Transactions in GIS, 2007, 11(1): 29–45

Research Article

Toward Optimal Calibration of the SLEUTH Land Use Change Model

Charles Dietzel and Keith C Clarke
Department of Geography
University of California – Santa Barbara

Abstract
SLEUTH is a computational simulation model that uses adaptive cellular automata to simulate the way cities grow and change their surrounding land uses. It has long been known that models are of most value when calibrated, and that using back-casting (testing against known prior data) is an effective calibration method. SLEUTH's calibration uses the brute force method: every possible combination and permutation of its control parameters is tried, and the outcomes tested for their success at replicating prior data. Of the SLEUTH calibration approaches tried so far, there have been several suggested rules to follow during the brute force procedure to deal with problems of tractability, most of which leave out many of the possible parameter combinations. In this research, we instead attempt to create the complete set of possible outcomes with the goal of examining them to select the optimum from among the millions of possibilities. The self-organizing map (SOM) was used as a data reduction method to pursue the isolation of the best parameter sets, and to indicate which of the existing 13 calibration metrics used in SLEUTH are necessary to arrive at the optimum. As a result, a new metric is proposed that will be of value in future SLEUTH applications. The new measure combines seven of the current measures, eight if land use is modeled, and is recommended as a way to make SLEUTH applications more directly comparable, and to give superior modeling and forecasting results.

1 Introduction

While the history of spatial modeling is filled with failed attempts to capture the dynamics of urban systems and subsequent land use change (Lee 1973, 1994), increased computational power and improved sources of spatial data have rekindled attempts to

Address for correspondence: Keith Clarke, Department of Geography, University of California-Santa Barbara, Santa Barbara, CA 93106, USA. E-mail: kclarke@geog.ucsb.edu

numerically model and simulate change. New techniques, including cellular automata (CA) and agent-based modeling, have shown potential in representing and simulating the complexity of the dynamic processes involved in urban growth and land use change, and can provide an additional level of knowledge and understanding of spatial and temporal change. One of the primary goals in the creation and use of these models is to provide an extra layer of insight to policy makers and planners; using the models to anticipate and forecast future changes or trends of development, to describe and assess impacts of future development, and to explore the potential impacts of different policies (Pettit et al. 2002, Verburg et al. 2002). The nature of this use for models requires that they be robust in every aspect of their operation – the data used, the calibration and mechanization of the model, and how change is forecast. Rarely are models tested with such rigor that weaknesses or flaws in these aspects are discovered, yet with the goal of impacting policy, models must be examined in more rigorous detail.

The calibration phase of modeling is one of, if not the most critical portions of a modeling exercise (Batty 1977, Batty and Xie 1994, Landis and Zhang 1998). It is during this time that the parameters which describe a system are determined, and upon which all forecasting and scenario simulation is based. In some spatial models, the 'goodness of fit' of a set of parameters is determined by a set of spatial metrics, each describing differently how accurately a set of parameters can replicate real-world historical patterns. While this provides a quantifiable measure of the ability of a parameter set to replicate historical patterns, there are too many spatial measures available to determine which measure is the appropriate one to use. This problem is compounded even further when a suite of metrics are calculated. An example of this is the suite of metrics that are calculated during the calibration of the SLEUTH urban growth model (Clarke et al. 1997), and the different metrics or combination thereof that researchers have used in different applications of the model (Silva and Clarke 2002, Yang and Lo 2003, Jantz et al. 2004). Robust modeling requires a clear consensus on what the measure of fit should be during calibration. To determine the optimal set of metrics for use in calibration of the SLEUTH model an exhaustive calibration parameter set was generated for three synthetic test data sets. Self-organizing maps (SOMs) (Kohonen 1995) were then used to explore the multi-dimensional parameter and metric spaces for multiple datasets that captured a wide array of spatial patterns. Analysis of these results led to the development of a new metric for use in future SLEUTH model calibrations.

2 The SLEUTH Model

The SLEUTH urban growth and land use change model is a cellular automaton (CA) model that has been widely used to model urbanization throughout various regions of the United States and the world (Clarke et al. 1997, Silva and Clarke 2002, Esnard and Yang 2002, Yang and Lo 2003, Jantz et al. 2004, Leão et al. 2004). The model has the ability to simulate urban/non-urban dynamics as well as urban-land use dynamics, although the latter has not been widely used; presumably due to the limitations of gathering consistently classified land use data. The dual ability has led to the development of two subcomponents within the framework of the model, one that models urban/non-urban growth, the urban growth model (UGM), and the other that models land use change dynamics (Deltatron). Regardless of which of these components is used, the model has the same calibration routine. The input of land use data during calibration activates the Deltatron part of SLEUTH. Due to computational complexity, land use was not a part of the testing in this study.

2.1 Required Data

SLEUTH is an acronym for the data required to calibrate and forecast in this urban growth model – **S**lope, **L**and Use, **E**xclusion, **U**rban, **T**ransportation, **H**illshade. The model requires topographic data in the form of Slope and Hillshade maps, although the hillshade is used only for visualization purposes, and does not play a role in determining model behavior. Land use with consistent classification for two time periods is needed to implement the Deltatron submodel, but land use data are not necessary to simulate urban growth. An Exclusion layer is used to place constraints on urban growth. Through the Exclusion layer, a user can specify where urban growth is allowed to occur, or where it is prohibited, for example in the ocean. This layer can also be a weighted layer so that ‘resistances’ against growth can be put in place in an attempt to slow or alter the rate of urbanization. Urban extent data is critical and necessary for this model. At least four different temporal layers are needed, showing the extent of urban areas at different points in time. These maps serve as the control points, against which the model is calibrated, and a goodness of fit is determined. The last layer required for using SLEUTH is Transportation. Historical maps of the transportation network show the evolution of the transportation network through time. Since different types of roads attract urban growth in different ways, roads in the transportation map are classified according to access. Roads classified with a value of 100 are generally high access roads such as freeways, interstates, and state routes. Primary local roads are given a weight of 50, and secondary roads are classified as 25. The creation of these input maps is typically done within a GIS, and then they are converted to digital images in GIF format which are the actual data used in the model.

