of uncertainty in simulation modeling that impact the model output, such as quality of input data, process understanding of the model, or error propagation. Sensitivity analysis helps to establish confidence in a model and its predictions (Liliburn and Tarantola 2009). Uncertainty analysis, on the other hand, is the measure of accuracy of the model output, which may result from uncertainties associated with modeling and model inputs. Uncertainties in a model and its inputs arise from different sources, and are even more difficult to quantify when the inputs consist of a combination of spatio-temporal geographic data (Liliburn and Tarantola 2009). Some models use methods such as Monte Carlo simulation to provide estimates of uncertainty.

The use of cellular automata (CA) models in the simulation of land use change from one time point to another is subject to the errors and uncertainties associated with CA models (Yeh and Li 2006) and with the spatial data on which its use is based (Crosetto and Tarantola

2001). Multiple studies (Urban et al. 2006; Yeh and Li 2006; Liliburn and Tarantola 2009) have identified the different sources of error in simulation modeling that can lead to the overall uncertainty in CA model outputs. Thus for long-term predictions using these models, quantifying the amount of accuracy in each of the multi-temporal predicted outputs will help to understand how the accuracy of the predicted maps change for the predicted years from the immediate future into the distant future. This information will provide better insight for practical application of the model. Contemporary and related land use change simulation models are also subject to similar errors and uncertainties that ultimately impact their model outputs.

Inaccuracies in the modeled outputs of land use change simulation models can be quantified using various map comparison techniques (Pontius 2000; Hagen 2002; Hagen-Zanker and Martens 2008; Pontius and Millones 2011). During the validation stage of modeling, modelers usually quantify the accuracy of the first predicted map and for subsequent prediction years the accuracies are extrapolated according to the results of the first. A study conducted by Pontius and Spencer (2005) showed that the prediction accuracy of the Geomod model in Central Massachusetts decreased by 90% over 14 years, and to near complete randomness over 200 years. Goldstein et al. (2004) estimated that SLEUTH can be run for forecasts of urban extent for a time period as long as the historical data, in their case 70 years. On the other hand, Candau and Clarke (2000) showed that the accuracy of forecast maps increases with input data from the immediate past, and decreases when the model is calibrated over longer historical time durations, even for short-term forecasts. The present study extends this theme of examining the temporal trend of prediction accuracy by performing an accuracy analysis of each of the first five predicted years of SLEUTH forecasts, to measure the trend in accuracy of predicted maps over time. Analysis of the results helps us to understand the causes of decreasing accuracy over time and its dependence on the regional characteristics of a simulation.

3 Map Comparison Techniques

Accuracy of prediction, in general, is determined by comparing a predicted land use map with an observed map of the same time period with an equal number of land use classes. This procedure is theoretically true when the observed map is perfectly accurate (Pontius and Lippitt 2006); but in reality the observed map itself is subjected to a certain amount of classification error compared with the reality on the ground. Multiple studies have been conducted to account for the errors in the observed maps (Pontius and Lippitt 2006; Pontius and Li 2010). This is particularly relevant when the input maps are from classified remotely sensed data, especially at medium spatial resolution, where measured classification accuracy may be only about 80% due to the mixed pixel problem. An accuracy analysis of the observed maps is essential to determine the credibility of the predicted maps and to provide reasonable explanations for the possible estimates of errors.

At present multiple methods have been used to compare the simulated maps and the actual maps to estimate the accuracy of the predicted land use maps (Hagen 2003; Hagen-Zanker et al. 2005; Hagen-Zanker 2006; Pontius et al. 2004, 2007; Van Vliet et al. 2011, Pontius and Millones 2011). The present study used three map comparison techniques from this literature: the Kappa coefficient (and its variations) (Pontius 2000; Hagen 2002), the quantity and allocation disagreement (Pontius and Millones 2011) and $K_{\text{simulation}}$ (and its variations) (Van Vliet et al. 2011) to assess the temporal trend of accuracy of the predicted urban maps modeled by SLEUTH. The results also help to compare these techniques and to verify the validity of the results. The next section provides a brief explanation of the three map comparison techniques that were applied in this study to test the uncertainty of the predicted maps.

The Kappa statistic has been the most popular measure of accuracy used in the field of remote sensing and map comparison. More recently, studies have explored the applications, advantages and disadvantages of Kappa (Congalton et al. 1983; Monserud and Leemans 1992; Congalton and Green 1999; Smits et al. 1999; Pontius 2000; Wilkinson 2005; Pontius and Millones 2011). The use of the kappa coefficient for assessing map accuracy has been subjected to intense criticism, mostly because of its allowance of chance agreement which leads to an underestimation of map accuracy (Brennan and Prediger 1981; Aickin 1990; Foody 1992, 2004, 2008; Ma and Redmond 1995; Stehman 1997; Stehman and Czaplewski 1998; Turk 2002; Jung 2003; Di Eugenio and Glass 2004; Allouche et al. 2006; Pontius and Millones 2011). Despite the criticism, Kappa indices are still considered important measures for the accuracy assessment of maps (Congalton and Green 2009), and are commonly provided as a measure of classification accuracy. Pontius (2000) derived a number of variations of Kappa in an attempt to overcome the flaws of the standard Kappa, but finally replaced the indices completely with a more useful and simpler approach that focused on the two components of disagreement between maps, the quantity and spatial distribution of the categories (Pontius and Millones 2011). Van Vliet et al. (2011) introduced $K_{\text{simulation}}$ which is similar to the kappa statistic but uses an adjusted stochastic model of random allocation of class transitions relative to the initial map.

3.1 Kappa Statistics

The Kappa statistic is computed from a confusion matrix derived from a cell-by-cell comparison of the observed map and the predicted map (Hagen-Zanker and Martens 2008). It can be measured from a sample, or from all pixels in the two maps. Kappa, based on the percentage of agreement, is corrected for the fraction of agreement that can be expected by pure chance. The Kappa statistic is calculated as:

$$Kappa = \frac{(p_o - p_c)}{(1 - p_c)} \tag{1}$$

where p_o is the observed proportion of the sample or pixels correct, and p_c is the expected proportion correct due to chance (Foody 2004).

The kappa coefficient values can generally be interpreted in the following way: if classification is perfect, then Kappa = 1; if the observed proportion correct is greater than the expected proportion correct due to chance, then Kappa > 0; if the observed proportion correct is equal to the expected proportion correct due to chance, then Kappa = 0; and if the observed proportion correct is less than the expected proportion correct due to chance, then Kappa < 0 (Landis and Koch 1977; Aickin 1990; Pontius 2000). To distinguish between the quantification disagreement and location disagreement between two maps Klocation (Pontius 2000) and Khistogram (Hagen 2002) were introduced. 'Klocation' compares the actual success space to the expected success rate relative to the maximum success space given that the total number of cells of each category does not change (Pontius 2000). It is expressed as:

$$K_{location} = \frac{(p_o - p_c)}{(p_{\max} - p_c)} \tag{2}$$

where Po is the observed proportion correct, p_c is the expected proportion correct due to chance and $p_{\max}$ is the total number of cells taken in by each class. Khistogram indicates similarity of the quantitative model results in the maximal similarity that can be found based upon the total number of cells taken in by each class ($p_{\max}$) (Hagen 2002). $p_{\max}$ can be put in the context of Kappa and Klocation by scaling it to p_c. Khisto can be calculated directly from the histograms of two maps (Hagen 2002) and is expressed as:

$$K_{histo} = \frac{(p_{\max} - p_c)}{(1 - p_c)} \tag{3}$$

These statistics are sensitive to the respective differences in location and in the histogram shape for all land use classes (Visser and de Nijs 2006). Kappa, Klocation and Khisto are connected through the multiplicative relation: *Kappa = Klocation * Khisto* (Visser and de Nijs 2006).

