Environ Model Assess (2009) 14:221–230
DOI 10.1007/s10666-008-9154-6

Performance Evaluation of the SLEUTH Model in the Shenyang Metropolitan Area of Northeastern China

Xiaoqing Wu · Yuanman Hu · Hong S. He · Rencang Bu · Jeff Onsted · Fengming Xi

Received: 7 February 2007 / Accepted: 14 May 2008 / Published online: 3 September 2008

Abstract Performance evaluation is crucial for the development and improvement of an urban cellular automata model, such as SLEUTH. In this paper, we employed multiple methods for map comparison and model validation to evaluate the simulation performance of the SLEUTH urban growth model in the Shenyang metropolitan area of China. These multiple methods included the relative operating characteristic (ROC) curve statistic, multiple-resolutions error budget, and landscape metrics. They were used to quantitatively examine model performance in terms of the amount and spatial location of urban development, urban spatial pattern and prediction ability. The assessment results showed that SLEUTH performed well in the way of the quantitative simulation of urban growth for this case study. Similar to other urban growth models, however, the simulation accuracy for spatial location of new development at the pixel scale and urban spatial pattern still needs to be improved greatly. These inaccuracies might be attributed to the structure and nature of SLEUTH, local urban development characteristics, and the temporal and spatial scale of its application. Finally, many valuable suggestions had been put forward to improve simulation performance of SLEUTH model for spatial location of urban development in the Shenyang metropolitan area.

X. Wu · Y. Hu (✉) · H. S. He · R. Bu · F. Xi
Institute of Applied Ecology, Chinese Academy of Sciences,
72, Wenhua Road, Shenhe District,
Shenyang 110016, China
e-mail: huym@iae.ac.cn

J. Onsted
Department of International Relations and Geography,
Florida International University,
11200 S.W. 8th Street,
Miami, FL 33199, USA

Keywords Multiple methods · SLEUTH model · Model performance · Urban growth

1 Introduction

With the increased computational power and the greater availability of spatial data, micro-simulation such as the agent-based and cellular automata simulation methods, has been developed by geographers, planners, and scholars, and it has shown great potential for representing and simulating the complexity of the dynamic processes involved in urban growth and land use change [9]. Cellular Automata (CA) is particularly well-suited for this sort of simulation. In the past 10 years, CA has witnessed significant technological advancements as many CA-based urban models have been developed [2, 6, 20, 35, 36]. These CA-based dynamic spatial urban models provide an improved ability to forecast and assess future urban growth and to create planning scenarios, allowing us to explore the potential impacts of simulations that correspond to urban planning and management policies [15, 17, 19].

The SLEUTH model is a well-known CA-based urban growth model coupled with a land-cover-change model [6–7], which can project urban growth based on historical trends with urban/non-urban data or with detailed categorized land use data under different development conditions [10]. SLEUTH is a moniker for the input data required to use the model: Slope, Land Use, Exclusion, Urban, Transportation, and Hillshade [10]. A more detailed explanation of the model can be found in the literature [6, 10, 30] and on the SLEUTH website, Project Gigalopolis [29]. Despite many examples of recent and ongoing applications, compared with other LUCC (Land-Use and Land-Cover Change) models, SLEUTH lacks a rigorous examination of performance and model validation, which is arguably the most challenging aspect of contemporary urban modeling [28]. Currently, the methods of accuracy assessment applied in the model include: the simple least squares regression, visual comparison, overall agreement and kappa coefficient based on the contingency table, spatial goodness-of-fit statistics such as the Lee–Sallee shape index [6], spatial metrics [15–16], as well as multiple-scale fitting [17]. The least squares regression scores, including the number of urban pixels, urban edges and clusters, are computed in the process of model calibration to measure the goodness of fit between the modeled results and historical urban and land use/cover data, which can only give good expression for the amount of linear urban growth. Though the Lee–Sallee shape index is an explicitly spatial metric in the calibration process, the higher Lee–Sallee value tends to be associated with little or no growth [17]. The combination of remote sensing and spatial metrics has been found to be very effective to assess the SLEUTH's performance, but it is still a challenge to incorporate the metrics into the calibration process of model [15]. The examination of the multiple-scale goodness of fit showed simply that the model was more appropriate for regional scale modeling by the least-square regression [17]. So, up to date, these assessments are one-fold and seldom consider the local characteristics of urban development. In recent years, the need for more robust methods for calibrating and validating CA models has been noted [8, 33]. And, in response, a series of map comparison and model validation methods have been introduced into the validation of LUCC models such as the fuzzy set map comparison, ROC curve statistic and the idea of Null and Random model [28]. Some shareware packages provide comparison algorithms, such as MCK [14], while a variety of statistical indices and tests have been developed, which improve the evaluation ability of the model and help conduct standardization and comparison between different models.

Specifically, the objectives of this paper are: (1) to simulate historical urban expansion in the Shenyang metropolitan area of China using SLEUTH; and (2) to evaluate quantitatively the performances of SLEUTH using multi-methods for map comparison and model validation.

2 Materials and Methods

2.1 Study Area and Data

The Shenyang metropolitan area lies on the transition zone from the branch-range of Changbai Mountain to the flood plain of Liao River in northeast China (Fig. 1).The western alluvial plain is its major topography, while low hilly lands are in the northeast and southeast parts. The Hun River runs through the center of the region. The Shenyang metropolitan area covers approximately 3.4×10^3 km^2, including the central city of the metropolitan area, namely Shenyang, which is the capital of Liaoning province and communication, commerce, science, and culture center of Northeast China. As a key investment and industrial base designated by the new government of China since its liberation from the Japanese in 1948, Shenyang has made significant contributions to China's heavy industrial development in the past several decades. Simultaneously, the central city has continuously expanded its borders with the growth of population and industry, especially after 2000. This rapid urban expansion has occupied much of the surrounding arable land and created substantial change to the region's landscape and ecosystems.