2.2 Urban Growth Model (UGM)

Calibration of SLEUTH produces a set of five parameters (coefficients), which describe an individual growth characteristic, and that when combined with other characteristics, can describe several different growth processes. For this model, the transition rules between time periods are uniform across space, and are applied in a nested set of loops. The outermost of the loops executes each growth period, while an inner loop executes growth rules for a single year. Transition rules and initial conditions of urban areas and land use at the start time are integral to the model because of how the calibration process adapts the model to the local environment. Clarke and Gaydos (1998) describe the initial condition set as the ‘seed’ layer, from which growth and change occur one cell at a time, each cell acting independently of the others, until patterns emerge during growth and the ‘organism’ learns more about its environment. The transition rules that are implemented involve taking a cell at random and investigating the spatial properties of that cell’s neighborhood, and then urbanizing the cell, depending on probabilities influenced by other local characteristics (Clarke et al. 1997). Five coefficients (with values 0 to 100 as integers) control the behavior of the system, and are predetermined by the user at the onset of every model run (Clarke et al. 1997, Clarke and Gaydos 1998, Candau 2000). The parameters drive the four transition rules that simulate spontaneous (of suitable slope and distance from existing centers), diffusive (new growth centers), organic (infill and edge growth), and road influenced (a function of road gravity and density) growth. These parameters are:

1. *Diffusion* – Determines the overall dispersiveness of the outward distribution. This parameter controls the number of times that a pixel will be randomly selected for possible urbanization, and is calculated as:

$$\text{diffusion value} = (\text{diffusion coefficient} * 0.005) * \sqrt{((\text{number of rows})^2 + (\text{number of columns})^2)}$$

   This means that the maximum diffusion value will be half of the input image diagonal.
2. *Breed Coefficient* – The likelihood that a newly generated detached settlement will start on its own growth cycle.
3. *Spread Coefficient* – Controls how much contagion diffusion radiates from existing settlements.
4. *Slope Resistance Factor* – Influences the likelihood of development on steep slopes.
5. *Road Gravity Factor* – An attraction factor that draws new settlements towards and along roads.

2.4 Model Calibration and Determining 'Goodness of Fit'

By running the model in calibration mode, a set of control parameters is refined in the sequential 'brute-force' calibration phases: coarse, fine and final calibrations (Silva and Clarke 2002), although other methods of calibration, including the use of genetic algorithms have been suggested and tested (Goldstein 2004). Initially the model was calibrated using hierarchical spatial resolutions, beginning with data of coarser resolution, narrowing the range of the parameters that most accurately described the growth of the system, and then using a finer resolution to narrow the parameter values to one distinct set. Advances in computing power have allowed a more timely calibration, but it has been shown that using the hierarchical spatial resolution may lead to parameter sets that do not as accurately describe the growth of the system as a calibration at full data resolution (Dietzel and Clarke 2004). The calibration process is done in three stages: coarse, fine, and final. The coarse calibration begins with parsing the parameter space into five areas and using the values of 1, 25, 50, 75, and 100 for each of the five parameters. This gives 3,125 (5^5) different parameter sets that are tested as initial conditions to start the cellular model.

For each outcome, thirteen parameters are computed and used to determine the goodness of fit of the run to the known prior data. Runs returning the highest values indicate that within the ranges given, there is a region of the five-dimensional parameter space that may contain a peak in the model's simulation ability. Results from the coarse calibration are examined to determine the goodness of fit for each of the parameter sets. Narrowing of the parameter set can be based on a variety of different goodness of fit measures or their combinations. Despite numerous applications of SLEUTH all over the world, there is no clear consensus as to which metrics are the appropriate ones to use during the calibration process. In their application of SLEUTH to the Washington-Baltimore metropolitan area, Jantz et al. (2004) used the *compare*, *population*, and *Lee-Sallee* statistics, while in Atlanta, Yang and Lo (2003) used a weighted sum of all the metrics, and Silva and Clarke (2002) used only the *Lee-Sallee* metric in modeling Porto and Lisbon. In past cases, the *Lee-Sallee* (Lee and Sallee 1970) metric has often been used to determine which

Table 1 Metrics that can be used to evaluate the goodness of fit of the SLEUTH model

Metric Name	Description
Product	All other scores multiplied together
Compare	Modeled population for final year/actual population for final year, or IF $P_{modeled} > P_{actual}$ {1 – (modeled population for final year/actual population for final year)}.
Pop	Least squares regression score for modeled urbanization compared to actual urbanization for the control years
Edges	Least squares regression score for modeled urban edge count compared to actual urban edge count for the control years
Clusters	Least squares regression score for modeled urban clustering compared to known urban clustering for the control years
Cluster Size	Least squares regression score for modeled average urban cluster size compared to known average urban cluster size for the control years
Lee-Salle	A shape index, a measurement of spatial fit between the model's growth and the known urban extent for the control years
Slope	Least squares regression of average slope for modeled urbanized cells compared to average slope of known urban cells for the control years
% Urban	Least squares regression of percent of available pixels urbanized compared to the urbanized pixels for the control years
X-Mean	Least squares regression of average x_values for modeled urbanized cells compared to average x_values of known urban cells for the control years
Y-Mean	Least squares regression of average y_values for modeled urbanized cells compared to average y_values of known urban cells for the control years
Rad	Least squares regression of standard radius of the urban distribution, i.e. normalized standard deviation in x and y
F-Match	A proportion of goodness of fit across landuse classes. {#_modeled_LU correct/(#_modeled_LU correct + #_modeled_LU wrong)}

parameter sets best describe the replication of the historical datasets. *Lee-Sallee* is the ratio of the intersection and the union of the simulated and actual urban areas, but others, including the *compare* statistic and *population* statistics, have been used. Table 1 provides a description of the 14 metrics that can be used to determine the goodness of fit of model calibration; they have a value range of 0.0 to 1.0, with 1 being a perfect fit.