3.2 *Quantity Disagreement and Allocation Disagreement*

Pontius and Millones (2011) developed a simpler and more useful approach to report disagreement between the observed map and predicted map. They sub-divided the cell-to-cell total disagreement between the two maps in quantity of disagreement and allocation disagreement. Quantity disagreement is the amount of difference between the reference map and a comparison map due to a less than perfect match in the proportions of the categories. Allocation disagreement is due to the imperfect match in the spatial allocation of the categories given the proportions of the categories in the reference and the comparison maps. The total disagreement of a comparison map is the summation of the allocation disagreement and the quantity disagreement (Pontius and Millones 2011). Details about the cross-tabulation matrix and equations to calculate quantity and allocation disagreement are well documented in Pontius and Millones (2011).

3.3 $K_{Simulation}$

Van Vliet et al. (2011) developed $K_{\text{Simulation}}$ which uses adjusted information from the original land use map to test the agreement between the simulated land use map and the actual land use map. In this method, Van Vliet et al. (2011) integrated conditional probabilities of expected agreement (which assumes that the probability of finding a certain class at a particular location is dependent on the class that originally occupied that location) with the original Kappa statistic. $K_{\text{Simulation}}$ and its variations can be expressed as:

$$K_{Simulation} = \frac{p_o - p_{e(Transition)}}{1 - p_{e(Transition)}} \tag{4}$$

$$K_{Transition} = \frac{p_{Max(Transition)} - p_{e(Transition)}}{1 - p_{e(Transition)}} \tag{5}$$

$$K_{Transloc} = \frac{p_0 - p_{e(Transition)}}{p_{Max(Transition)} - p_{e(Transition)}} \tag{6}$$

where $P_{e(Transition)}$ is the expected fraction of agreement, given the sizes of the class transitions and $P_{Max(Transition)}$ is the maximum accuracy that can be achieved given the sizes of the class transitions (Van Vliet et al. 2011). The result of $K_{Simulation}$ and $K_{Transloc}$ range from –1 to 1, where 1 indicates perfect agreement between simulated and actual land use class, 0 represents chance agreement due to random distribution and below 0 indicates less accurate than chance agreement due to random distribution. The value of $K_{Transition}$ varies from 0 to 1, where 0 indicates no class transition and 1 indicates perfect agreement between size of class transition in simulation and reality.

4 The SLEUTH Model

SLEUTH is a CA model for the computational simulation of urban growth and land use changes that are caused by urbanization. The model has been applied to different cities and in most regions of the world (Clarke et al. 1997, 2007; Clarke and Gaydos 1998; Clarke 2008; Chaudhuri and Clarke 2013). SLEUTH is an acronym for the gridded map input data layers required by the model: **S**lope, **L**and-use, **E**xclusion, **U**rban extent over time, **T**ransportation, and **H**ill-shade, and simulates land use dynamics as a physical process (Gazulis and Clarke 2006).

SLEUTH is a tightly coupled model involving two CAs, the Deltatron Land Cover Model and the Urban Growth Model. The two cellular automata run in sequence and the output of the newly urbanized cells determines the number of times the deltatron code will be executed. Thus, when urban growth is stagnant, land change pressure is reduced and alternatively when other land use classes are being consumed by rapid urban growth, more inter-class transitions are created (Clarke 2008). The urban areas inside this CA model behave as a living organism trained by a finite set of transition rules that influence the state changes within the two CAs within a set of nested loops. During model calibration, the outer control loop executes Monte Carlo iterations on historical maps and searches for the parameters that best replicate the transitions between the first year of input data (the seed layer) and the last (usually the present day), retaining cumulative statistical data. The second or the inner loop executes the growth rules to replicate the growth and transitions between the individual input periods (Clarke and Gaydos 1998; Sietchiping 2004; Gazulis and Clarke 2006; Dietzel and Clarke 2007).

To model the physical differences that exist in a study area, SLEUTH calibrates the historical data input to derive a set of five control parameter coefficients (dispersion coefficient, breed coefficient, spread coefficient, road gravity, and slope resistance factor) which control the behavior of the system and encapsulate the past urbanization trends of that region (Clarke et al. 1997; Gazulis and Clarke 2006). The impact of these coefficient values determine the degree to which each of the four growth rules influences urban growth in the system (Clarke et al. 1997; Gazulis and Clarke 2006).

The most commonly used calibration process is known as brute force calibration, and during this mode of the modeling, a set of control parameters are refined by three sequential calibration phases: coarse, fine and final calibrations (Silva and Clarke 2002; Dietzel and Clarke 2007). The Optimal SLEUTH Metric (OSM) (Dietzel and Clarke 2007) is used to derive the best fit (degree of similarity between simulated images and control years) and to provide the most robust results for SLEUTH calibration (Clarke 2008). The optimal set of parameters based on the OSM produces an output map that most closely resembles the control data (Dietzel and Clarke 2007; Clarke 2008) and is used in the next step of calibration. The combination of parameters with the highest OSM value in the final calibration phase is then used for prediction, after adjustment to reflect their values at the end of the calibration period rather than the start. Finally, the accuracy of the predicted maps is measured outside the model using different map comparison techniques and an observed map of the predicted year (if available).

The model simulation is made up of a series of growth cycles and four types of growth can take place in the model: Spontaneous, Diffusive, Organic, and Road influenced growth of the non-urbanized cells (Clarke and Gaydos 1998). Apart from the initial growth rules there is a second level of rules, which controls the behavior of the macro-system called the 'self-modification' rules. These rules respond to the aggregate growth rate, they start to increase or decrease the growth control parameters in each of the following growth cycles (Sietchiping 2004). Self-modification is important to avoid linear or exponential growth of the area in the model (Silva and Clarke 2002).

5 Data

This study used the data produced by Chaudhuri and Clarke (2012) for the application of SLEUTH in Gorizia and Nova Gorica in Italy and Slovenia, respectively. SLEUTH uses topographic data in the form of slope and hill-shade (for visualization) maps derived from digital elevation models; two land use layers (1985, 2004) for forecasting land use in the deltatron land use model part; at least four urban layers (classified from Landsat 5 TM 1985, 1991, Landsat 7 ETM+ 1999 and Aster 2004), for statistical calibration of the model; and two or more weighted road maps (1969, 1998) from different time periods (Chaudhuri and Clarke 2012). For the urban layer, the built-up areas (red-roofed houses and concrete buildings) and the residential areas were considered as urban and the remainder of the area was classified as non-urban. The road layers were created from 1969 and 1998 topographic maps and were weighted according to their functional classification by level of road. The observed maps of 2005, 2006, 2007, 2009, and 2010 were derived from the classified Landsat 5 TM images for 2005, 2006, 2007, 2009, and 2010 respectively and were used for the validation of the predicted images. A supervised maximum-likelihood classification was performed on the images for land use classification. The objective of supervised classification is to categorize every image pixel into one of several pre-defined land type classes (Jensen et al. 2009).