In this study, many data sets have been collected, including: multi-temporal city maps, historical traffic/tourism maps, aerial photography; and a time series of Landsat Thematic Mapper data, population census and statistical data (Table 1). The Landsat Thematic Mapper image in 1997 was geometrically corrected using the topographic map (1:100,000) of 1981. Then the image-to-image method was used for the geo-referenced registration of other images with the total Root Mean Squared (RMS) error of less than 0.5 pixels. Visual interpretation (with local knowledge) from Landsat TM images for the years 1988, 1992, 1997, 2000, and 2004 was carried out to form a binary map of urban/non-urban classes, with the help of ancillary data including the aerial photography of the central city in 1997 and 2001, the topographic map, and ground survey information. In this study, the urban extent included not only all built-up land use types but also the unattached transportation and industrial land outside of the central city which had played a role in urban development. The multiple-temporal urban extent images, called observed maps, were used to perform the calibration and evaluation of the model.

We mapped the 1981 and 2004 road layer based on the topographic map of 1981 and a 2004 map of Shenyang's (1:210,000) detailed transportation network, respectively. These roads were then weighted according to their relative urban attractiveness. Based on the local knowledge available from master and transportation planning schemes and tourism maps, the expressways, national and provincial highways, and local primary roads were classified as the primary road class, which were then given a value of 100, while non-road cells had a value of 0. The secondary roads, including local roads and those reserved for special use, were given a value of 50. We then created 1988, 1992, 1997, and 2000 transportation networks by overlaying the 2004 and 1981 roads with 1988, 1992, 1997, and 2000 Landsat TM satellite images and removing the roads that did not appear in these years or modifying the weight value by referencing the transportation construction data and the historical maps. Dates selected for transportation input were intended to match urban layers.

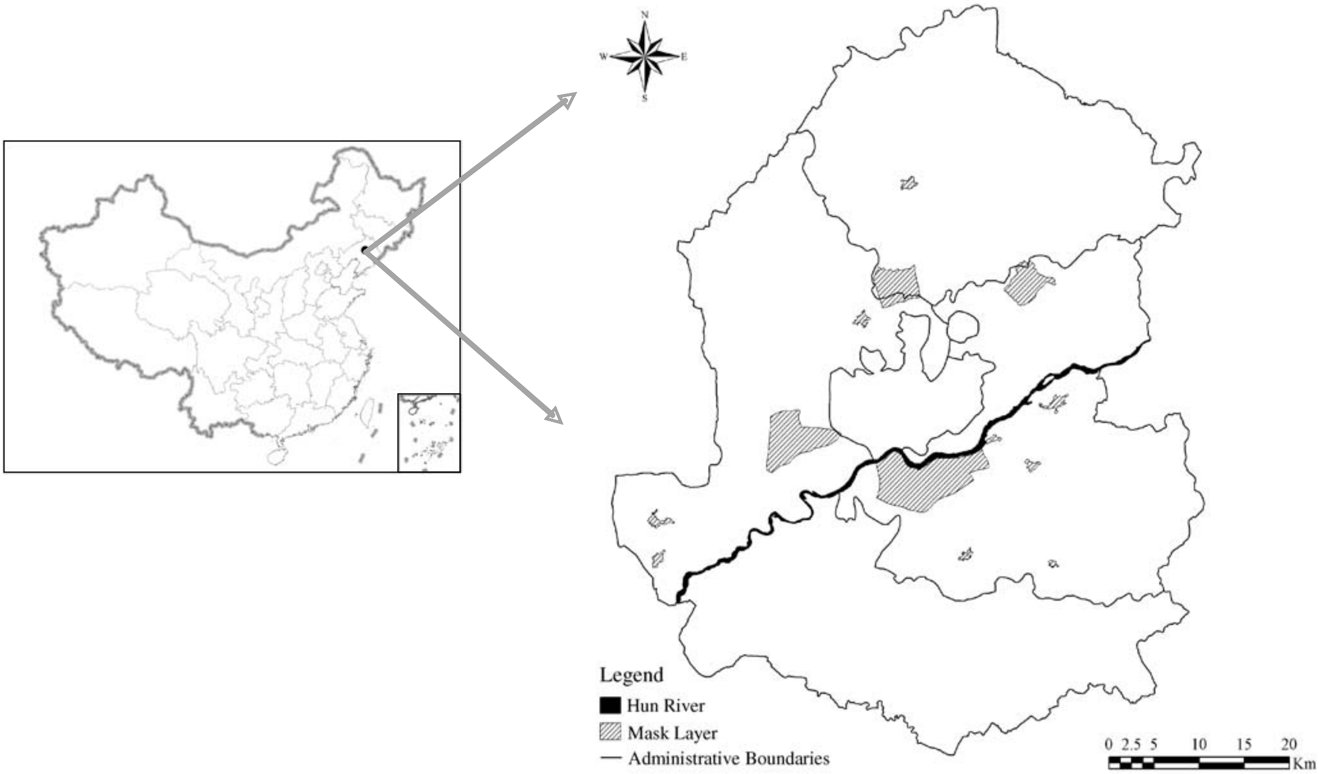

Fig. 1 Location and administrative districts of Shenyang metropolitan area in China and mask layer for assessment of model performance