After determining the parameter set that best fits the historical data, a range of values bracketing that set of parameters is selected and the calibration is repeated. The goodness of fit of the second calibration is evaluated and an even narrower range of parameters is selected ideally with unit increments. The best fit of the parameters from this third calibration are then the parameters that are used in forecasting urban growth and land use change.

3 Determining the Optimal SLEUTH Metric (OSM)

To determine the optimal metrics for model calibration, three datasets were exhaustively calibrated (meaning that every single combination of parameter values was tested)

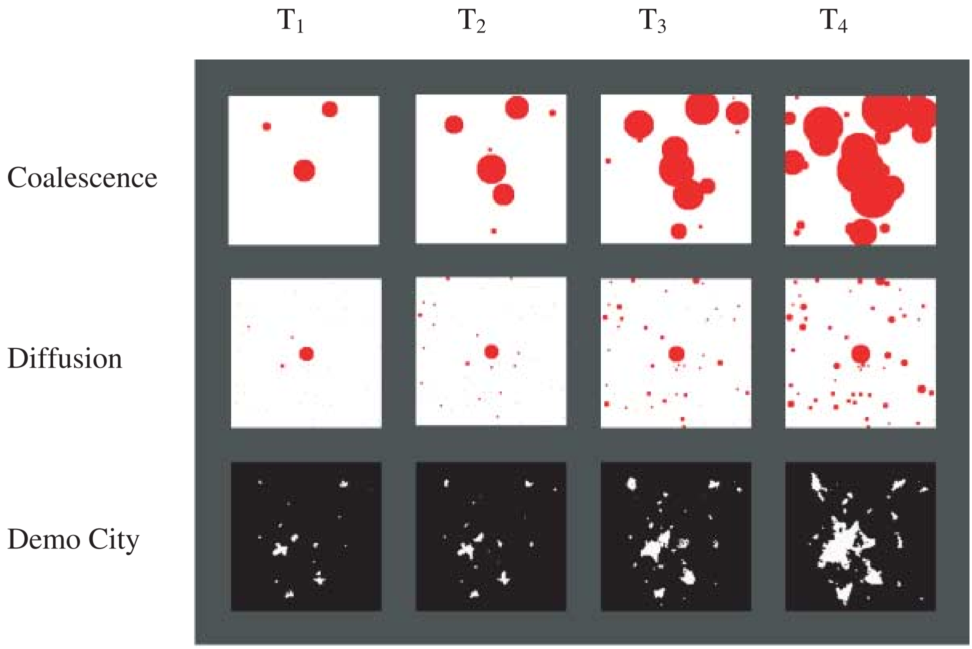

Figure 1 Three spatial datasets used in determining the optimal metrics during calibration of the SLEUTH urban growth model. The top dataset represents coalescent growth, whereby urban areas grow together. The middle dataset depicts a diffusive system where urban growth diffuses outwards from a distinct core. The lower dataset represents both diffusion and coalescence of an urban system

across their parameter spaces, and the resultant metrics analyzed. Two of the data sets were representative of the two major spatial processes in urban growth, diffusion and coalescence (Dietzel et al. 2005). A third dataset, the "demo city" data available with SLEUTH as a test calibration data set, was used as a combination of the two with greater (though still hypothetical) realism (Figure 1). Since these data only represent urban growth, the land use-related *F-match* statistic was not calculated or evaluated. In addition to this, the *product* metric was not used in the analysis since it is directly dependent on all of the other metrics. It should be assumed that whenever land use data is used, that the *F-match* statistic be used in calibration because it is specific to land use change.

For the diffusive and coalescent datasets, no road or slope data were used: growth occurred on an isotropic plane. Because of this, the slope and road gravity parameters were held constant with a value of 1. The *% slope* metric was not included in the analysis because there was no slope variation. These two datasets were exhaustively calibrated for the diffusion, breed, and spread parameters (101^3 or 1,030,301 different parameter combinations), calculating all the metrics for each parameter set. For the demo city data set, the parameter space was parsed into blocks of five (meaning that the parameter values were tested in increments of five, from 0 to 100), and then exhaustively calibrated (21^5 or 4,084,101 parameter combinations), calculating all of the metrics. Since the demo city dataset that comes with SLEUTH has both slope and transportation data, the slope and road gravity parameters were used during the calibration of this dataset, and the % *slope* metric was calculated.

Upon running the calibration of each of these datasets, the result was an 18-dimensional dataset that consisted of 13 metrics and five parameter values, with over a million entries in each dimension. The sheer size of this dataset is not readily handled by most of today's statistical analysis packages. Due to this multitude of data, and the uncertainty in understanding the meaning of the parameter values, standard statistical methods for examining the relationships between variables were not used. Instead it was decided that data mining techniques, namely the self-organizing map (SOM) algorithm (Kohonen 1995), would serve as a better analytical tool for determining which metrics were related to one another, and which ones would be best in evaluating model fit.

3.1 SOM Creation and Clustering

The self-organizing map (SOM) is created using a neural network algorithm that is based on unsupervised learning and is used to cluster data without knowing the class memberships of the data (Kohonon 1995). This technique of data analysis and mining has been widely applied with regard to data in other disciplines, including engineering, economics, sociology, and library sciences (Kaski 1997), and provides a highly visualized approach to see rough relations between variables (Vesanto 1997).

