Exploring the DNA of Our Regions: Classification of Outputs from the SLEUTH Model

Nicholas Gazulis and Keith C. Clarke

University of California, Santa Barbara, 3611 Ellison Hall, University of California, Santa Barbara, CA, 93106
{ngazu, kclarke}@geog.ucsb.edu

Abstract. The SLEUTH urban growth model is a cellular automata model that has been widely used by geographers to examine the rural to urban transition as a physical process and to produce forecasts of future urban growth [1]. Previous SLEUTH applications have generally been limited to individual model applications, with little to no comparison of model results [2]. Building upon research by Silva and Clarke [3], and borrowing from their metaphorical comparison of urban growth characteristics to genetic DNA, this research distills a combination of actual city and model behavior in a controlled environment to provide for comparisons between disparate model applications. This work creates a digital "petri dish" capable of producing normalized model forecasts from previously incomparable results. Results indicate that despite the inherent differences between actual model results, sufficient similarities were observed among the forecasts to warrant the creation of an urban behavioral taxonomy, providing for direct comparison of the results.

1 Introduction

Cellular Automata (CA) models are increasingly being used for representing geographical processes, including many applications within the field of urban and regional modeling [1], [4]. Spatial processes, such as urban growth, exploit the natural analogy between two-dimensional CA and time-sequenced grid representations of two-dimensional geographic space. As geographer Waldo Tobler realized [5], the grid cells of a CA lattice can represent the "state" in areas of land while the lattice of the CA can foster geographical processes such as distance decay and spatial autocorrelation. While the idea of simulation with CA in the field of geography can be traced to Tobler [5], a more formal declaration of the use of CA applicability for representing urban systems was by Couclelis [6], leading to a major new modeling paradigm in recent years [7]. More recently, CA models have broadened to multiple states, and so to land-use change modeling [8].

Innovation within the computer and geographical sciences, coupled with increased access to quality and affordable remotely sensed data, has led to the use of these new urban growth models in both a policy and theoretical context [9]. Responding to heavy criticism of the first generation of urban computer models [10], [11], CA models have demonstrated practical success in urban planning. Due to data-driven

S. El Yacoubi, B. Chopard, and S. Bandini (Eds.): ACRI 2006, LNCS 4173, pp. 462 – 471, 2006.
© Springer-Verlag Berlin Heidelberg 2006

issues, such as inconsistent scale and resolution, urban and regional models are generally limited to specific policy situations where little emphasis is placed on comparisons between successive applications. Urban areas exhibit extremely different characteristics due to the complexity of the processes underlying urban growth. Consequently, modeling results produced from heterogeneous applications are generally incomparable [12].

Silva and Clarke [3] suggest that despite overwhelming differences between urban areas, there exist fundamental elements that are common to each urban area. This variability manifests itself in unique patterns of urban growth for each particular city as determined by the local environment or site. Many of these common elements relate to an area's particular dependence on transportation, how technologically feasible construction is on steep slopes, to what extent new urban centers develop within a system, how likely new spreading centers are to develop their own growth cycle, and how quickly spreading centers are to grow. Silva and Clarke [3] further suggest that many of these common elements can be empirically quantified for individual urban systems. The reduction of the characteristics that describe an urban area's uniqueness bears a resemblance to the biological notion of genetic DNA, a complete set of which fully describes a living organism's growth and development cycle, and as such, can metaphorically be considered to be the "DNA of our Regions". Like individual creatures, all cities are unique, yet share common building-blocks that permit replication and growth.

Given the DNA analogy, an experiment was created to distill a combination of actual city and model behavior in a controlled environment out of the data-dependent context of typical applications. DNA fragments were selected from cities and then grown under controlled circumstances in a digital "Petri dish." As in the work by Silva and Clarke [3], the SLEUTH urban growth and land use change model was used to quantify differences among worldwide urban areas. To do this, two sets of input data were used, including an anisotropic plane representing geographic variability and individual parameter sets as fit to various real cities. The anisotropic plane was held fixed throughout the experiments while only the SLEUTH control parameter sets were varied. The overall goal of this work was to create an experiment that allowed for comparison of previously incomparable results. As a means for comparing the results, a simple taxonomy was created based on visual and quantitative model results.