Designated Excluded layer areas are partially or wholly excluded from urban development with a range of values (0~100) [17]. In the excluded layer, the large river and lakes or reservoirs had a 100% probability of exclusion, while primary parks had an 80% probability of exclusion according to local environmental characteristics and correlated applications of SLEUTH model. These exclusion factors were derived from the 1988 Landsat TM images, the topographic map and map of Shenyang city. Percentage slope and hillshade maps were computed from the DEM with 25 m resolution in GIS. The hillshade layer was used as a background image for model image output. All the input data layers were geo-referenced to the local coordinate system and clipped to the same map extent and then

Table 1 List of model input data for SLEUTH model in Shenyang and mask layer used for model performance evaluation

Theme No.(no. of layers)	Year	Source(s)
Urban extent (5)	1988 1992 1997 2000 2004	Classified from Landsat5 TM satellite images (1988,1992,1997, 2000, 2004) The ancillary data: the Topographical map (1:100,000) of 1981; the Shenyang central ciy maps derived from air photos (1:10,000) of 1997 and of 2001(1:15,000); the map of Shenyang city (1:10,000) of 2000 and the ground survey information (GPS points)
Roads (5)	1988 1992 1997 2000 2004	Digitalized from the Topographical map (1:100,000) of 1981 and the map of Shenyang (1:210,000) of 2005 and updated with Landsat5 TM satellite images (1988,1992,1997, 2000) and transportation construction data and the historical maps
Exclusion (1)	1988	Classified from Landsat5 TM satellite image of 1988
Slope (1) and Hillshade (1)	1981	Derived from 25 m DEM
Mask (1)	–	Mapped according to the map of Shenyang (1:210,000) of 2004, the Landsat5 TM satellite image of 2004, and socioeconomic statistical data for the Shenyang metropolitan area

transformed to raster grids at 60 m resolution. Then, for model calibration, we resampled each data layer into 120 m and 240 m spatial resolution.

2.2 Urban Growth Simulation

We calibrated SLEUTH with the historical urban extent maps for the years 1988, 1992, 1997, and 2000, all of which were extracted from a time series of Landsat TM images for Shenyang. Because of time and computational constraints, we applied the "Brute Force" calibration method (see Project Gigalopolis website) to refine the model parameters in the three sequential calibration phases suggested by the website. This allowed us to find the optimal coefficient sets that could effectively simulate growth during the historic time period. The product metric (the product of all other metrics output during calibration) was used to evaluate the performance of the model in the process of model calibration.

Then, in order to achieve accuracy assessment, based on the optimal coefficients, the model was initialized with 1988 urban extent and growth was predicted out to the year 2004 with one hundred Monte Carlo (MC) iterations. In this study, we do not separate the validation data from calibration data, because of the prediction of less change than observed during 2000–2004, the availability of data for this study area, and cumulative probability of the SLEUTH model. The simulated urbanization probability images for 1992, 1997, 2000, and 2004 were transformed into binary images of urban extent using a probability threshold of 50%, which passed the test in the process of model calibration. These images were then compared with the observed urban extent derived from satellite images and additional statistical validation of the model predictive performance using multiple methods was performed. Simulated urban extent for 2004 served as the main comparison in order to better evaluate the performance of SLEUTH.

2.3 Assessment Methods of Model Performance

2.3.1 ROC Curve Statistic

The Relative Operating Characteristic (ROC) [32], a method used to compare a Boolean variable versus an order variable, has been widely applied in the medical field as well as in the validation of logistic regression models. The ROC compares the likelihood of a given class occurring in a given location to a reference Boolean layer that denotes where the class exists in reality [24]. In this study, we reclassified the urbanization probability image for 2004 into twelve new development maps according to the different probability thresholds (1, 5, 10, 20, 30, 40, 50, 60, 70, 80, 90, and 95) and performed a spatial overlay with observed urban development maps to create the contingency tables. Then, using the tables to calculate the true-positive proportion and false-positive proportion, we were able to plot the ROC curve and compute the area under the curve [34]. The ROC are interpreted as 0.5 (completely random) and 1 (perfect fit). A score between 70–90% shows reliable precision while a score above 90 or under 70% shows high or low precision, respectively.

2.3.2 Multiple-resolutions Error Budget

It is very important to examine the ability of spatially explicit urban growth models to specify accurately both quantity and location of urban growth [23]. It is well known that the Kappa coefficient is a measure of similarity between two maps based on a contingency table and is used for accuracy assessment as a whole [4, 13]. But the Kappa coefficient fails to distinguish clearly between quantification error and location error [24], and cannot quantify the components of agreement and disagreement between two compared maps [28]. The error budget at multiple resolutions not only complements the deficiency, but also can reflect the effect of spatial aggregation on model performance. The details of this philosophy of map comparison can be found in the several papers published by Pontius [24, 25, 27]. Using methods suggested by Pontius [28], we budget sources of agreement and disagreement at multiple resolutions between the simulated and observed urban growth to examine the performance of SLEUTH.

2.3.3 Landscape Metrics

Landscape metrics are increasingly used to study the urban environment [12], which can help improve understanding and representation of urban spatial structure, urban dynamics, and urban modeling. In terms of urban growth modeling, landscape metrics lead to enhanced interpretation and evaluation of modeling results [1]. The type and number of metrics used vary among studies, and different metrics have been found useful in describing different characteristics of model performance and results [16]. In this study, we use four metrics to evaluate the model's performance in terms of urban landscape pattern. First, the number of patches (NP) metric quantifies the number of individual urban clusters. The largest patch index (LPI) metric, second, describes the percentage of the total urban land represented in the largest urban blob [15]. Third, area weighted mean patch of the fractal dimension (AWMPFD) values of all urban patches describes the complexity and the fragmentation, whose expression can be found in the Technical Report of Fragstats software program and correlated literatures [22]. Fourth, we introduce the compactness index (CI) to evaluate the compactness of observed and simulated urban development. The CI is the reciprocal of modified perimeter–area proportion. In this study, the metric is adjusted to minimize the bias towards the large number of small compact urban patches rather than the large complex ones [21]. The revised compactness index (CI′) is given as follows:

$$CI' = \frac{CI}{n} = \frac{\sum_j 2 \times \sqrt{Sj \times \pi}/Pj}{n^2} \tag{1}$$

where CI is the value of the compactness index, *Sj* and *Pj* are the area and perimeter of urban patch *j*, and *n* is the total number of urban patches.