Two of the input layers have been subjected to extensive accuracy analysis. Beekhuizen and Clarke (2010) worked on the Landsat 5 TM 1991 and ASTER 2004 images of Gorizia with the goal of improving the classification system. With the help of Google Earth they visually interpreted independently derived high-resolution satellite imagery to collect reference data. Google Earth provided access to a good coverage of QuickBird imagery between 2003 and 2007, with a resolution of 0.61 m by 0.61 m. Beekhuizen and Clarke (2010) omitted the sample sites where it was hard to judge land use with Quickbird imagery in order to avoid false reference data. Furthermore, reference data pixels were compared with pixels at the same location in the ASTER satellite imagery to avoid possible misregistration in Google Earth or change in land use between Quickbird and ASTER imagery. In the case where the labeled class of a reference pixel was uncertain, the pixel was removed from the reference data set. However, they gave preference to a bias in the selection of reference data when this effectively decreased the chance of using false reference data. The sampling units consisted of all intersection points in a 500 x 500 m UTM grid, created and overlaid in Google Earth. A second 500 x 500 m grid, with an offset of 250 m from the first grid, was used to collect additional reference data for the classes water and artificial in order to provide sufficient sample sites for both reference datasets (Beekhuizen and Clarke 2010). The same method was implemented to collect ground truth points for the remainder of the images and kappa coefficients were calculated for each of them. Classification accuracy of the input images (1985–2004) varied from 0.74 to 0.80 and the images used for validation (2005–2010) varied from 0.79 to 0.87.

6 SLEUTH Application in Gorizia-Nova Gorica

SLEUTH was applied in Gorizia-Nova Gorica (Figure 1) by Chaudhuri and Clarke (2012) to study the effect of differential policies and political history on urbanization. Gorizia is a small town on the Isonzo River at the foothills of the Italian Alps, astride Italy's northeastern border with Slovenia. It was originally a single city, which over the last century has been occupied by multiple neighboring countries such as Italy, Germany, and Yugoslavia at different time periods. Under each of the regimes, the city was destroyed during warfare, re-built, and finally partitioned into Gorizia (Italy) and Nova Gorica (Slovenia). The international border, on the other hand, has changed from a highly restricted border during the Cold War period to a mere symbolic landmark in 2004 after the inclusion of Slovenia in the European Union (Chaudhuri and Clarke 2012).

To understand whether the trend of urbanization has changed from independent urban growth of each of the cities during the restricted border situation to agglomerated urban

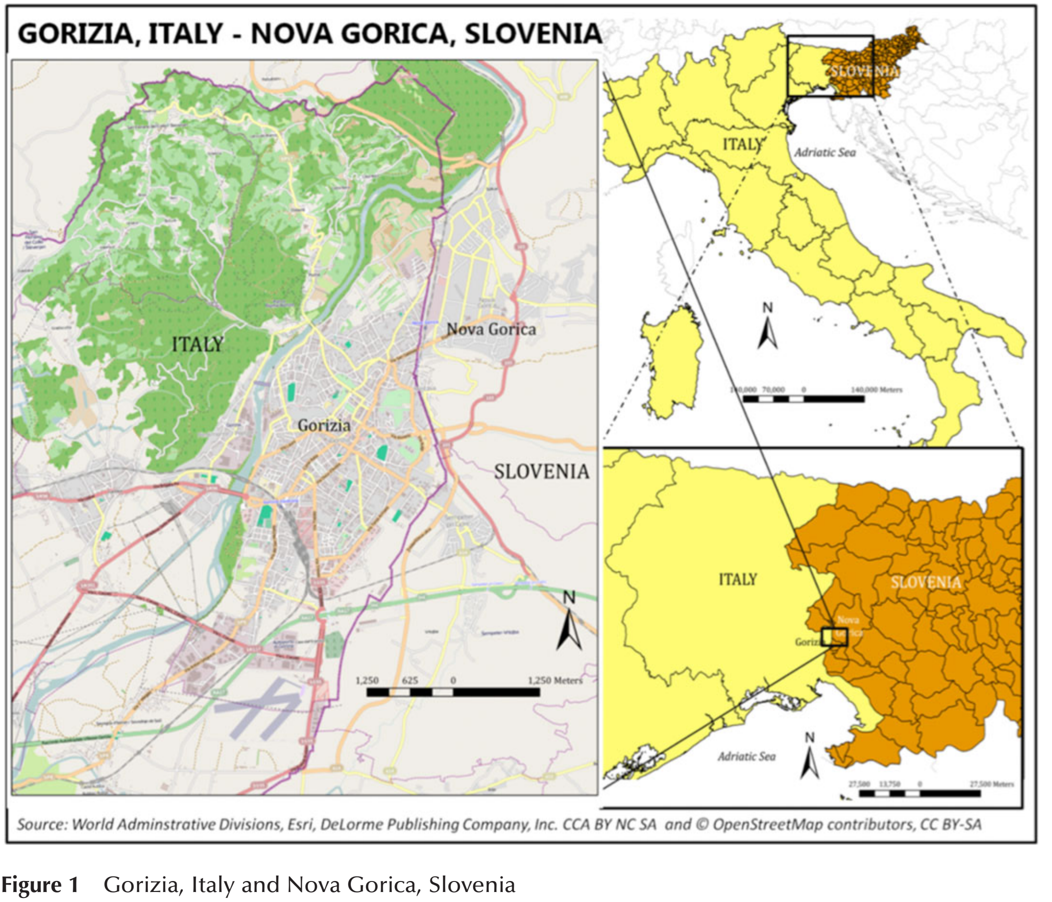

Figure 1 Gorizia, Italy and Nova Gorica, Slovenia

growth after territorial cohesion, the urban growth was simulated under three scenarios. In the SLEUTH model, the exclusion layer was used to define areas where urban growth cannot take place (e.g. on water bodies, or within protected areas). For the first scenario the whole of Gorizia-Nova Gorica was used for simulation and only the water bodies were excluded from future urban growth. For the second and the third scenarios, the city region of Nova Gorica and Gorizia were excluded, respectively, along with water bodies. Thus in the first scenario, both the cities were simulated and forecasted jointly as if they were a single city with uniform characteristics, which represents the situation after the territorial cohesion of the region. The other two scenarios considered and simulated each of the cities independently. These two scenarios represent the situation before the territorial cohesion with a highly restricted international border. The intent of such scenario forecasting was to find out which scenario result was more accurate when compared with the observed map for any given time period, and thus to help understand the trend of urbanization in the region (Chaudhuri and Clarke 2012). Finally, an accuracy assessment of the forecasted maps from each of the scenarios showed relatively higher accuracy for the latter two scenarios, in which each of the cities were simulated independently. This fact led the study to conclude that even after the territorial cohesion both Gorizia and Nova Gorica were growing independently (Chaudhuri and Clarke 2012). To test the second hypothesis, the results from all the scenarios were tested but only results from Scenario 1 are reported in this article. The results from all the scenarios showed a similar trend in temporal uncertainty with only slight variation in the absolute values of accuracy.