SOMs consist of a regular two-dimensional grid of neurons connected in the form of either hexagons or rectangles. The grid of neurons is typically rectangular, but other arrangements including toroids have also been used (Andreu et al. 1997). As in the case of CA, the immediate neighbors are considered to be in the neighborhood of the adjacent neuron. There are various shapes for these neighborhoods; typically the bubble and Gaussian neighborhoods are used, but there are several other variations. The number of neurons in the grid is set at the beginning of the SOM creation, and this granularity does have an impact on the accuracy and generalization capability of the SOM (Simula et al. 1999). Skupin (2002) created a relatively large SOM in order to "create a base configuration in which the SOM recognizes and geometrically preserves many of the finer differences" among the training vectors. At the other extreme, a very small SOM more suitable for clustering might have every neuron acting as one cell. With our purpose being to visually examine the relations among variables as revealed on the component planes, we chose to use 50,000 nodes, the maximum number of vectors possible in the software used (Viscovery SOMline Plus). This necessarily involved sampling the parameter space output from the full exploration set. Training of the SOM is conducted in two stages. In the first stage a relatively large neighborhood radius is used to train the SOM. During the second stage the radius is decreased to train the SOM in a more precise manner.

Visualization is one of the major advantages of the SOM. Results are presented in a manner that make them readily understandable and in an easy method from which to draw out relationships between variables. There are two main methods for the visual analysis of SOMs: component and cluster analysis. Each of the independent variables can be viewed as its own component plane. This provides information on the distribution of component values. By viewing several component planes together at the same time, it is easy to see simple correlations among different components. Additionally, all of the components can be visualized together by displaying the distances between the vectors. This distance matrix is termed the *U*-matrix (Ultsch 2003). In the *U*-matrix, distances between the weight vectors of map units and their neighborhood are calculated. The map can then be displayed in two or three dimensions, and the relative distances between adjacent map units can be seen. While there is no clear-cut method aside from visual interpretation of SOMs, visualization of the component planes readily allowed for the identification of relationships between the metrics used in calibrating SLEUTH, making it possible to determine which were optimal for use in calibration.

3.2 Generating SOMs

Calibration of the model for the three datasets resulted in extremely large derivative datasets, through which to mine (ranging from 1 million to 4.1 million entries for 18 dimensions). The magnitude of these datasets made them difficult to manipulate, and no software that could prepare the data for the SOM routines was available for a dataset of that size. Due to these limitations, and because of the maximum number of vectors possible in the software, three random samples of the metric values for 50,000 different parameter sets were taken from the results of the coalescence, diffusion, and demo city SLEUTH calibrations (nine samples total, three for each dataset). The separate samples were taken to demonstrate that there were no significant differences among SOMs from the different samples. Vectors on the metric values for *compare*, *population*, *edges*, *clusters*, *cluster size*, *Lee Sallee*, *% urban*, *X-mean*, and *Y-mean* were then used as input into Viscovery SOMine Plus (Eudaptic Software).

One method of evaluating the quality of the resulting SOM component plane map is to calculate the quantization error over the input samples. Viscovery SOMine computes two values, the average quantization error and the normalized distortion (Vesanto and Sulkava 2002). The average quantization error represents the average distance between the best matching units and the sample data vectors, i.e. the average square distance of all records in the data set to their respective projections onto the map. The normalized distortion is the degree of fit of the map with respect to the shape of the data distribution. These values are reported in section 4.

Before data were put into the SOM, a test dataset from the coarse calibration was used to determine the best values for the number of nodes and the map tension. We used the maximum number of nodes (20,000) and established that the best visual interpretation corresponded to a tension setting (influence radius of the neighborhood). The SOMs were then trained with 2,000 nodes, after preliminary tests were tried with 20,000 and 5,000. In the final application, where clustering was of greater importance, we used a hexagonal grid, with 2,000 nodes in a mesh about 40 cells in y and about 50 cells in x. The value of increasing the number of nodes was negligible, so 2,000 were used. At this size map tension was set at 0.5. Decreasing the tension resulted in a higher quantization error or normalized distortion, while increasing it produced maps that appeared overly smooth and regular. Average training time for the SOMs was 39 minutes on a 2.5 GHz PC.

4 Results

Calibration of the three datasets used in the SLEUTH model led to data too extensive for running the SOM algorithm in a timely manner with reasonable computing resources. To overcome the computational limitations, three random samples of the

Table 2 Summary statistics of the results from coalescence dataset calibration

	Mean	Standard Deviation	Minimum	Maximum
Compare	0.9258	0.0967	0.7102	1.0000
Population	0.6897	0.3559	0.0000	1.0000
Edges	0.3557	0.2572	0.0000	1.0000
Clusters	0.9151	0.1679	0.0000	1.0000
Cluster_size	0.2935	0.0585	0.1653	0.5472
Leesalee	0.9258	0.0967	0.7102	1.0000
%urban	0.0367	0.0670	0.0000	0.9999
Xmean	0.9425	0.0720	0.7843	1.0000
Ymean	0.4861	0.1725	0.0910	1.0000
Rad	0.9258	0.0967	0.7102	1.0000

Table 3 Summary statistics of the results from diffusion dataset calibration

	Mean	Standard Deviation	Minimum	Maximum
Compare	0.7735	0.1644	0.4844	0.9997
Population	0.5678	0.3219	0.0000	1.0000
Edges	0.7595	0.3159	0.0000	1.0000
Clusters	0.4100	0.1955	0.0000	1.0000
Cluster_size	0.0773	0.0240	0.0483	0.3197
Leesalee	0.7735	0.1644	0.4844	0.9997
%urban	0.6860	0.3186	0.0000	1.0000
Xmean	0.8232	0.1288	0.5732	0.9978
Ymean	0.0521	0.0790	0.0238	1.0000
Rad	0.7735	0.1644	0.4844	0.9997

output statistics were generated, and comparison between the samples shows that there was no bias between any of the samples (Tables 2–4).