In addition, in order to examine the impacts of urban planning and governmental polices on SLEUTH's model performances in the study area, we created a mask layer to perform another statistical comparison between simulated and observed urban growth outside the masked areas (Table 1). The layer contained the planned Economic Development Zones (EDZs) and the rural settlements whose administrative system had been upgraded to urban between 1988 and 2004.

3 Results

3.1 Comparison Based on ROC Curve Statistic

The dashed line in Fig. 2 illustrates the expected ROC (50%) for a model that selects grid cells at random [24]. The ROC value assessing urban growth between 1988 and 2004 equals 76.5%, demonstrating that model performance is better than random and that it, therefore, has a credible precision. The mask analysis result shows a value of 81.8%, when excluding the EDZs and the rural settlements with a new urban administrative system between 1988 and 2004.

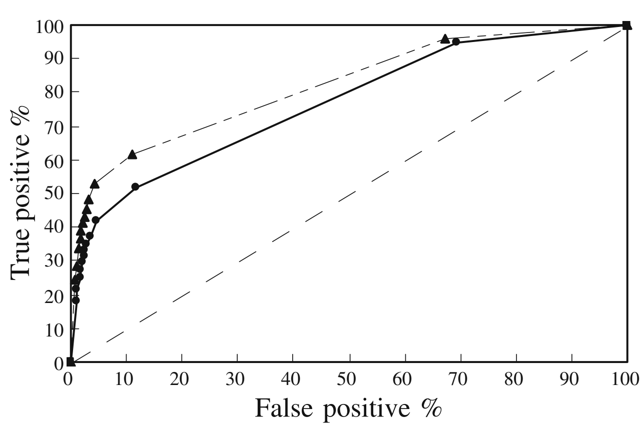

Fig. 2 ROC curve to assess simulation accuracy for urban growth in Shenyang between 1988 and 2004, based on the SLEUTH model (middle ROC=76.5%). Random location (bottom ROC=50%), and the mask analysis result (upper ROC=81.8%)

In order to give a more robust validation, it is best to apply the data sets differed by calibration data to validate the model performance. For example, in this study, we carry out another simulation where the model is initialized with the urban extent of 2000 and predicted out to 2004, and then the ROC value is computed to validate the simulation of urban development between 2000 and 2004. The value amounts to 68.2%, which is lower than the results from the validation between 1988 and 2004. Though this is not encouraging for SLEUTH, the value is still higher than the random model.

3.2 Comparison Based on the Multiple-resolutions Error Budget

Figure 3 shows the error budget results based on the comparison between simulated and observed urban extent maps. Because of only two land use categories, the agreement due to chance is the largest component, which accounts for 50% of the landscape at 60 m resolution. The agreement due to quantity and location equal 27.3% and 17.6%, respectively, the disagreement due to quantity and location accounting for 0.53% and 4.52%, respectively. The above results demonstrate that SLEUTH offers robust simulation for the amount of urban growth at the simulation resolution. The mask analysis results show the disagreement due to location drops to 2.8%, but the disagreement due to quantity rises to 0.7%.This reflects the better agreement between observed and simulated urban spatial development in 2004, when we exclude the above-mentioned EDZs, and the rural settlements with new urban administrative system between 1988 and 2004.

In order to be thorough, it is necessary to examine the model performance at multiple resolutions. Figure 3 shows that the percentages in agreement increase as resolution becomes coarser with the pixel size growing at each subsequent level of aggregation by five cells (300 m). However, the agreement due to quantity acts in just the reverse. The disagreement due to quantity is independent of resolution because resolution relates only to the location of the categories, not the quantity [33]. The disagreement due to location drops resulting from the spatial aggregation procedure, which suggests the SLEUTH model could give good expression for neighborhood relationships at a coarser resolution. And, at any resolution, the disagreement due to location is consistently greater than the disagreement due to quantity. The mask analysis results also show similar trends at various resolutions.

3.3 Comparison Based on Landscape Metrics

Figure 4 shows the temporal signatures for four different landscape metrics. The significant differences in the metric