7 Methodology

The input data were calibrated using data up to 2004 and predicted from 2005 to 2040. Further details about calibration and prediction are explained in Chaudhuri and Clarke (2012). After rigorous calibration of the SLEUTH model, the best-fit parameter values were used to run predictions from 2004 to 2040 (Chaudhuri and Clarke 2012). A pair-wise map comparison was conducted between observed maps of 1985 and 2005, the observed map of 1985 and the predicted map of 2005 and finally observed change and predicted change of 2005 to compare the actual land transition versus predicted land transition (Chen and Pontius 2010; Pontius et al. 2011).

To test the first hypothesis, the prediction was run with the same set of parameters as the first scenario for three additional prediction ranges: 1970–2040, 1985–2040, and 2000–2040. The predicted maps of 2010 from all the prediction ranges were then validated using the observed map of 2010 and compared against each other. Note that the 2010 map was for accuracy assessment only, and was not part of the model calibration. At present, SLEUTH uses the last input urban layer as the start date for a prediction run. The urban and road images used to initialize growth during prediction are those with dates equal to, or greater than, the prediction start date. If the prediction start date is greater than any of the urban dates, the last urban file on the list is used. Similarly, if the prediction start date is greater than any of the road dates, the last road file on the list is used. The prediction run terminated at the prediction stop date (http://www.ncgia.ucsb.edu/projects/gig/About/data_files/scenario_file.htm). For the second objective, the calibration was run from 1985 to 2004 and the model was used to predict from 2005 to 2040. The accuracy of the predicted maps (reclassified into urban/ non-urban) of 2005, 2006, 2007, 2009 and 2010 (most recent available data) were validated against the observed map of those years.

Validation was performed independently, using the Kappa coefficient (and its variations) and Kappa Simulation (and its variations) in the Map Comparison Kit (MCK), (RIKS and

MNP – RIVM, 2004) (Visser and Nijs 2006). Further map comparison in terms of quantity disagreement and allocation disagreement (Pontius and Millones 2011) was performed between the classified urban map from the Landsat imagery and the predicted urban map from the model.

8 Results

Before analyzing the results of the three map comparison techniques, a pair-wise map comparison (Chen and Pontius 2010) was performed using the observed map of 1985 (initial map), observed map of each year from 2005 to 2010 and the predicted map of each year from 2005 to 2010. The comparisons between the three pairs of maps help to reveal the inaccuracies caused by land persistence versus land change (Pontius et al. 2011).

For clarifying the concept, maps from 2005 were used as an example as explained below. The first pair of comparison maps, the observed urban maps of 1985 and 2005 (Figure 2), showed the observed change of the land categories. The second pair of comparison maps, the observed urban map of 1985 and predicted urban map of 2005, showed the predicted change of the land categories, based on 1985. Finally the third pair of comparison maps, the observed urban map of 2005 and predicted urban map of 2005, shows the level of consistency between the two maps. Figure 2 shows the observed change in urban area from 1985 to 2005 where 19.96% of the non-urban land has changed into urban by 2005.

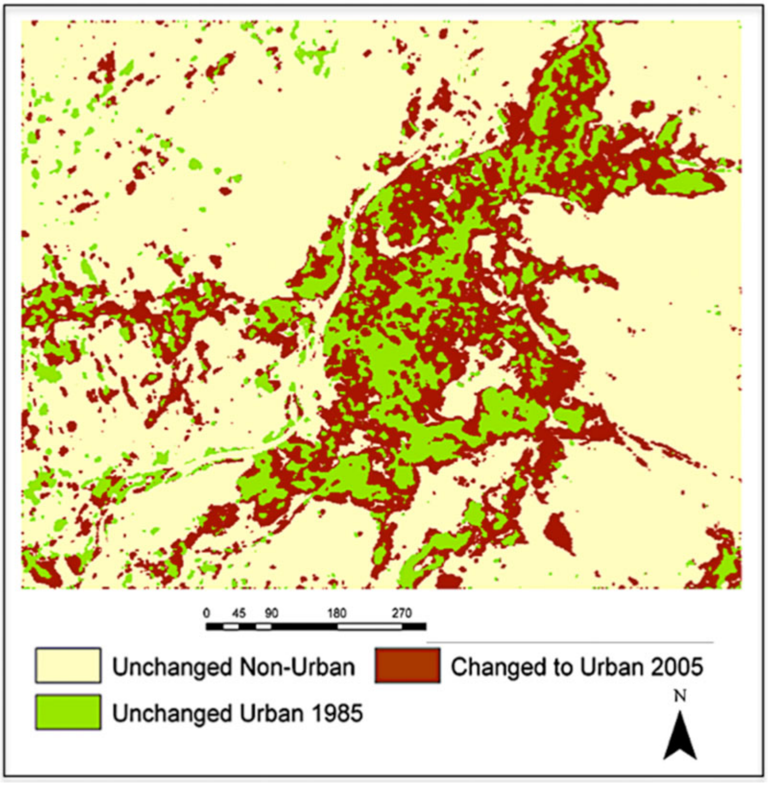

Figure 2 Spatial distribution of change in urban areas based on observed maps of 1985 and 2005

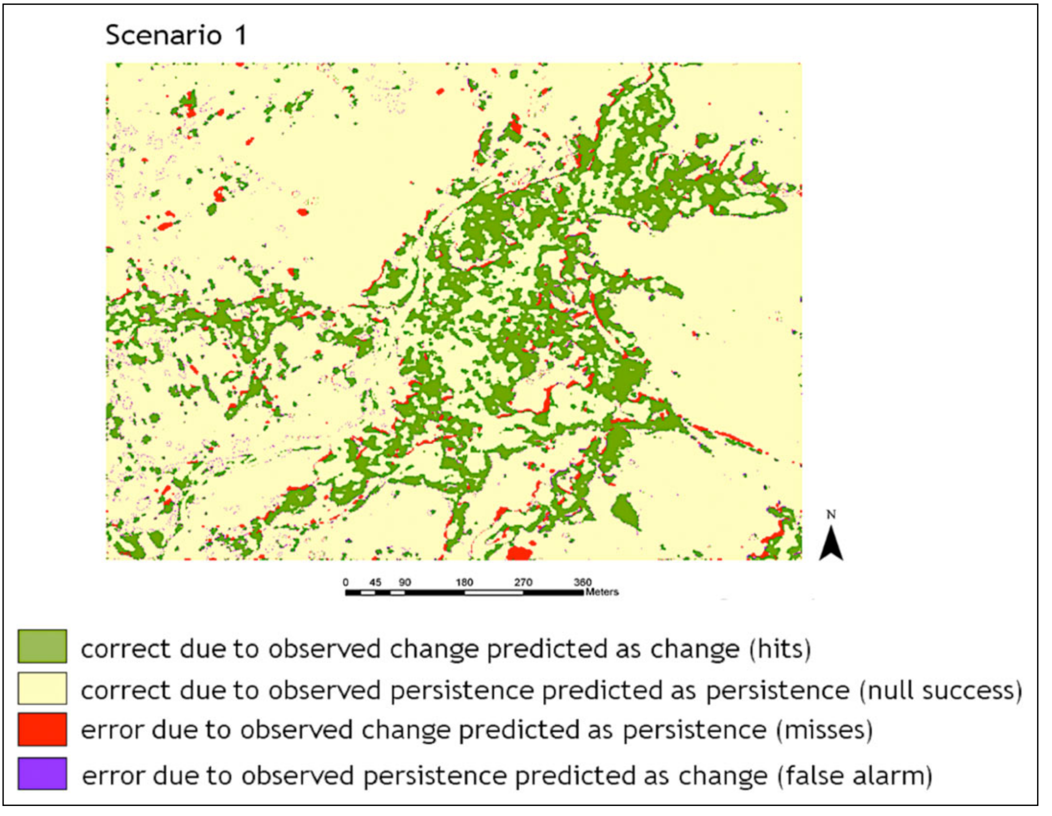

Figure 3 Spatial distribution of prediction correctness and error based on three map comparison (observed map of 1985 and 2005, and predicted map of 2005)

