Don't stop 'til you get enough - Sensitivity testing of Monte Carlo Iterations for Model Calibration

Noah Charles Goldstein[1], Charles Dietzel[2], Keith Charles Clarke[3]

Department of Geography, University of California at Santa Barbara,
Santa Barbara, CA 93106-4060
Tel: +1 805-893-2705
Fax: +1 805-893-7782
[1]Email noah@geog.ucsb.edu

[2]Email dietzel@geog.ucsb.edu

[3]Email: kclarke@geog.ucsb.edu

Abstract

One of the powerful tools geospatial modeling uses is the Monte Carlo Method. However, little work has been done on measuring the optimal number of Monte Carlo iterations to be performed. In this work, we present the new utility of two metrics for deriving the number of Monte Carlo iterations needed to calibrate the CA-based SLEUTH Urban Growth Model. SLEUTH calibration is the process of choosing the best set of parameters to forecast urban growth into the future. SLEUTH calibration is performed on historical urban layers. The two metrics used are the OSM metric, which is the optimal combination of available SLEUTH metrics (like comparative size and dispersal of urban growth) and the MCAWS-derived Diversity metric that accounts for the individual model run area in summarizing the Monte Carlo results. We applied these two metrics on the calibration of three different cities; Tampa, FL, Merced, CA, and the Ellwood region of Santa Barbara, CA. We found that for SLEUTH calibration of historical data sets, one would need between 10 and 25 Monte Carlo iteration for the optimal variation in the calibration process. We discuss the far-reaching consequences of discovering that "less is more" in terms of Monte Carlo iterations in urban modeling, Geocomputation, and beyond.

1. Introduction

One of the more powerful tools in geosimulation is Monte Carlo simulation. Assuming a Gaussian behavior, a stochastic model that is run multiple times will produce a distribution of outputs that describes the randomness and variability in the model. However, in geostatistics, there are no standard heuristics to identify the appropriate – or best – number of Monte Carlo iterations a model should run. The tradeoffs are clear; too few, and the diversity in the model results is not represented. If one runs too many Monte Carlo iterations, valuable computational resources are used up. Following in the footsteps of Oreskes et al (1994), who outlined the importance of model validation and testing, this research presents a methodology for interpreting spatial Monte Carlo results that are uniform in nature, using results from the calibration of the SLEUTH urban growth model as an example. In this research, new metrics of gauging the diversity of Monte Carlo runs are introduced, as well as new metrics for testing the fit to the SLEUTH urban growth model.

The SLEUTH urban growth model is a cellular automata model of urban and land use change (NCGIA 2003). To date, it has been successfully applied to a variety of international urban regions including San Francisco, Porto and Lisbon, Portugal, and many others (Clarke et al. 1997, Silva and Clarke 2002). SLEUTH models urban change with four forms of growth: Spontaneous Growth, New Spreading Center Establishment, Edge Growth, and Road-Influenced Growth. Five coefficients are used to parameterize the four growth behaviors. In the calibration process, the model tries to replicate historical urban extent as the method for determining the best parameter set that captures the "flavor" of that city's growth.

During the calibration process, the SLEUTH user selects the number of Monte Carlo iterations that they would like to use. Traditionally this value has ranged from 3 to 100. In this research the goodness of fit during calibration was determined by the OSM metric (Optimal SLEUTH Metric) (Dietzel and Clarke, In Review). This metric was derived after exhaustive calibration of multiple datasets based on theoretical patterns of urban growth and data reduction through the use of self-organizing maps.

The use of what we now call the Monte Carlo Method for simulation modeling exploded in the late 1930's as a way of simulating a suite of possibilities of nuclear reactions (Kalos and Whitlock 1986). Using the Monte Carlo approach works best when the simulation behaves in a uni-modal manner (one mean), and multiple iterations of the simulation are available. A common method of viewing the results of a simulation with a stochastic element is to take the spatial average of all of the Monte Carlo runs, producing an average or probability map. The advantage of spatial averaging of spatial spread model runs is that they are easy to understand and visualize, as well as being a well-understood view of the possibilities of the spread process.