2 The SLEUTH Model

SLEUTH is capable of modeling the complex dynamics of any urban growth or land use change system given a set of historical input data. SLEUTH is an acronym for the six required data inputs, **S**lope, **L**anduse, **E**xclusion, **U**rban extent, **T**ransportation,and **H**illshade, and simulates land use dynamics as a physical process [13].

During forecasting with SLEUTH, the model is initialized with the most recent data as the "seed" layer. SLEUTH then executes a finite set of transition rules that influence state changes within the CA. The transition rules involve selecting cells at random and investigating the spatial properties of that cell's neighborhood. Based on an urbanization probability derived from the local characteristics of a particular cell, that cell is either urbanized or not urbanized. Monte Carlo simulation is employed to reduce stochastic bias, and it has been shown that a large number of iterations does not always result in improved results [14]. As such, users typically define between 15 and 30 iterations.

Before forecasting, the model must account for the physical differences that exist among individual study areas. To do so, SLEUTH employs a calibration routine that examines the historical data input to derive a set of parameters representing past urbanization trends for each unique region. As the CA iterates, a dozen statistical descriptors are computed that relate model behavior to the known historical data. The calibration phase of SLEUTH produces a set of five coefficients, each of which describes an individual growth characteristic of an urban area, plus their statistical goodness of fit to the historical data. A complete set of five calibration coefficients (each with an integer value ranging from 0 to 100) influences the degree to which each of the four growth rules influences urban growth in the system. These coefficients include:

1. *Dispersion* – controls the overall dispersive nature of the distribution.
2. *Breed* – determines the likelihood that an urbanized cell will start its own growth cycle.
3. *Spread* – determines the likelihood that the pixels that comprise a new spreading center will continue to generate new urban pixels.
4. *Slope* – influences the likelihood that a cell will be urbanized on a slope.
5. *Road Gravity* – a factor that encourages growth along the road network.

This set of parameters drives the four transition rules that govern urban growth within the system, which simulate *spontaneous* growth in suitable urban areas, *diffusive* growth in new spreading centers, *organic* growth in infill and edge areas, and *road-influenced* growth along the transportation network.

Calibration of the model is based on comparing model output and initial model inputs for a variety of parameter combinations. The model is initialized with the earliest available time period and "forecasts" urban extent using a coefficient set for the time period corresponding to the distance between the first and last data inputs. Images of urban extent are produced using many different parameter combinations and compared to the control data available for "goodness of fit". The degree of similarity between the simulated images and the control years is determined through a set of metrics that are calculated and stored in a log file. The analyst must examine the log file to determine the optimal set of parameters based on the calculated metrics, deducing which set of parameters produces an image that most closely resembles the control data images. Recent work has determined an optimal metric of fit, known as the Optimal SLEUTH Metric (OSM) specifically for use in determining best fit [15]. As a substitute for an exhaustive search, SLEUTH employs a 'Brute Force' method of coefficient optimization, which explores the parameter space in successively finer intervals. This structured brute force approach has been shown to reduce model overfit. Computation time is still a major factor in calibration, and other methods have been explored such as genetic algorithms [16].

3 Input Data

The data inputs for this research are twofold, including (1) a set of simulated images, and (2) a set of calibration parameters derived from the twenty SLEUTH model applications in the data repository for which parameters were reported.

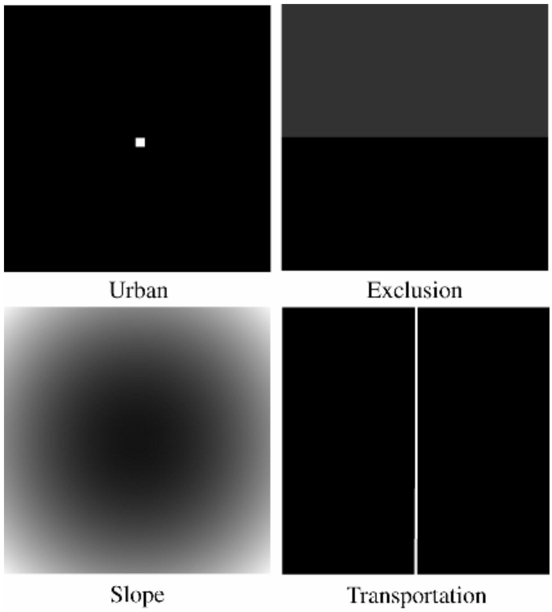

Fig. 1. Inputs: GIS images used in exploring the SLEUTH model parameters

For the first input, Nystuen [17] suggests that geographical problems must be assessed in a uniform representation of abstract space devoid of geographic variability, a surface he refers to as the isotropic plane. Contrary to Nystuen's [17] experimental isotropic data space, this work proposes that the isotropic plane, while preferable for some applications, is not preferable for others, and as such, cannot be considered to be a uniform standard for the modeling of geographic processes.