According to Chen and Pontius (2010), changes between the observed and the predicted maps of the urban areas can be classified into four types: (1) correct due to observed persistence predicted as persistence (null success); (2) error due to observed persistence predicted as change (false alarm); (3) correct due to observed change predicted as change (hits); and (4) error due to observed change predicted as persistence (misses). Figure 3 shows the spatial distribution of the four types of prediction correctness and error. The pixels which are categorized under hits and misses are the observed change of non-urbanized pixels to urban pixels in 2005 and the pixels which are categorized under false alarm and hits are the predicted change of non-urbanized pixels to urban pixels by the model. Null success accounts for 79.20% of the landscape in the study area. As mentioned above, the observed change for the area was 19.96%, whereas the predicted change was 19.03%. So the model under-predicts the increase in urban areas, but by less than 1%.

Figure 4 is a graphical representation of the three-map comparison (Chen and Pontius 2010) for years 2005 to 2010 to show the overall prediction correctness and error in the landscape. For each of the years, the 1985 urban map was considered as the initial map. The figure shows that over the years, as the region is experiencing change, the null success of the model decreases and the proportion of hits, misses and false alarms increases. The hits (correct due to observed change predicted as change) and misses (error due to observed change predicted as persistence) when combined together represent the observed change in the landscape. A

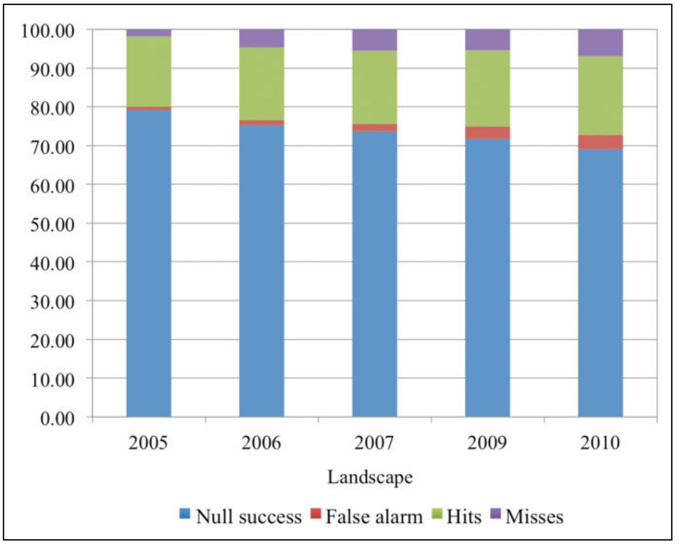

Figure 4 Overall prediction correctness and error in the landscape (Data tables are provided in Appendix)

combination of the hits (correct due to observed change predicted as change) and false alarms (error due to observed persistence predicted as change) represents the predicted change in the landscape. For all the years (2005–2010), the proportions of hits are greater than the combined false alarms and misses. The comparison between the predicted change and the observed change in the maps also shows that the model under-predicts urban growth by approximately 1% (in 2005) – and approximately 3% (in 2010). The associated table for the figure is provided in the appendix. This microanalysis of the pairwise map comparison for all the maps is helpful to understand the reliability of the accuracy values for 2005 and the future years that are discussed below.

8.1 Impact of Prediction Date Range on the Accuracy of the Predicted Maps

The result of the test, as conducted to test the first hypothesis, shows that by changing the prediction start date, the accuracy of the 2010 predicted map decreases (Figure 5). Accuracy analysis of the 2010 predicted map from the different prediction ranges was performed using the $K_{Simulation}$ method, where the 1985 urban map was used as the initial map. For the 1970–2040 prediction range, the values of $K_{TransLoc}$ (0.66) and $K_{Transition}$ (0.64) suggest that the model did not capture the type of land use change and the allocation of the changed pixels very well. For the other prediction ranges, the higher $K_{Transition}$ values and lower $K_{TransLoc}$ values suggest that the model simulates the type of land use change better than the location of the change. With approximately 10% error in the observed map, the predicted map of 2010 using the prediction range of 2004–2040 better explained the land use change than the 2010 map using the prediction range of 1970–2040. This suggests that the prediction start date does have a substantial impact on the accuracy of the predicted maps. As hypothesized, if the start date of prediction is further back in the past from the last input map date, the accuracy of the maps predicted for the future will be less.

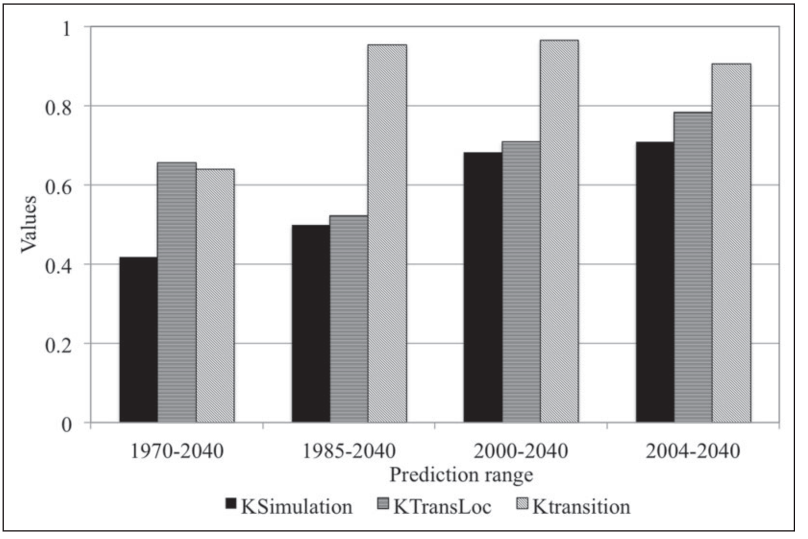

Figure 5 Trend of accuracy of the 2010 predicted maps when the prediction start dates are altered

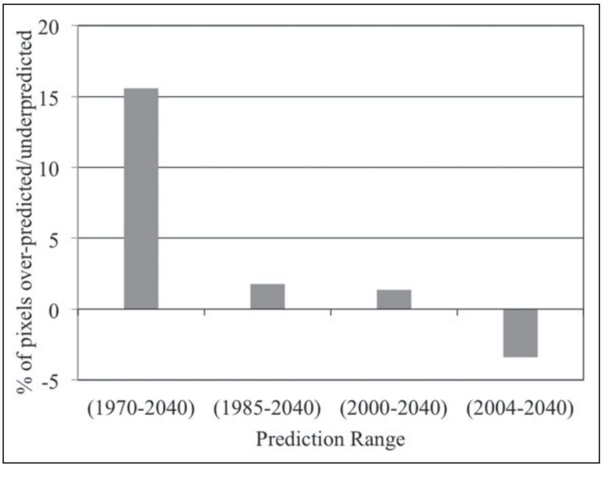

Figure 6 Overestimation/underestimation of the percentage of urban pixels

A comparative analysis of the 2010 predicted map across all prediction date ranges (Figure 6) showed that those starting in 1970, 1985 and 2000 overestimate the percentage of urban pixels in their respective 2010 images, whereas the 2004 prediction start date underestimates the percentage of urban pixels.