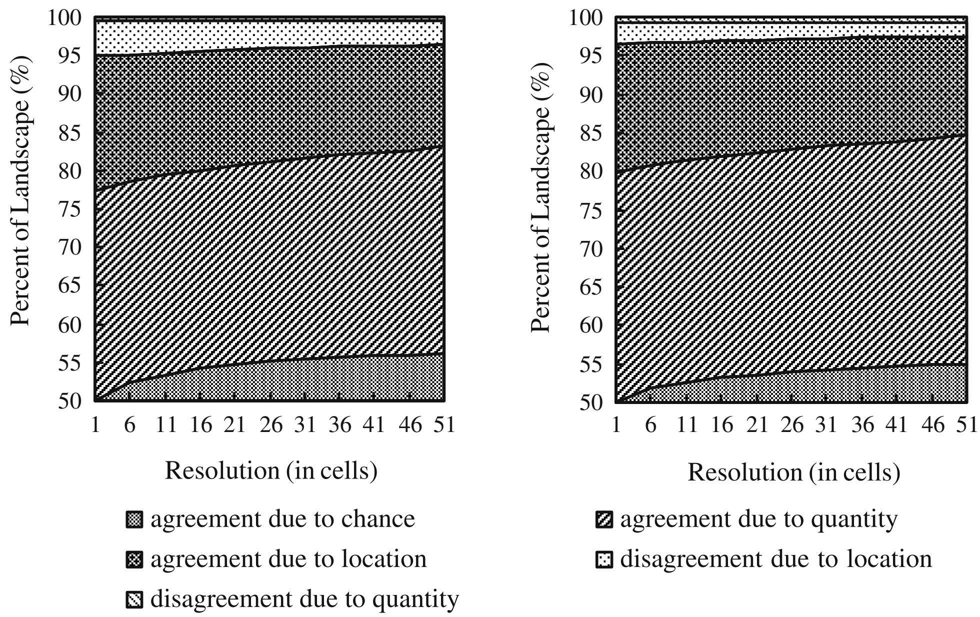

Fig. 3 Multiple-resolutions budget for components of agreement and disagreement between the observed and simulated urban extent map for 2004; SLEUTH model assessment results (*left*) and Mask analysis results (*right*)

graphs refer to variations between modeled results and observed data sets, which represents the model performance for every calibration year. The NP and the fractal dimension in observed maps both increased over the last sixteen years, which indicates the diffuse urban sprawl and "leapfrog-style" urban development during that time (Fig. 4a,b). The corresponding metrics value of simulated urban landscape pattern reversed this general trend. The decreased NP and AWMPFD values for simulated urban growth indicated that continued growth was focused on the regions between the sprawled and fragmented urban areas and connected the areas of diffuse growth to the central urban area. This can be confirmed by minor differences of the LPI values between observed and simulated data sets (Fig. 4c). The *CI'* values also show significant difference between observed and simulated urban extent maps (Fig. 4d). The higher values in the simulated results show that SLEUTH tends to predict urban growth as being too compact. Therefore, SLEUTH can accurately represent the spatial growth of the urban core and the increasing connection of individual urban patches to the central urban area. But there is a general trend to underestimate the spatial growth and

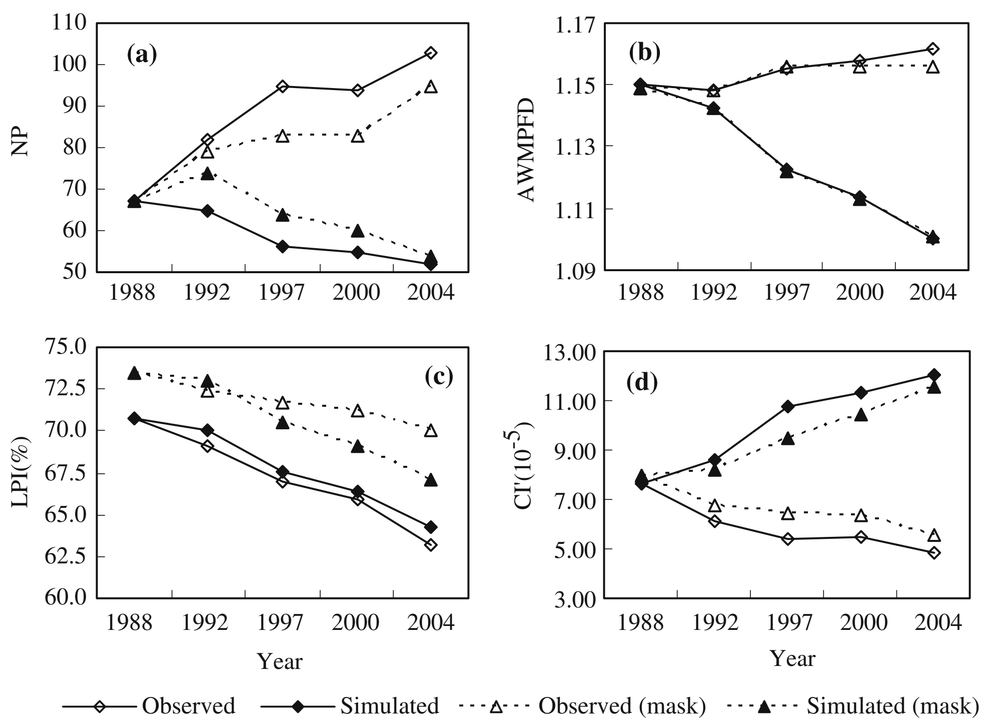

Fig. 4 Temporal urban growth signatures of landscape metrics derived from the SLEUTH model results and observed urban extent maps; (**a**) *NP* number of patches; (**b**) *AWMPFD* area weighted mean patch fractal dimension;(**c**) *LPI* largest patch index; (**d**) *CI* compactness index

complexity; i.e., the growth in most parts is too compact [16]. Mask analysis results show the discrepancy of the above-mentioned metrics values between observed and simulated results decreased.

4 Discussion and Conclusion

4.1 The Effectiveness of the Multi-methods for Assessment of Model Performances

The multi-methods approach for map comparison and assessment of model performances offers a refined assessment of SLEUTH's performance.

ROC curve statistic can examine the accuracy of prediction at several different threshold levels, then aggregate the information at all thresholds into one measure of agreement [26]. So, this method can give overall assessment for prediction ability of a stochastic probability model, for example, SLEUTH. The more probability groups classified, the more accurate an assessment we can obtain. But ROC method cannot account for spatial arrangement of urban development.