In training the SOM, both the quantization error (the average square distance of all records in the data set to their respective projections onto the map) and the normalized distortion (the fitting of the map with respect to the shape of the data distribution) were calculated as a measure of the quality of the SOMs (Table 5, Figures 2 and 3). For the diffusion SLEUTH dataset, the mean quantization error was 0.0333344, and the normalized distortion of the SOM was 0.05597. For the coalescence and demo city SLEUTH datasets, the mean quantization error was 0.02294 and 0.055534, with a normalized distortion of 0.040915 and 0.085265.

The SOM algorithm was run for one of the three samples for each of the three SLEUTH datasets (Figures 4–6). From visual inspection we concluded that there was no measurable difference between the three different samples taken from each of the three SLEUTH datasets, but that there were similarities among the three SLEUTH datasets. The similarities are discussed in the next section.

Table 4 Summary statistics of the results from demo city calibration

	Mean	Standard Deviation	Minimum	Maximum
Compare	0.1388	0.0702	0.1158	0.9969
Population	0.6972	0.1692	0.5544	1.0000
Edges	0.6197	0.2559	0.0000	1.0000
Clusters	0.8170	0.2838	0.0000	1.0000
Cluster_size	0.6078	0.3090	0.0000	1.0000
Leesalee	0.0918	0.0414	0.0645	0.4279
Slope	0.7360	0.1424	0.0001	1.0000
%urban	0.7205	0.1556	0.5695	1.0000
Xmean	0.6342	0.1908	0.0000	1.0000
Ymean	0.8973	0.0888	0.0000	1.0000
Rad	0.7608	0.1296	0.6512	1.0000

Table 5 Final quantization error and normalized distortion for data that the SOM algorithm was applied from the diffusion, coalescence, and demo city SLEUTH datasets

SLEUTH Dataset	Sample	Quantization Error	Normalized Distortion
Diffusion	1	0.033291	0.055780
	2	0.033423	0.056271
	3	0.033318	0.055859
Coalescence	1	0.023114	0.040988
	2	0.022800	0.040942
	3	0.022905	0.040814
Demo City	1	0.055079	0.084839
	2	0.055809	0.085648
	3	0.055713	0.085309

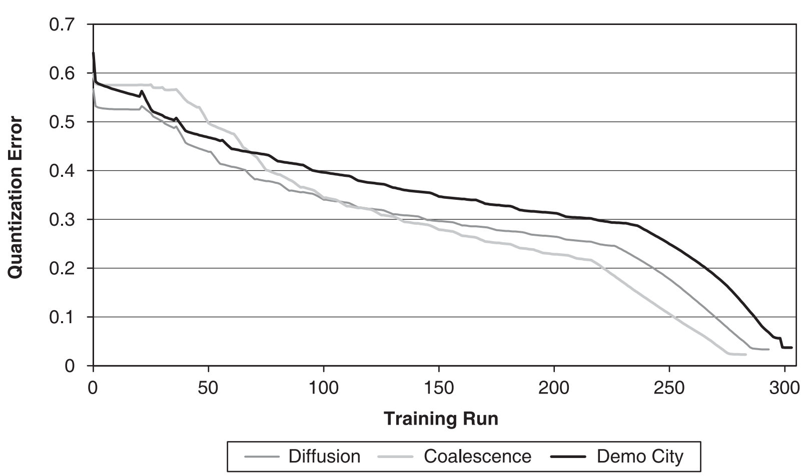

Figure 2 Trajectory of the mean quantization error during the SOM training process for the calibration results from the diffusion, coalescence, and demo city SLEUTH datasets

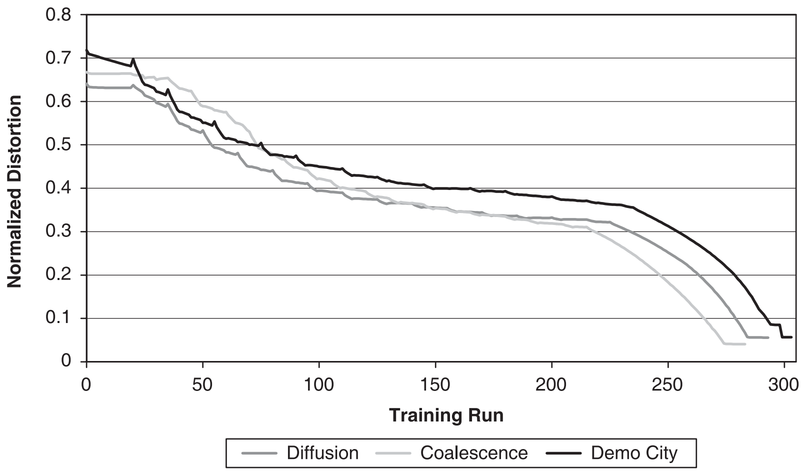

Figure 3 Trajectory of the mean normalized distortion during the SOM training process for the calibration results from the diffusion, coalescence, and demo city SLEUTH datasets

Table 6 Summary of whether or not spatial metrics were necessary in the calibration of the SLEUTH model

	SLEUTH Dataset			
Metric Name	Diffusion	Coalescence	Demo City	Metric Needed
Compare	Yes	Yes	Yes	Yes
Population	Yes	Yes	Yes	Yes
Edges	Yes	Yes	Yes	Yes
Clusters	Yes	Yes	Yes	Yes
Cluster Size	Yes	Yes	No	No
Lee-Sallee	No	Yes	No	No
Slope	N/A	N/A	Yes	Yes
% Urban	No	No	No	No
X-mean	Yes	Yes	Yes	Yes
Y-mean	Yes	Yes	Yes	Yes
Rad	No	No	No	No
F-Match	N/A	N/A	N/A	If LU is modeled

5 Discussion and Conclusions

This section discusses the interpretation of the SOM component plane visualizations. Visual interpretation of the SOMs ultimately revealed that of the 12 spatial metrics that SLEUTH calculates during the calibration process, only seven are needed to determine the goodness of fit for the majority of urban systems (Table 6). If land use is modeled in addition to urban growth, the *F-match* metric is additionally required since it is specific to land use.