At present, each model run is completed using Monte Carlo methods, with each simulation being run from about five to 100 times (http://www.ncgia.ucsb.edu/projects/gig/About/data_files/scenario_file.htm). For calibration, measurements of simulated data are taken for years of known data, and are averaged over the total number of Monte Carlo iterations. The averaged values were compared with the known data, and multiple coefficient measures

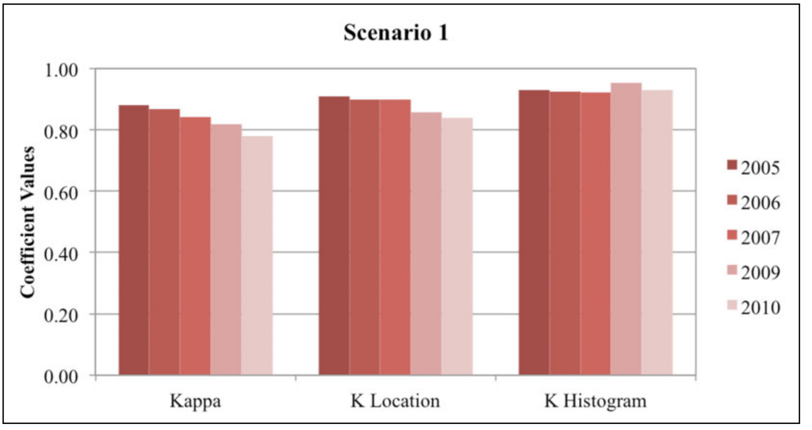

Figure 7 Kappa indices for SLEUTH predicted urban maps of Gorizia-Nova Gorica from over time

are calculated, the product of which gives the Optimum SLEUTH Metric or OSM (see appendix for OSM equation) (Dietzel and Clarke 2007). The set of coefficient values having the highest OSM from the final calibration are considered as the 'best-fit values' and are used to initiate each simulation in a prediction run along with the SLEUTH images and a random seed value. After a simulation is complete, the initializing seed that began that simulation is reset and a new simulation is run (http://www.ncgia.ucsb.edu/projects/gig/About/bkStrPrediction.html). Thus, for Gorizia, in the first three cases, when the prediction start dates are in the distant past, with the average period of growth at 20 years, the model overestimated urbanization for 2010, whereas when the prediction is from the last time point of input data the model underestimated urbanization in 2010. This shows that in reality the region experienced major fluctuations in the rate of urbanization and at present the growth rate is higher than the average.

8.2 *Trend of Accuracy of the Predicted Maps from the Immediate Future to the Distant Future*

In the second test, the predicted urban maps from 2005 to 2010 were compared with the actual land use maps of those years, to show the trend of accuracy over time. The accuracies were calculated using all three methods of map comparison as discussed previously.

8.2.1 *Kappa coefficient results*

The result of measuring the Kappa coefficient (and its variants) for the years from 2005 to 2009 (Figure 7) varies from 0.81 to 0.85, which indicates a high level of prediction accuracy, but for 2010 the overall Kappa decreases to 0.78, although the *k*histo and *k*location remains in the high accuracy class range. As expected, 2005 has the highest level of agreement whereas 2010 has the least. Interestingly, for Kappa and Kloc there is a decreasing trend in accuracy but Khisto, after an immediate decrease in 2006, remains similar until 2010 with only minor deviation.

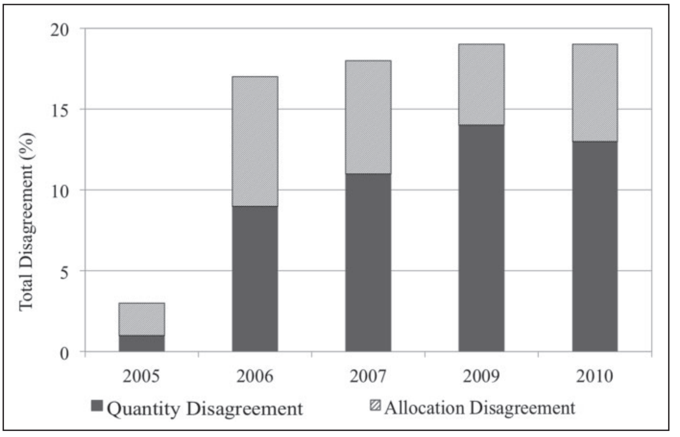

Figure 8 Quantity and allocation disagreement between observed map and SLEUTH predicted urban maps of Gorizia-Nova Gorica over time

Table 1 $K_{Simulation}$, $K_{TranLoc}$ and $K_{Transition}$ for the urban maps of Gorizia-Nova Gorica

Scenario 1	2005	2006	2007	2009	2010
$K_{Simulation}$	0.91	0.82	0.78	0.75	0.71
$K_{TransLoc}$	0.94	0.91	0.88	0.81	0.78
$K_{Transition}$	0.97	0.89	0.89	0.94	0.90

8.2.2 Quantity disagreement and allocation disagreement results

The amounts of quantity and allocation disagreement (Figure 8) show that, the total disagreement between the observed images and the predicted images increases with time. In general, the proportion of quantity disagreement is equal to or greater than quantity disagreement. With errors ranging from 10 to 12% for the observed maps, the results (difference between observed urban area and predicted urban area) show that the model underestimates the amount of urban area for all the forecast years. The underestimation ranges from 2.64% in 2005 to 8.95% in 2010.

8.2.3 Kappa simulation results

Table 1 shows the results of the $K_{Simulation}$ from 2005 to 2010. For 2005, high $K_{Simulation}$ values and its variants suggest that the model was able to explain most of the urban change correctly. Over the years, the $K_{Simulation}$ value decreased from 0.91 in 2005 to 0.71 in 2010. The trend of the relationship between $K_{Transition}$ and $K_{TransLoc}$ remains almost identical, i.e. $K_{TransLoc}$ decreases from 2005 to 2010 whereas the $K_{Transition}$ remains high and almost constant over the span of five years. This suggests that the model simulates the type and amount of land use change correctly but the changed pixels are not correctly located.