The error budget method not only separates the agreement due to location from the agreement due to quantity, but also can give us the magnitude for different components of overall agreement. The disagreement due to location and the disagreement due to quantity are the two most important components for researchers trying to improve the model [28]. This insight can help Geographic Information (GI) scientists decide whether to dedicate energy improving a simulation's ability to either specify quantity or location [24]. Similar to other LUCC models, the prediction performance of SLEUTH for quantity is better than that for location of urban development. But it is very difficult for SLEUTH to obtain simultaneously a high goodness of fit for location and quantity. So deciding which index will be applied for assessment of model performance depends on the specific goals of the research.

The multiple-resolution budget for components of agreement and disagreement and map comparison based on the landscape metrics, on the other hand, can examine model performance at various resolutions and for urban spatial pattern. The landscape metrics help assess the goodness of fit in terms of spatial structure and highlight specific problems, uncertainties or limitations of the model results [16]. Although the error budget at 60 m resolution for the observed and simulated urban extent presents better agreement, the assessment results based on landscape metrics show SLEUTH still need to improve their expression of urban spatial structure and urban morphology at specific resolution. As the resolution becomes coarser, SLEUTH can give better expression for spatial location and neighborhood relationships of urban development.

In this study, the mask analysis showed the disagreement level between observed and simulated urban growth resulting from the construction of EDZs and newly upgraded urban areas. However, in fact, there exist some types of causal relations between the urban growth and the planned EDZs and the newly upgraded urban areas. Thus, the impact of this causal relationship on the assessment of model performance need be discussed in the future study.

4.2 Key Factors Affecting Model Performance

4.2.1 Structure of SLEUTH Model

SLEUTH betrays many disadvantages when attempting to accurately simulate urban growth in Shenyang. First, SLEUTH gives more precedence to the edge growth transition rule, limiting the ability of the model to simulate other urban development processes that arise from less organic origins. This includes leapfrog development pattern resulting from the potential impacts of regional development plans and polices as well as the transfer of administration for some rural settlements from the country to the town in the study area, for example. Discovering a mechanism to simulate the potential impacts of these incentive policies is not currently possible and, therefore, a topic for further research [18]. In addition, the self-modification parameters (SMPs) of SLEUTH are also important to the calibration and prediction of the model. In this study, we do not modify these critical values and, instead, only apply the default values of the model derived from the application to the Washington–Baltimore metropolitan area. Unfortunately, the geometrically larger computational time and resources necessary to calibrate both coefficients and SMPs is prohibitive [17]. Consequently, there is currently no scientific process to evaluate and determine these values in a manner similar to the calibration of the growth coefficients. Any selection, therefore, of these parameters is arbitrary, or at the very least, requires a process involving arbitrary methods of trial and error.

Second, the calibration of the SLEUTH model is very time-consuming, subjective and sensitive. Though the sequential calibration method improves efficiency, it excludes the optimal parameters sets to match historical data sets in fact [3, 17]. The determination of calibration metrics and the method of sorting order are user-selected. Often, the different calibration metrics selected cause to different parameter sets, so it is difficult to decide which are optimal for prediction. In addition, the model is sensitive to the number of MC iterations in the calibration process [37]. In this study, the optimal parameter sets for future prediction were derived from a robust calibration of 20 MC iterations. We discovered this number not only affects computation time but also could affect the parameter sets and simulation preference to some extent.

Third, the randomness and cumulative probability of the model affects its performance. Consequently, some randomly spontaneous growth can not be recorded into the statistical output and/or development at some highly suitable location can not take place at all because the probability threshold used in the allocation of new growth partly affects the simulation accuracy [15].

In addition, for most CA-based urban growth models, the simplicity of cellular automata has been considered a great weakness in representing real cities [38]. But any model is only an abstract representation of the real world and, therefore, flawed. Consequently, it is very difficult for any model to simulate the complex dynamics of urban systems with their attendant uncertainty and emergent nature.

4.2.2 Local Urban Development Characteristics

Every region has its own history, with its own unique attributes that have influenced urban growth [11]. Though SLEUTH attempts to reflect this through its "digital DNA", i.e., the parameter sets derived through calibration, there are still unique urban development characteristics in the Shenyang metropolitan area for which the model has a difficult time accounting. Except for the driving of the regional economic development, the scale and spatial pattern of urban growth has been affected by strong administrative intervention from the local and central government. The regional economic and urban development policies, urban planning, and the transfer of administrative systems are the primary manifestations of this intervention. For example, the extensive construction of EDZs from 1988 to 2004, the large-scale transfer and reconstruction of the Tiexi industrial zone after 2001, and the implementation of the state policy of reviving the Old Northeast Industrial Base. The regional development polices and measures resulted in rapid urban expansion during 1992–1997 and 2000–2004, as well as the transformation of development patterns from the cricoids to an axial urbanization of multi-centers in the central city. SLEUTH, in its current incarnation, can not explicitly account for this planned growth, since the only zoning it incorporates is land officially off-limits to development or land with a weighted resistance against development. Otherwise, as the mask analysis illustrates, model performance of SLEUTH shows a greater improvement when excluding the EDZs and those rural settlements whose administration system has been updated to an urban system during the study period.

4.2.3 The Temporal and Spatial Scale of its Application

Because the calibration process of the model is very time-consuming, we had to apply a resolution of sixty meters, which might have affected the expression of some local growth process [5]. We are also mindful that when budgeting for components of agreement and disagreement at multiple resolutions, the procedure of spatial aggregation from fine to coarse resolution might compound the error for accuracy evaluation.