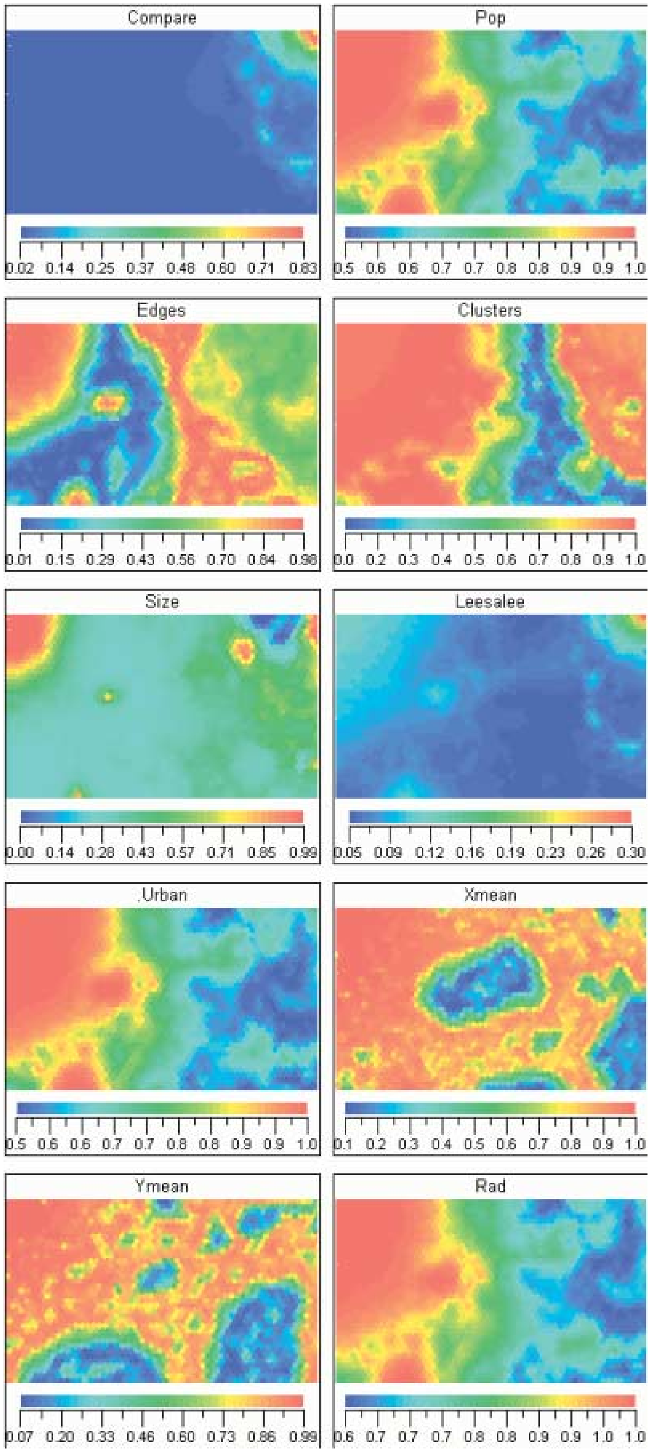

Figure 4 Component planes from the self-organizing maps of the metric space of the diffusion SLEUTH dataset

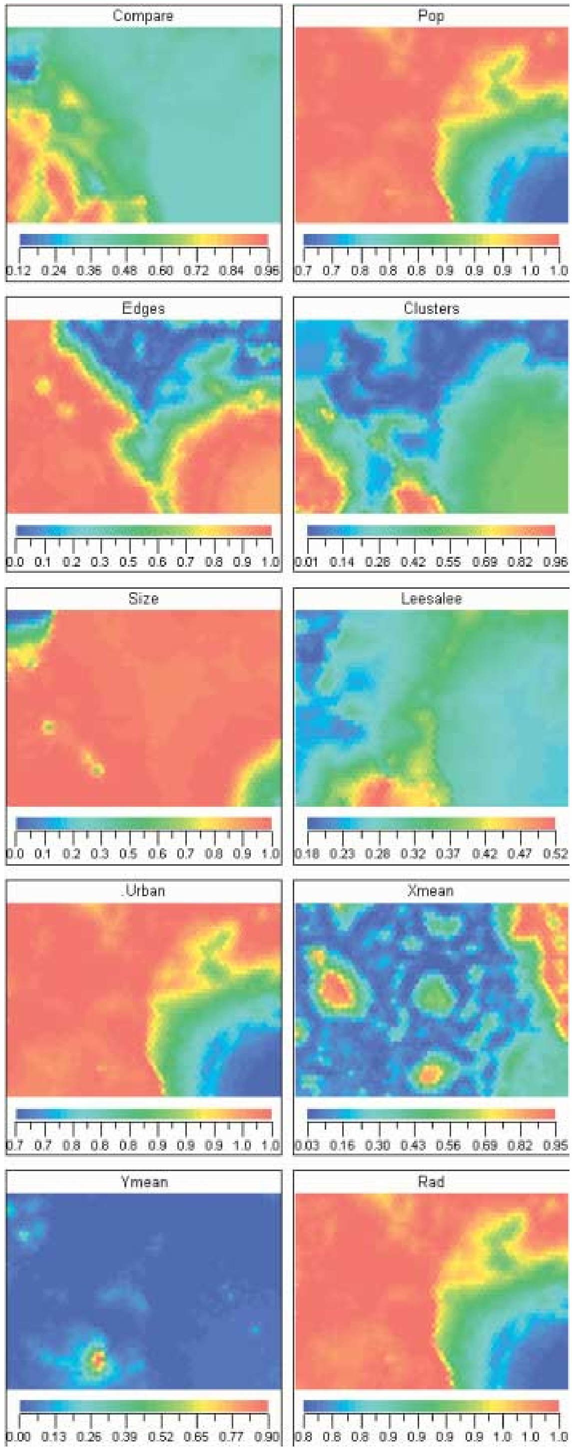

Figure 5 Component planes from the self-organizing maps of the metric space of the coalescence SLEUTH dataset