Due to the continually variable conditions of the Earth's surface, urban areas necessarily develop under sub-optimal conditions. As a result, the reduction of urban growth behavior into a set of parameters requires a variable surface from which a forecast can be derived. Otherwise, modeling results would only yield information about the overall spread and growth rates of the system. To allow for a robust comparison of the results, input data was created that mimicked the variable conditions experienced in actual urban systems.

The initial urban input consists of a single urban cell in the middle of the image with all other cells beginning as non-urban (see figure 1). The exclusion layer is divided lengthwise from east to west, with the southern portion bearing no resistance

Table 1. SLEUTH parameters derived from studies documented in the SLEUTH repository

Application Location	Diffusion	Breed	Spread	Slope	Road
Atlanta, GA	55	8	25	53	100
Austin, TX	47	12	47	1	59
Chiang Mai, Thailand	1	4	88	1	25
Colorado Front Range	11	35	41	1	91
Houston, TX	1	3	100	22	17
Lisbon, Portugal	19	70	62	38	43
Mexico City, Mexico	24	100	100	1	55
Netherlands	2	80	5	4	5
New York, NY	100	38	41	1	42
Oahu, HI	5	96	12	1	50
Porto, Portugal	25	25	51	100	75
San Joaquin Valley, CA	2	2	83	10	4
Santa Barbara, CA	40	41	100	1	23
Santa Monica Mts, CA	31	100	100	1	33
Seattle, WA	87	60	45	27	54
Sioux Falls, SD	1	1	12	34	29
Tampa/S. Florida	90	95	45	50	50
Tijuana, Mexico	3	8	70	42	22
Washington, DC	52	45	26	4	19
Yaounde, Cameroon	10	12	25	42	20

to urbanization and the northern portion bearing 50% resistance. No portion of the exclusion layer was 100% excluded from urbanization. The center of the slope layer has a slope of 0% while the far diagonal corners have the maximum slope of 100%. The slope increments radially, appearing as a spherical depression centered at the initial urban area with a slope of 0 %. There is a threshold above which urbanization cannot occur due to a high degree of slope, resulting in an effective circular constraint on growth (the edges of the Petri dish). A threshold was included to represent the physical barrier of increased slope, above which urbanization is rarely permitted, or even possible. The slope threshold was defined as 23% in this study. The transportation layer consists of a single road running north-south through the center of the image. The hill-shade layer, which consists of a simple white background, adds only to the visual output of SLEUTH.

For the second input, a collection of parameters from the approximately one-hundred papers, presentations, theses, and dissertations about the more than 80 domestic and international SLEUTH applications was performed. Recently, studies completed by Gazulis et al. [18] and Clarke et al. [2] have sought to compile, catalog, and analyze this wealth of information, resulting in the creation of the SLEUTH online data repository, http://www.ncgia.ucsb.edu/projects/gig/.

4 Results

Initial modeling of each parameter set with the hypothetical anisotropic plane resulted in more than 2000 SLEUTH output images for the twenty unique parameter sets included in this study. Each parameter set produced time-series images in one year increments, displaying the probability of urbanization over twenty-five Monte Carlo iterations. The result was 100 probabilistic images of urban growth for each parameter set outside of the specific data driven environment that usually underlies disparate SLEUTH applications.

The results indicate that the anisotropic plane defined for this study was capable of producing conditions to which each parameter set must adapt in order to grow. Each unique parameter set adapted differently, producing a distinct urbanization pattern both spatially and temporally. However, despite the heterogeneous results produced by each parameter set, some spatial and temporal similarities did arise among particular applications.