The time-step and temporal duration of the data can also influence the accuracy of calibration [3] Considering the availability and consistency of data sources and the purpose of model validation, historical urban development in only a 12-year span (1988–2000) were used to calibrate the model. This calibration period, unfortunately, is too short to pick up the large-scale spatial processes. Moreover, during the calibration period, the urban development had been relatively steady, but after 2000, a sharp growth largely compromised the prediction accuracy for the 2004 urban extent. In all fairness though, this emergent development would be difficult for any urban growth model to capture correctly.

4.3 Suggestions for Improvement of Model Performance

We might improve model performance based on information provided by the multiple methods for map comparison and assessment of model performance. Regardless of the approach used to improve model performance, it can be helpful to perform a thorough and deep examination of the historical urban development of a study area to identify key processes or patterns and driving forces of urban growth. This can help estimate the appropriateness of applying SLEUTH to a study area, and assist in selecting the appropriate spatial resolution and the goodness-of-fit statistics. Most importantly, historical examination can provide the most relevant information for improving the model's performance. In this study, based on the historical development characteristics and results from accuracy evaluation in Shenyang, we propose the following methods to improve the simulation accuracy, especially for spatial location of urban development.

First, we may improve the simulation accuracy by modifying the parameters and input data layers or changing the calibration method, including the calibration with full resolution, increasing MC iterations, determining the optimal self-modification constraints, displaying some seeds in urban extent layer, etc. Second, we can perform the sensitivity and uncertainty analysis of SLEUTH model to provide information for the improvement of the model itself as well as how it is applied. The sensitivity analysis for SLEUTH in this particular study area should include testing the temporal sensitivity of the data inputs [3] and an examination of the calibration sensitivity to scaling as well as the influences of growth parameters and the number of MC iterations on model performance. Third, it is necessary to perform urban growth modeling with multiple land use categories, or combine more socioeconomic factors affecting urban growth into SLEUTH as well as other modeling methods and techniques, for example, incorporating a suitability map of urban development derived from a logistic regression model. Lastly, we can modify SLEUTH or develop a new model better suited for the particular characteristics of urban growth in China. The new urban growth model could handle noncontiguous, leapfrog urban development patterns [31], and incorporate the planning objectives and policy factors from local government, which needs to further develop and integrate existing theories into new methodological frameworks.

Acknowledgments We thank Professor Keith C. Clarke, University of California Santa-Barbara, for providing helpful suggestions. This work was jointly sponsored by the Knowledge Innovation Programs of the Chinese Academy of Sciences (KSCX2-SW-133) and the Knowledge Innovation Programs of Institute of Applied Ecology, Chinese Academy of Sciences (O6LYQY1001). We are grateful to the anonymous reviewers for their constructive suggestions and comments.