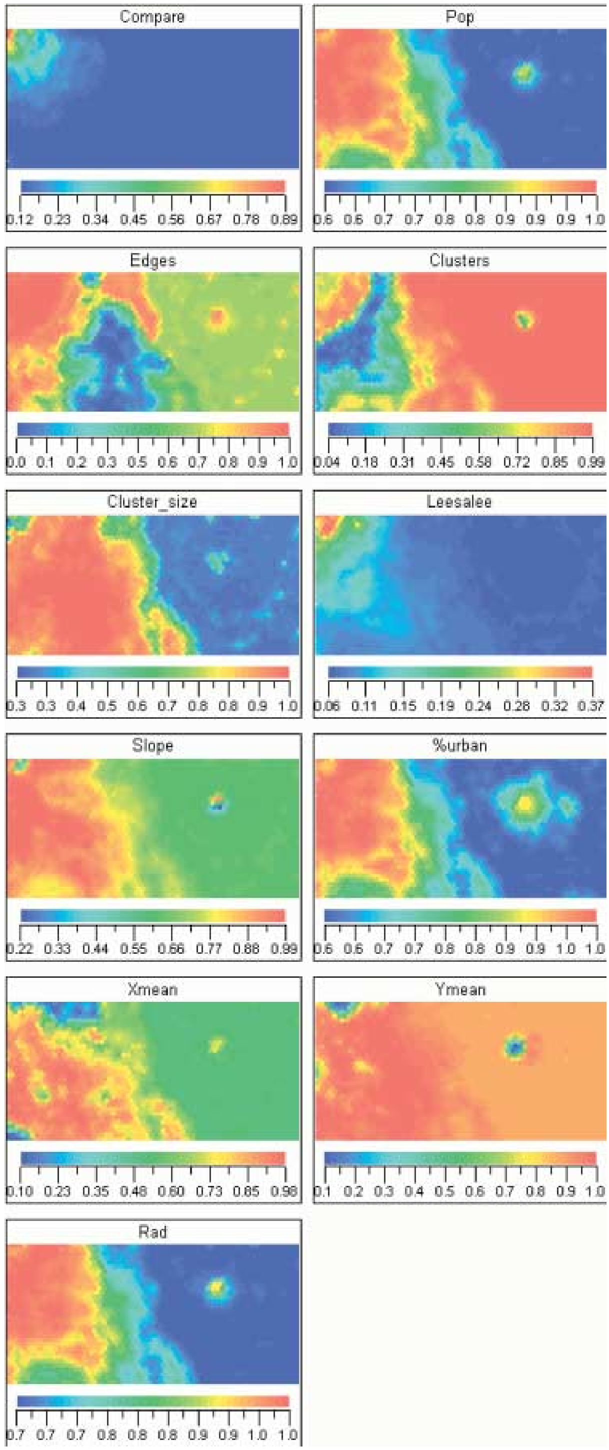

Figure 6 Component planes from the self-organizing maps of the metric space of the demo city SLEUTH dataset

Determination of which metrics were necessary was done through visual evaluation of the resultant component planes from each SOM, for each sample from each SLEUTH dataset, then comparing between the three datasets to determine what the necessary metrics were. If the interpretation showed that there were multiple metrics that were similar, then the metric that was first calculated by SLEUTH was deemed necessary, and the redundant metric was disregarded. The choice of the first metric was arbitrary, and the interchangeability within the redundancies was not explored.

For the diffusion dataset, the *compare* metric and the *Lee-Sallee* produced similar results, so only the *compare* metric is needed in future work. Additionally the *population*, *% urban*, and *rad* metrics had nearly identical component planes, so only the *population* metric was necessary. In the case of the coalescence dataset, the *population*, *% urban*, and *rad* metrics had similar component planes, so only the *population* metric was needed. The demo city dataset had comparable component planes for the *compare* and *Lee-Sallee* metrics, as well as for the *population*, *cluster size*, *% urban*, and *rad* metrics.

Based on the three datasets examined, which are believed to comprise the major spatial patterns of urban growth, a goodness of fit metric, known as the OSM (optimal SLEUTH metric, the product of the *compare*, *population*, *edges*, *clusters*, *slope*, *X-mean*, and *Y-mean* metrics) will provide the most robust results for SLEUTH calibration. *Cluster size* and *Lee-Sallee* were not included in this new metric because it was not clear from the three datasets if these metrics produced unique results from other metrics that SLEUTH calculates. This uncertainty led to them not being included in the suggested metric.

Development of the OSM for use in the calibration of SLEUTH was the first attempt to reach a clear conclusion as to what the best goodness of fit measure is for this model. The method used attempted to evaluate the model under the major spatial processes that are seen in urban growth data, diffusion and coalescence, and should be viewed as a metric necessary for calculation during the calibration process.

While prior research (Silva and Clarke 2002, Yang and Lo 2003, Dietzel and Clarke 2004, Jantz et al. 2004) has suggested the use of a variety of metrics, it is now clear that they have not been the optimal ones for determining a rigorous goodness of fit measure. In most cases (Yang and Lo 2003, Dietzel and Clarke 2004, Jantz et al. 2004), the method for determining goodness of fit was redundant, including two or more metrics that were closely related, possibly biasing the results. Other cases (Silva and Clarke 2002) have used too few metrics in evaluating the goodness of fit, not including others that would have made the calibration process more robust. Yet the development of the OSM and the conclusion that one measure for the goodness of fit will suffice should not be viewed as undermining previous research, as each has made its own contribution to the development and use of the SLEUTH model.