9 Discussion and Conclusions

This study addressed three issues in validation of predicted maps from the SLEUTH simulation model. First, evaluating the effect of different prediction date ranges (1970–2040, 1985–2040, 2000–2040 and 2004–2040) showed that using the last urban or road data point as the start date for prediction (2004–2040) produces maximum accuracy in the predicted maps. This has been the usual SLEUTH practice, yet the rationale behind the usage has not been tested before. Secondly, testing the accuracy of Gorizia-Nova Gorica simulations for the years 2005, 2006, 2007, 2009 and 2010 showed that the accuracy of the predicted maps decreased as the model predicts further into the future from the start date of the prediction. Clearly, the most accurate prediction runs only a few years into the future, starting at the present, or at least the time with the most recent data. Lastly, by the application of three types of map comparison techniques (Kappa coefficient and its variants, quantity-allocation disagreement and $K_{simulation}$ and its variants) the study was not only able to capture the overall trend of accuracy of the predicted maps, but also to review in detail the type of agreement/disagreement between the observed map and the predicted map and the over/under-estimation of urban pixels for each of the time periods in the predicted maps.

A possible explanation of the decrease in accuracy of the predicted maps for Gorizia-Nova Gorica can be attributed to the trend of actual land use change in the region, which is caused by changes in the socio-economic and political status of the individual cities. Gorizia-Nova Gorica, with its dynamic political history and recent softening of the international border, is experiencing rapid urbanization. Thus the historical data for Gorizia that belongs to the period from the Cold War and before the softening of the international border followed a different behavior in its land transition than the present day. In regions where there is no or very little urban growth, the accuracy of the predicted map may remain almost the same for near future predictions. Comparison among the three pairs of maps (Figure 4) in the study showed that as the model prediction years move further in the future the null success of the model decreases due to the actual land use changes in the region. Moreover, SLEUTH is a mechanistic model, which tries to capture the changes in a region through morphological changes of the landscape. Thus one might argue that the lack of socio-economic-demographic data in the modeling leads to the model's inability to capture the real world processes successfully, which in turn results in increased uncertainty of the predicted maps of distant future. So it can be said that the trend of accuracy of the predicted maps in the future depends not only on the model structure and performance but also on the actual amount of land use change on the ground during the prediction period. Though the results in this study are specific to SLEUTH application in Gorizia-Nova Gorica, which is going through a slow but steady socio-political transformation, yet it surely shows the need for quantifying the accuracy of the other land-use change model outputs for different time points in the future if the results are to be used for long term decision-making. The study also shows that to increase confidence in the prediction results, multiple measures of accuracy or map comparison techniques should be used to cross-validate results for multiple predicted years.

The above analysis revealed that the overall Kappa coefficient and $K_{simulation}$ values decrease as we go further into the distant future. Comparison of the observed and predicted maps does not reveal any trend with the total amount of disagreement between the maps over the years, but it does show higher agreement in 2005 compared with the remainder of the years. If the trend of decreasing accuracy is extrapolated further into the future, the results suggest that the predicted maps 10 years from the prediction start date will be still within the acceptable levels of accuracy. Beyond 10 years, the future prediction becomes increasingly more uncertain.

References

Aickin M 1990 Maximum likelihood estimation of agreement in the constant predictive probability model, and its relation to Cohen's Kappa. *Biometrics* 46: 293–302

Allouche O, Tsoar A, and Kadmon R 2006 Assessing the accuracy of species distribution models: Prevalence, Kappa and true skill statistic (TSS). *Journal of Applied Ecology* 43: 1223–32

Beekhuizen J and Clarke K C 2010 Towards accountable land use mapping: Using geocomputation to improve classification accuracy and reveal uncertainty. *International Journal of Applied Earth Observation and Geoinformation* 12: 127–37

Brennan R and Prediger D 1981 Coefficient kappa: Some uses, misuses, and alternatives. *Educational and Psychological Measurement* 41: 687–99

Candau J and Clarke K C 2000 Probabilistic land cover modeling using deltatrons. In *Proceedings of the Thirty-eighth Annual Conference of the Urban Regional Information Systems Association*, Orlando, Florida

Chaudhuri G and Clarke K C 2012 How does land use policy affect urban growth? A case study on the Italo-Slovenian border. *Journal of Land Use Science* 7: in press

Chaudhuri G and Clarke K C 2013 The SLEUTH land use change model: A review. *International Journal of Environmental Resources Research* 1: 88–105

Chen H and Pontius Jr R G 2010 Diagnostic tools to evaluate a spatial land change projection along a gradient of an explanatory variable. *Landscape Ecology* 25: 1319–31

Clarke K C 2004 The limits of simplicity: Toward geocomputational honesty in urban modeling. In Atkinson P, Foody G, Darby S, and Wu F (eds) *Geodynamics*. Boca Raton, FL, CRC Press: 215–32

Clarke K C 2008 Mapping and modeling land use change: An application of the SLEUTH Model. In Pettit C, Cartwright W, Bishop I, Lowell K, Pullar D, and Duncan D (eds) *Landscape Analysis and Visualization: Spatial Models for Natural Resource Management and Planning*. Berlin, Springer: 353–66

Clarke K C and Gaydos L 1998 Loose-coupling a cellular automaton model and GIS: Long-term urban growth prediction for San Francisco and Washington/Baltimore. *International Journal of Geographical Information Science* 12: 699–714

Clarke K C, Gazulis N, Dietzel C K, and Goldstein N C 2007 A decade of SLEUTHing: Lessons learned from applications of a cellular automaton land use change model. In Fisher P (ed) *Classics from IJGIS: Twenty Years of the International Journal of Geographical Information Systems and Science*. Boca Raton, FL, CRC Press: 413–25

Clarke K C, Hoppen S, and Gaydos L 1997 A self-modifying cellular automaton model of historical urbanization in the San Francisco Bay Area, *Environmental and Planning B* 24: 247–61

Congalton R G and Green K 1999 *Assessing the Accuracy of Remotely Sensed Data: Principles and Practices*. Boca Raton, FL, Lewis Publishers

Congalton R G and Green K 2009 *Assessing the Accuracy of Remotely Sensed Data: Principles and Practices* (Second Edition). Boca Raton, FL, CRC Press

Congalton R G, Oderwald R G, and Mead R A 1983 Assessing Landsat classification accuracy using discrete multivariate analysis statistical techniques. *Photogrammetric Engineering and Remote Sensing* 49: 1671–78

Crosetto M and Tarantola S 2001 Uncertainty and sensitivity analysis: tools for GIS-based model implementation. *International Journal of Geographical Information Science* 15: 415–37

Di Eugenio B and Glass M 2004 The Kappa statistic: A second look. *Computational Linguistics* 30: 95–101

Dietzel C and Clarke K C 2007 Towards optimal calibration of the SLEUTH land use change model. *Transactions in GIS* 11: 29–45

Foody G M 1992 On the compensation for chance agreement in image classification accuracy assessment. *Photogrammetric Engineering and Remote Sensing* 58: 1459–60

Foody G M 2004 Thematic map comparison: evaluating the statistical significance of differences in classification accuracy. *Photogrammetric Engineering and Remote Sensing* 70: 627–33

Foody G M 2008 Harshness in image classification accuracy assessment. *International Journal of Remote Sensing* 29: 3137–58