References

1. Alberti, M., & Waddell, P. (2000). An integrated urban development and ecological simulation model. *Integrated Assessment*, *1*, 215–227. doi:10.1023/A:1019140101212.
2. Batty, M., & Xie, Y. (1994). From cells to cities. *Environment and Planning B: Planning and Design*, *21*, 31–48. doi:10.1068/b21s031.
3. Candau, J. C. (2002). Temporal Calibration Sensitivity of the SLEUTH Urban Growth Model. M.A. Thesis, University of California, Santa Barbara, USA
4. Carletta, J. (1996). Assessing agreement on classification tasks: the kappa statistic. *Computational Linguistics*, *22*(2), 249–254.
5. Claggett, P. R., Jantz, C. A., Goetz, S. J., & Bisland, C. (2004). Assessing development pressure in the Chesapeake Bay watershed: an evaluation of two land use change models. *Environmental Monitoring and Assessment*, *94*, 129–146. doi:10.1023/B:EMAS.0000016884.96098.77.
6. Clarke, K. C., Hoppen, S., & Gaydos, L. (1997). A self-modifying cellular automaton model of historical urbanization in the San Francisco Bay Area. *Environment and Planning B: Planning and Design*, *24*, 247–261. doi:10.1068/b240247.
7. Clarke, K. C., & Dietzel, C. (2004). *A Decade of SLEUTHing: Lessons learned from applications of a cellular automaton land use change model*. Paper presented at the Joint International Workshop on Integrated Assessment of the Land System: The Future of Land Use. Institute for Environmental Studies, Amsterdam.
8. Dietzel, C. (2004). Spatio-temporal difference in model outputs and parameter space as determined by calibration extent. In P. M. Atkinson, G. M. Foody, S. E. Darby, & F. Wu (Eds.), *Geodynamics* pp. 233–247. Boca Raton: CRC.
9. Dietzel, C., & Clarke, K. C. (2004). Spatial differences in multi-resolution urban automata modeling. *Transactions in GIS*, *8*(4), 479–492. doi:10.1111/j.1467-9671.2004.00197.x.
10. Dietzel, C., & Clarke, K. C. (2006). The effect of disaggregating land use categories in cellular automata during model calibration and forecasting. *Computers, Environment and Urban Systems*, *30*, 78–101. doi:10.1016/j.compenvurbsys.2005.04.001.
11. Goldstein, N. C., Candau, J. T., & Clarke, K. C. (2004). Approaches to simulating the "March of Bricks and Mortar". *Computers, Environment and Urban Systems*, *28*, 125–147. doi:10.1016/S0198-9715(02)00046-7.
12. Gustafson, E. J. (1998). Quantifying landscape spatial pattern: what is the state of the art? *Ecosystems*, *1*, 143–156. doi:10.1007/s100219900011.
13. Hagen, A. (2002). *Multi-method assessment of map similarity.* Paper presented at the 5th AGILE Conference on Geographic Information Science, Palma, Spain.
14. Hagen-Zanker, A. (2006). Map comparison methods that simultaneously address overlap and structure. *Journal of Geographical Systems*, *8*(2), 165–185. doi:10.1007/s10109-006-0024-y.
15. Herold, M., Goldstein, N. C., & Clarke, K. C. (2003). The spatiotemporal form of urban growth: measurement, analysis and modeling. *Remote Sensing of Environment*, *86*, 286–302. doi:10.1016/S0034-4257(03)00075-0.
16. Herold, M., Couclelis, H., & Clarke, K. C. (2005). The role of spatial metrics in the analysis and modeling of urban land use change. *Computers, Environment and Urban Systems*, *29*, 369–399. doi:10.1016/j.compenvurbsys.2003.12.001.
17. Jantz, C. A., Goetz, S. J., & Shelley, M. K. (2003). Using the SLEUTH urban growth model to simulate the impacts of future policy scenarios on urban land use in the Baltimore–Washington metropolitan area. *Environment and Planning B: Planning and Design*, *31*, 251–271. doi:10.1068/b2983.
18. Jantz, C. A., & Goetz, S. J. (2005). Analysis of scale dependencies in an urban land use change model. *International Journal of Geographical Information Science*, *19*(2), 217–241. doi:10.1080/13658810410001713425.
19. Klostermann, R. E. (1999). The what if? Collaborative planning support system. *Environment and Planning B: Planning and Design*, *26*, 393–408. doi:10.1068/b260393.
20. Li, X., & Yeh, A. (2000). Modeling sustainable urban development by the integration of constrained cellular automata and GIS. *International Journal of Geographical Information Science*, *14* (2), 131–152. doi:10.1080/136588100240886.
21. Li, X., & Yeh, A. (2004). Analyzing spatial restructuring of land use patterns in a fast growing region using remote sensing and GIS. *Landscape and Urban Planning*, *69*, 335–354. doi:10.1016/j.landurbplan.2003.10.033.
22. McGarigal, K., & Marks, B. J. (1995). *FRAGSTATS: spatial pattern analysis program for quantifying landscape structure*. Gen. Tech. Rep. PNW-GTR-351. Portland, OR: U.S. Department of Agriculture, Forest Service, Pacific Northwest Research Station. 122 p.
23. Pontius Jr., R. G. (2000). Quantification error versus location error in comparison of categorical maps. *Photogrammetric Engineering and Remote Sensing*, *66*(8), 1011–1016.
24. Pontiusx Jr., R. G., & Schneider, L. C. (2001). Land-cover change model validation by an ROC method for the Ipswich watershed, Massachusetts, USA. *Agriculture, Ecosystems & Environment*, *85*, 239–248. doi:10.1016/S0167-8809(01)00187-6.
25. Pontius Jr., R. G. (2002). Statistical methods to partition effects of quantity and location during comparison of categorical maps at multiple resolutions. *Photogrammetric Engineering and Remote Sensing*, *68*(10), 1041–1049.
26. 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. doi:10.1111/1467-9671.00159.
27. Pontius Jr., R. G., & Suedmeyer, B. (2004a). Components of agreement between categorical maps at multiple resolutions. In R. S. Lunetta, & J. G. Lyon (Eds.), *Remote Sensing and GIS Accuracy Assessment* (pp. 233–251). Boca Raton: CRC.
28. Pontius Jr., R. G., Huffaker, D., & Denman, K. (2004b). Useful techniques of validation for spatially explicit land-change models.

Ecological Modelling, *179*(4), 445–461. doi:10.1016/j.ecolmodel.2004.05.010.
29. Project Gigalopolis: Urban and Land Cover Modeling. URL: www.ncgia.ucsb.edu/projects/gig/project_gig.htm.
30. 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. doi:10.1016/S0198-9715(01)00014-X.
31. Sui, D. Z., & Zeng, H. (2001). Modeling the dynamics of landscape structure in Asia's emerging desakota regions: a case study in Shenzhen. *Landscape and Urban Planning*, *53*, 37–52. doi:10.1016/S0169-2046(00)00136-5.
32. Swets, J. A. (1986). Indices of discrimination or diagnostic accuracy: their ROC's and implied models. *Psychological Bulletin*, *99*, 100–17. doi:10.1037/0033-2909.99.1.100.
33. Torrens, P. M., & O'Sullivan, D. (2001). Cellular automata and urban simulation: where do we go from here. *Environment and Planning B: Planning and Design*, *28*, 163–168. doi:10.1068/b2802ed.
34. Verburg, P. H., de Nijs, T. C. M., Ritsema van Eck, J. R., Visser, H., & de Jong, K. (2004). A method to analyse neighborhood characteristics of land use patterns. *Computers, Environment and Urban Systems*, *28* (6), 667–690. doi:10.1016/j.compenvurbsys.2003.07.001.
35. White, R., & Engelen, G. (1993). Cellular automata and fractal urban form: a cellular modeling approach to the evolution of urban land use patterns. *Environment and Planning A*, *25*, 1175–1199. doi:10.1068/a251175.
36. Wu, F. (1998). Simulating urban encroachment on rural land with fuzzy logic controlled cellular automata in a geographical information system. *Journal of Environmental Management*, *53*, 293–308. doi:10.1006/jema.1998.0195.
37. Xian, G., Crane, M., & Steinwand, D. (2005). Dynamic modeling of Tampa Bay urban development using parallel computing. *Computers & Geosciences*, *31*(7), 920–928. doi:10.1016/j.cageo.2005.03.006.
38. Yang, X., & Lo, C. P. (2003). Modelling urban growth and landscape change in the Atlanta metropolitan area. *International Journal of Geographical Information Science*, *17*(5), 463–488. doi:10.1080/1365881031000086965.