This research has further demonstrated the importance of the calibration procedure in spatial modeling, and how often it is a phase that is overlooked or regarded as being insignificant for major research. While the ambition of modelers is to use their models to impact policy and change, the models themselves must first be rigorously tested before they are ready. This is no more evident than in the case of the SLEUTH model which has been under development for a decade (Clarke et al. 2006), yet it has taken this long to fully understand the calibration procedure and develop a faster and more robust calibration method (Goldstein 2004). Only once there is complete understanding of the mechanical nature of a model can it be used properly and reach its promise of "honesty in modeling."

Acknowledgements

This research was funded in part by the National Science Foundation Doctoral Dissertation Grant Program (Award 0424916) and by the Public Policy Institute of California. Both authors greatly appreciate the generosity of both funding organizations.

References

Andreu G, Crespo A, and Valiente J M 1997 Selecting the toroidal self-organizing feature maps (TSOFM) best organized to object recognition. In *Proceedings of the ICNN-97 International Conference on Neural Networks*. Piscataway, NJ, IEEE: 1341–6

Batty M 1977 *Urban Modelling*. Cambridge, Cambridge University Press

Batty M and Xie Y 1994 Modeling inside GIS: Part 2, Selecting and calibrating urban models using Arc/Info. *International Journal of Geographical Information Systems* 8: 470–541

Candau J 2000 Calibrating a cellular automaton model of urban growth in a timely manner. In Parks B O, Clarke K C, and Crane M P (eds) *Proceedings of the Fourth International Conference on Integrating Geographic Information Systems and Environmental Modeling: Problems, Prospects, and Needs for Research*, Banff, Alberta. Boulder, CO, University of Colorado, Cooperative Institute for Research in Environmental Science: CD-ROM

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 Geographic Information Science* 12: 699–714

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

Clarke K C, Gazulis N, Dietzel C K, and Goldstein N C 2006 A decade of SLEUTHing: Applications of a cellular automaton land use change model. In Fisher P F (ed). *Classics from IJGIS: Twenty years of the International Journal of Geographical Information Science and Systems*. Boca Raton, FL, CRC Press: 413–26

Dietzel C K and Clarke K C 2004 Spatial differences in multi-resolution urban automata modeling. *Transactions in GIS* 8: 479–92

Dietzel C K, Herold M, Hemphill J J, and Clarke K C 2005 Spatio-temporal dynamics in California's Central Valley: Empirical links to urban theory. *International Journal of Geographical Information Science* 19: 175–95

Esnard A M and Yang Y 2002 Descriptive and comparative studies of 1990 urban extent data for the New York Metropolitan Region. *URISA Journal* 14: 57–62

Goldstein N C 2004 Brains vs. Brawn: Comparative strategies for the calibration of a cellular automata–based urban growth model. In Atkinson P, Foody G, Darby S, and Wu F (eds) *GeoDynamics*. Boca Raton, FL, CRC Press: 00–00

Jantz C A, Goetz S J, and Shelley M K 2004 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* 31: 251–71

Kaski S 1997 Data Exploration Using Self-organizing Maps. Unpublished Ph.D. Dissertation, Department of Computer Science and Engineering, Helsinki Technical University

Kohonen T 1995 *Self-Organizing Maps*. Berlin, Springer Series in Information Science No. 30

Landis J and Zhang M 1998 The second generation of the California urban futures model: Part 2, Specification and calibration results of the land use change submodel. *Environment and Planning B* 25: 795–842

Leão S, Bishop I, and Evans D 2004 Spatial-temporal model for demand allocation of waste landfills in growing urban regions. *Computers, Environment and Urban Systems* 28: 353–85

Lee D B 1973 Requiem for large-scale models. *Journal of the American Institute of Planners* 39: 163–78

Lee D B 1994 Retrospective on large-scale urban models. *Journal of the American Planning Association* 60: 35–40

Lee D R and Sallee G T 1970 A method of measuring shape. *The Geographical Review* 60: 555–63

Pettit C, Shyy T-K, and Stimson R 2002 An on-line planning support system to evaluate urban and regional planning scenarios. In Geertman S and Stillwell J (eds) *Planning Support Systems in Practice*. Heidelberg, Springer-Verlag: 331–47

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

Simula O, Vesanto J, Alhoniemi E, and Hollmen J 1999 Analysis and modeling of complex systems using the self-organizing map. In Kasabov N and Kozuma R (eds) *Neuro-Fuzzy Techniques for Intelligent Information Systems*. Heidelberg. Springer-Verlag Studies in Fuzziness and Soft Computing No. 30: 3–22

Skupin A 2002 A cartographic approach to visualizing conference abstracts. *IEEE Computer Graphics and Applications* 22: 50–8

Ultsch A 2003 *U-matrix: A Tool to Visualize Clusters in High Dimensional Data*. Marburg, Germany, University of Marburg, Department of Mathematics and Computer Science Report No. 10

Verburg P H, Soepboer W, Veldkamp A, Limpiada R, Espaldon V, and Mastura S S A 2002 Modeling the spatial dynamics of regional land use: The CLUE-S model. *Environmental Management* 30: 391–405

Vesanto J 1997 *Data Mining Techniques Based on the Self-organizing Map*. Helsinki, Helsinki Technical University, Department of Engineering Physics and Mathematics Report No. 71

Vesanto J and Sulkava M 2002 Distance matrix based clustering of the Self-Organizing Map. In Dorronsoro J R (ed) *Artificial Neural Networks*. Heidelberg, Springer-Verlag Lecture Notes in Computer Science No. 2415: 951–6

Yang X and Lo C P 2003 Modelling urban growth and landscape change in the Atlanta metropolitan area. *International Journal of Geographical Information Science* 17: 463–88