Based on an analysis of the growth rates of individual regions, and coupled with the spatial distribution of urban pixels in each image, sufficient similarities were observed among the resulting forecasts to warrant the creation of an urbanization behavioral taxonomy. The latter was tabulated by counting growth pixels by quadrant in the four principal directions. The initial conditions of the anisotropic plane were identical for the east and west quadrants, and as a result, the relative population of these two quadrants was averaged, giving three growth dimensions. Growth rates were calculated for each of the regions and averaged over the user-specified twenty-five Monte Carlo iterations. These growth rates were plotted and examined for similarities among the individual regions.

Plots of relative quadrant counts for each of the forecasts were created and examined for clustering. A three dimensional plot of north, south, and average east/west quadrant population revealed a distinct cluster of points with high values in the southern quadrant relative to both the north and average east/west quadrants as well as a cluster of points at or near the origin (see figure 2).

A third cluster appeared with relatively high values in the north and average east/west quadrants, as well as near complete urbanization of the southern quadrant. However, within this cluster were three parameter sets that produced results that never reached full saturation in the southern quadrant – indicating a separate cluster of points.

Each of these clusters represented a different urban growth behavior dependency, characteristic, or constraint that could be easily determined through a visual inspection of the time series images. As a result, each cluster was given a name representing the dependency, characteristic, or constraint that best described the cluster's urban behavior: these were (1) slope resistant growth; (2) transportation network dependent growth; (3) little to no growth; or (4) full build out growth.

An examination of the growth rates for each model application revealed that the clusters indicated above tended to have similar growth rates. Growth rates were calculated by dividing the number of newly urbanized pixels at each time step by the total number pixels urbanized during the simulation and then converted to a percent increase. Growth rates for slope resistant regions tapered off exponentially but

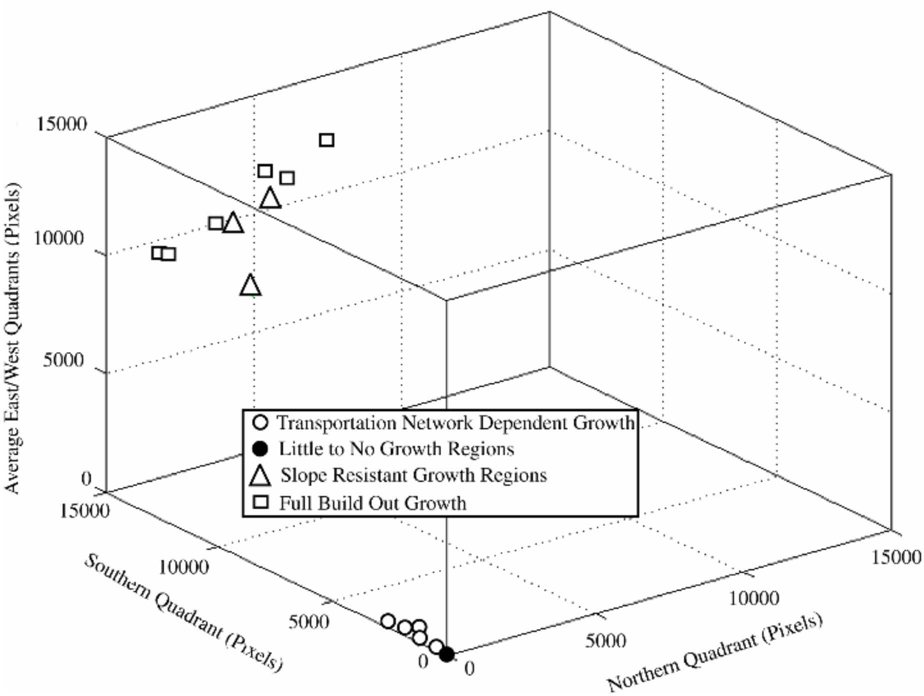

Fig. 2. Parameter set by growth direction and behavior class

generally did not drop to zero by model year 100. Growth rates for transportation network-dependent parameter sets also tapered off exponentially, but not in a smooth fashion. Year to year variability was extremely high relative to the other classes, especially in the initial growth stage between model years 0 and 30. In all fully-built-out parameter sets, growth rates dropped to zero by model year 100 and experienced little to no growth after model year 75. Finally, growth rates of little to no growth parameter sets were highly variable in all regions over the course of all 100 model years, with extreme year to year variability in the early model years. Rates for these sets dropped close to zero by model year 100 and also became less variable.