However, using spatial averaging of Monte Carlo iterations in spatial models is not always the easiest nor the most appropriate method to use in understanding the model or the outputs. If the simulation of spread produces multiple types of solutions, with a bi-, tri- or multi-modal distribution, spatial averaging will "wash out" the fine details of individual runs. For this research, we used the Monte Carlo Area Weighted Summation Metric (MCAWS) (Goldstein 2004), a metric that takes the individual iteration's extent into account when assessing the overall results of a group of Monte Carlo runs.

The MCAWS spatially distributed metric is an alternative to the simple averaging of multiple Monte Carlo spread model results, and is designed to be used with binary output. The MCAWS metric is defined as follows (Equation 1):

$$MCAWS = \sum_{n}\sum_{i}\sum_{j}\left(\frac{\left(\frac{\overline{Area} - Area_n}{MaxArea - MinArea}\right) + 1}{n}\right) \quad (1)$$

Where n is the number of Monte Carlo iterations, i and j are the number of rows and columns, $Area_n$ is the total aggregate area populated by the output of an individual run (either 1 or 0), $\overline{Area}$ is the average area of all the Monte Carlo iterations, *MaxArea* and *MinArea* are the respective maximum and minimum areal extents of all the Monte Carlo runs. MCAWS is calculated for a set of binary map outputs, on the entire grid. The MCAWS metric can range from 0 to 1, a 1 indicating regions that are populated in every single Monte Carlo runs, a zero for regions that do not appear in any output.

An added advantage of the MCAWS is the Diversity metric, derived from examining the number of unique values in the MCAWS map. The Diversity metric can give an aspatial measure to compare different models runs of different spatial resolutions.

2. Methods

In this research, the SLEUTH model was used to model urbanization of three cities, Merced, CA (from 1974 to 2000), the Ellwood subset of Santa Barbara, CA (from 1968 to 1986), and Tampa, FL (from 1980 to 1992). While these three urban areas have very different histories and characteristics determining the nature of their growth, there are some general similarities. Merced has developed in California's open Central Valley, yet Tampa and Santa Barbara have been constrained by the ocean. All three cities were modeled at different spatial resolutions. The models were run in calibration mode, and 3,125 parameter sets were tested in groups of 2, 3, 4, 7, 10, 25, 50, and 100 Monte Carlo iterations (Figure 1). The goodness of fit of the model for each of these runs was measured by the OSM. The diversity of the model runs was quantified by comparing the MCAWS-derived Diversity metric. The thesis of this work is that if more Monte Carlo iterations provide better metrics of fit, and capture the spatial diversity of the results, then more are needed. If more runs do not improve performance, fewer are needed.

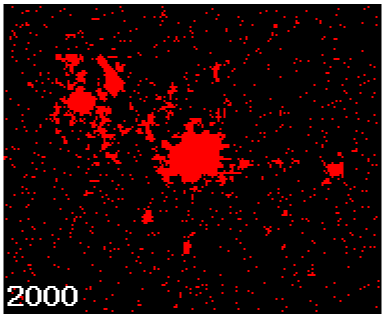

Figure 1 A sample SLEUTH simulation of Merced's urban extent in 2000

3. RESULTS

The different groups of Monte Carlo runs of each city's urban growth were evaluated using the OSM metric. As evident in Figure 2, the OSM metric behaves in an asymptotic nature, indicating that the more Monte Carlo runs are repeated, the overall set of results does not improve.