Gazulis N and Clarke K C 2006 Exploring the DNA of our regions: Classification of outputs from the SLEUTH model. In El Yacoubi S, Chapard B, and Bandini S (eds) *Cellular Automata*. New York, Springer Lecture Notes in Computer Science Vol. 4173: 462–71

Goldstein N C and Clarke K C 2004 Approaches to simulating the 'March of Bricks and Mortar'. *Computers, Environment and Urban Systems* 28: 125–47

Hagen A 2002 Approaching human judgment in the automated comparison of categorical maps. In *Proceedings of the Fourth International Conference on Recent Advances in Soft Computing*, Nottingham, United Kingdom

Hagen A 2003 Fuzzy set approach to assessing similarity of categorical maps. *International Journal of Geographic Information Science* 17: 235–49

Hagen-Zanker A H 2006 Map comparison methods that simultaneously address overlap and structure. *Journal of Geographical Systems* 8: 165–85

Hagen-Zanker A and Martens P 2008 Map comparison methods for comprehensive assessment of geosimulation model. In Karimipour F, Delavar M R, and Frank A U (eds) *Computational Science and Its Applications: ICCSA 2008*. Berlin, Springer Lecture Notes in Computer Science Vol. 5072: 194–209

Hagen-Zanker A, Straatman B, and Uljee I 2005 Further developments of a fuzzy set map comparison approach. *International Journal of Geographic Information Science* 19: 769–85

Jensen J R, Im J, Hardin P, and Jensen R R 2009 Image classification. In Warner T A, Foody G M, and Nellis M D (eds) *The Sage Handbook of Remote Sensing*. Thousand Oaks, CA, Sage Publications: 269–81

Jung H-W 2003 Evaluating inter-rater agreement in SPICE-based assessments. *Computer Standards and Interfaces* 25: 477–99

Lambin E F and Geist H J 2006 *Land-Use and Land-Cover Change: Local Processes and Global Impacts*. Berlin, Springer-Verlag

Landis J R and Koch G C 1977 The measurement of observer agreement for categorical data. *Biometrics* 33: 159–74

Liliburne L and Tarantola S 2009 Sensitivity analysis of spatial models. *International Journal of Geographical Information Science* 23: 151–68

Ma Z and Redmond R L 1995 Tau coefficients for accuracy assessment of classification of remote sensing data. *Photogrammetric Engineering and Remote Sensing* 61: 435–39

Monserud R A and Leemans R 1992 Comparing global vegetation maps with the Kappa statistic. *Ecological Modeling* 62: 275–93

Pontius Jr R G 2000 Quantification error versus location error in comparison of categorical maps. *Photogrammetric Engineering and Remote Sensing* 66: 1011–16

Pontius Jr R G and Li X 2010 Land transition estimates from erroneous maps. *Journal of Land Use Science* 5: 31–44

Pontius Jr R G and Lippitt C D 2006 Can error explain map differences over time? *Cartography and Geographic Information Science* 33: 159–71

Pontius Jr R G and Spencer J 2005 Uncertainty in extrapolations of predictive land change models. *Environment and Planning B* 32: 211–30

Pontius Jr R G and Millones M 2011 Death to Kappa: Birth of quantity disagreement and allocation disagreement for accuracy assessment. *International Journal of Remote Sensing* 32: 4407–29

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

Pontius Jr R G, Peethambaram S, and Castella J-C 2011 Comparison of three maps at multiple resolutions: A case study of land change simulation in Cho Don District, Vietnam. *Annals of the Association of American Geographers* 101: 45–62

Pontius Jr R G, Walker R, Yao-Kumah R, Arima E, Aldrich S, Caldas M, and Vergara D 2007 Accuracy assessment for a simulation model of Amazonian deforestation. *Annals of Association of American Geographers* 97: 677–95

Saltelli A, Chen K, and Scott E M 2000 *Sensitivity Analysis*. New York, John Wiley and Sons

Sietchiping R 2004 Geographic Information Systems and Cellular Automata-Based Model of Informal Settlement Growth. Unpublished Ph.D. Dissertation, School of Anthropology, Geography and Environmental Studies, University of Melbourne

Silva E A and Clarke K C 2002 Calibration of the SLEUTH urban growth model for Lisbon and Porto, Portugal. *Computers, Environment and Urban Systems* 26: 525–52

Smits P C, Dellepiane S G, and Schowengerdt R A 1999 Quality assessment of image classification algorithms for land-cover mapping: A review and proposal for a cost-based approach. *International Journal of Remote Sensing* 20: 1461–86

Stehman S V 1997 Selecting and interpreting measures of thematic classification accuracy. *Remote Sensing of Environment* 62: 77–89

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

Turk G 2002 Map evaluation and ‘chance correction’. *Photogrammetric Engineering and Remote Sensing* 68: 123–33

Urban D, McDonald R, Minor E, and Treml E 2006 Causes and consequences of land use change in the North Carolina Piedmont: The scope of uncertainty. In Wu J, Jones K B, Li H, and Loucks O (eds) *Scaling and Uncertainty Analysis in Ecology: Methods and Applications*. Berlin, Springer: 239–58

Van Vliet J, Bregt A K, and Hagen-Zanker A 2011 Revisiting Kappa to account for change in the accuracy assessment of land-use change models. *Ecological Modeling* 222: 1367–75

Verburg P H, Kok K, Pontius Jr R G, and Veldkamp A 2006 Modeling land-use and land-cover change. In Lambin E F and Geist H (eds) *Land Use and Land Cover Change: Local Processes and Global Impacts*. Berlin, Springer-Verlag: 117–31

Visser H and de Nijs T 2006 The Map Comparison Kit. *Environmental Modeling and Software* 21: 346–58

Wilkinson G G 2005 Results and implications of a study of fifteen years of satellite image classification experiments. *IEEE Transactions on Geoscience and Remote Sensing* 43: 433–40

Yeh A G and Li X 2006 Errors and uncertainties in urban cellular automata. *Computers, Environment and Urban Systems* 30: 10–28

Appendix

Table 1 and Figure 4

	2005	2006	2007	2009	2010
Null success	79.20	75.35	73.72	71.76	69.11
False alarms	0.84	1.27	1.82	3.15	3.55
Hits	18.19	18.65	19.00	19.70	20.38
Misses	1.77	4.73	5.46	5.39	6.96
Observed Change (Hits + Misses)	19.95	23.38	24.46	25.09	27.34
Predicted Change (Hits + False Alarms)	19.03	19.92	20.82	22.85	23.93

[1]*Optimum SLEUTH Metric (OSM) Calculation*

OSM = compare × population × edges × clusters × slopes × Xmean × Ymean;
compare: modeled population for final year / actual population for final year, or if Pmodeled > Pactual { 1 – (modeled population for final year / actual population for final year)}.
population: least squares regression score for modeled urbanization compared with actual urbanization for the control years
edges: least squares regression score for modeled urban edge count compared with actual urban edge count for the control years
clusters: least squares regression score for modeled urban clustering compared with known urban clustering for the control years
slope: least squares regression of average slope for modeled urbanized cells compared with average slope of known urban cells for the control years
Xmean: least squares regression of average x_values for modeled urbanized cells compared with average x_values of known urban cells for the control years
Ymean: least squares regression of average y_values for modeled urbanized cells compared with average y_values of known urban cells for the control years