Of some interest is the little or no growth class. An interpretation of this group is that these are cities which are unable to sustain growth at all given the starting conditions for the geographical location. Of course the actual conditions differ from those we used as hypothetical examples, but nevertheless, these could be interpreted as cities that required some other impetus than normal growth to get started, perhaps planning, government incentives, or a convergence of factors such as existence of a port, railhead or other factor, including chance. Houston, for example, had the advantage of oil finds in the surrounding area adding an external impetus to growth.

The piece of information that has been missing thus far is the direct influence of time on the different categories of urbanization. The model runs allow temporal

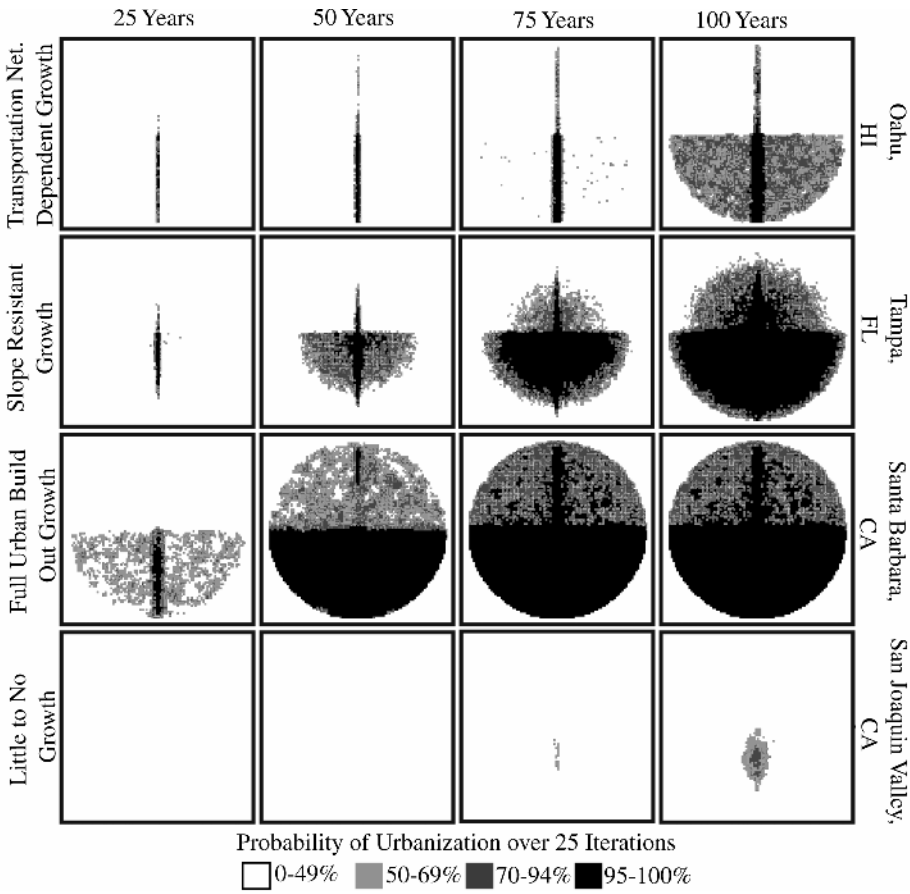

Fig. 3. Example of Archetypal Growth Patterns

comparison. For example, the sprawl category seems to accelerate early, while the slope resistant class begins sprawl-like infill only after all the available flat land is taken, about half way through the model run.

5 Conclusion

This research has shown that the output from SLEUTH calibration can provide for data-independent modeling of the urban growth of individual areas. In separating the behavior of urban growth from the city environment into a set of parameters, we gain the ability to experiment with the growth form in time and space using the SLEUTH model. An advantage of the approach is that the growth behavior is then directly comparable, and, as we have shown, is subject to classification and generalization. Exactly why these cities fall into the classes remains the topic of future research. Similarly, a more robust analysis would have hundreds or thousands of parameter sets to compare and contrast. Nevertheless, the study shows that models can be used above and beyond their traditional role (i.e. forecasting), and we have added a new role as experimental platforms for abstract behavior characterization. CA behaviors (e.g. extinction, stability, dynamic stability, growth) were expected from a CA model, nevertheless we were not aware prior to the analysis that cities could be grouped in this way.