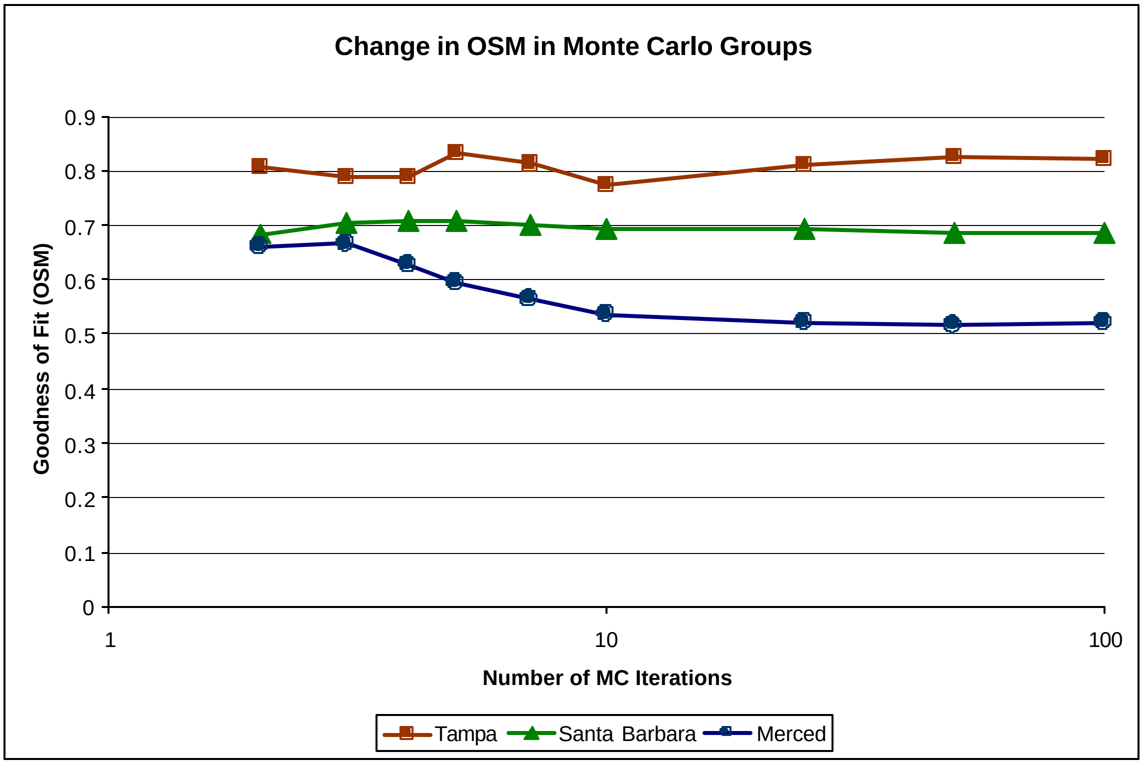

Figure 2. Goodness of fit for each calibration using 2, 3, 4, 7, 10, 25, 50, and 100 Monte Carlo iterations. Note that the X-axis is logarithmic in scale.

The graphical MCAWS output for the three study regions can be seen in Figure 3. Tampa's growth was concentrated around urban cores. Merced shows inconsistent growth from the urban cores, and a significant amount of random growth. The Ellwood subset of Santa Barbara showed concentrated development along the coast and flat regions, with slope exerting negative pressure for urbanization.