Uninvestigated in this research were the impacts of resolution, temporal sensitivity, data sensitivity, or land use. Some of these factors have been the topic of research work on SLEUTH and other urban models, yet we feel that these issues can now be the topic of further work.

References

1. Clarke, K. C., S. Hoppen, et al. (1996). Methods And Techniques for Rigorous Calibration of a Cellular Automaton Model of Urban Growth. Third International Conference Workshop on Integrating GIS and Environmental Modeling, Santa Fe, New Mexico, National Center for Geographic Information and Analysis.
2. Clarke, K. C., N. Gazulis, et al. (2006). A Decade of SLEUTHing: Lessons Learned from Applications of a Cellular Automaton Land Use Change Model. Twenty Years of the International Journal of Geographic Information Science and Systems. T.& F. London.
3. Silva, E. and K. C. Clarke (2005) Complexity, emergence and cellular urban models: lessons learned from applying SLEUTH to two Portuguese metropolitan areas. European Planning Studies January 2005, vol. 13, no. 1, pp. 93-115 (23).
4. Agarwal, G., Grove, Evans, Schweik (2003). A review and assessment of land-use change models: Dynamics of space, time, and human choice. UFS Technical Report NE-297. Burlington, VT, U.S. Department of Agriculture Forest Service, NE Research Station.
5. Tobler, W. (1979). Cellular Geography. Philosophy in geography. S. Gale and G. Olsson. Dordrecht; Boston, D. Reidel Pub. Co.**:** 379-386.
6. Couclelis, H. (1985). "Cellular worlds: a framework for modeling micro-macro dynamics." Environment and Planning A **17**: 585-596.
7. Batty, Michael, Helen Couclelis, and M. Eichen. 1997. Special issue: urban systems as cellular automata. *Environment and Planning B* 24 (2).
8. Verburg, P. H. (2006) Modeling Land-Use and Land-Cover Change. Chapter 5 in Lambin, E. F. amd Geist, H. (eds) Land-Use and land-Cover Change: Local Processes, Global Impacts. Springer-Verlag: Berlin, Heidelberg.
9. Brail and Klosterman (2001). Planning Support Systems: Integrating Geographic Information Systems, Models, and Visualization Tools. Redlands, CA, ESRI Press.
10. Lee, J. D. B. (1973). "Requiem for Large-Scale Models." AIP Journal **May**: 163-77.
11. Lee, J. D. B. (1994). "Retrospective on Large-Scale urban Models." Journal of the American Planning Association **60**(1): 35-40.
12. Benenson, I. and P. M. Torrens (2004). Geosimulation: automata-based modelling of urban phenomena. Hoboken, NJ, John Wiley & Sons.
13. Clarke, K. C., and L. Gaydos (1998) "Loose Coupling A Cellular Automaton Model and GIS: Long-Term Growth Prediction for San Francisco and Washington/Baltimore" International Journal of Geographical Information Science, vol. 12, no. 7, pp. 699-714.
14. Goldstein, N., Dietzel, C., Clarke, K.(2005) Don't Stop 'Til You Get Enough - Sensitivity Testing Of Monte Carlo Iterations For Model Calibration. Proceedings, 8th International Conference on GeoComputation University of Michigan, Eastern Michigan University.

15. Dietzel, C. (2005). From Parameters to Scenarios: Using the Behavior of CA to Create New Features. Geography. Santa Barbara, University of California, Santa Barbara. **PhD:** 166.
16. Goldstein, N. C. (2005). Brains versus Brawn - Comparitive Stratagies for the Calibration of a Cellular Automata-Based Urban Growth Model. GeoDynamics. P. M. Atkinson. Boca Raton, CRC Press**:** 249-272.
17. Berry, B. J. L. and D. F. Marble (1968). Spatial analysis; a reader in statistical geography. Englewood Cliffs, N.J., Prentice-Hall.
18. Gazulis, N., N Goldstein, and K. C. Clarke (2005). The SLEUTH Online Data Repository. Annual Meeting of the Association of American Geographers. Denver, CO.