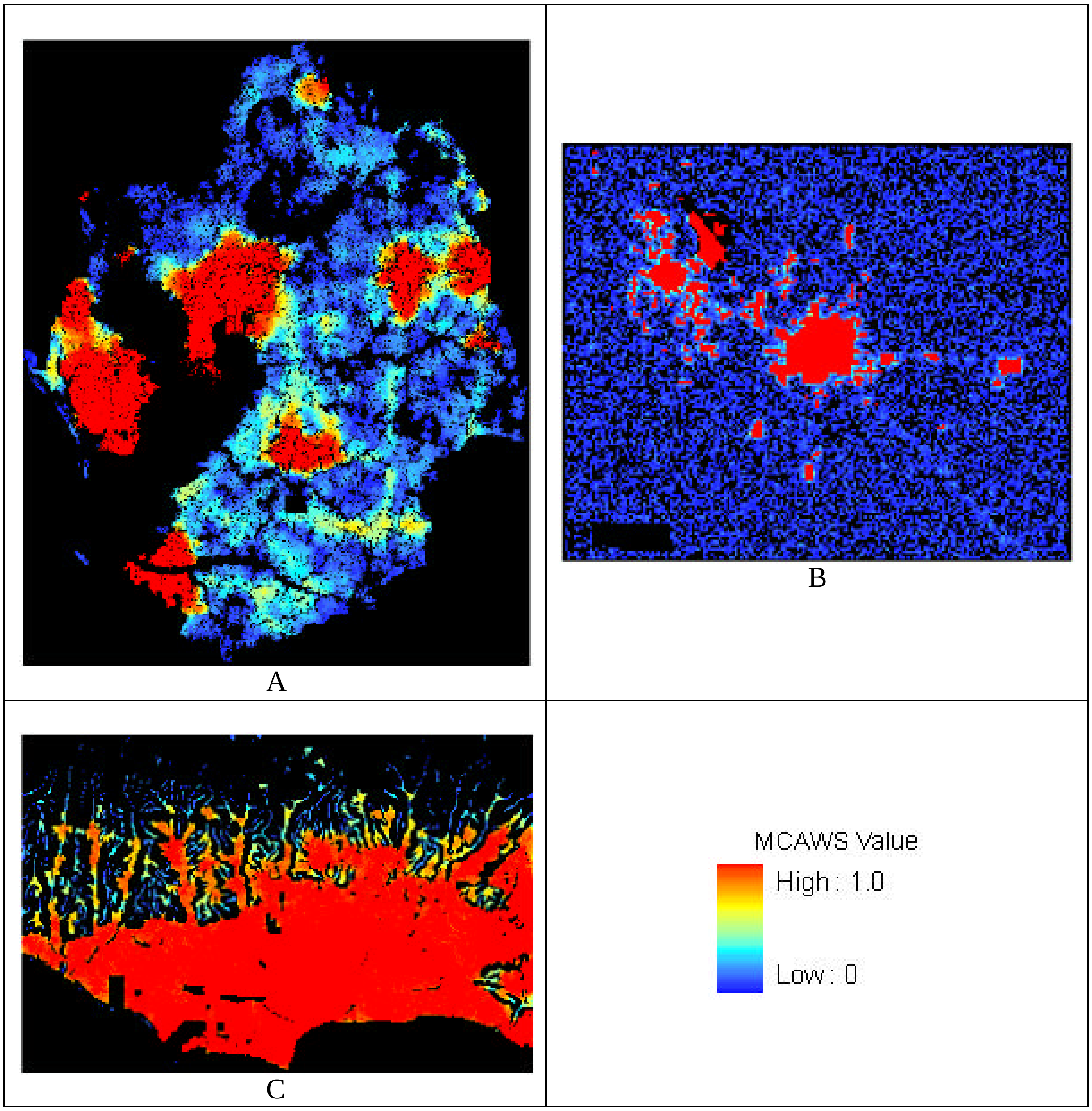

Figure 3. MCAWS maps for the three test regions. A. Tampa, FL, B. Merced, CA, C, Santa Barbara, CA (Ellwood subset). The MCAWS values are indicated by the legend. Black indicates no urbanization due to any urbanization pressure, extreme slopes, or exclusion regions (like the ocean).

As the number of Monte Carlo iterations used increases, the diversity of values increases in both the MCAWS map and the Average Map, due to spatial differences of the separate model runs. The MCAWS-derived Diversity metric and the number of unique values in the average map ("Diversity of the Average Map") are presented in Table 1.

Table 1. Diversity of MCAWS and the Average Map for all cities and Monte Carlo Groups

	Merced, CA							
Number of Monte Carlo Runs	**2**	**3**	**4**	**7**	**10**	**25**	**50**	**100**
MCAWS Diversity	3	7	15	89	172	358	348	342
Diversity of Average Map	2	3	4	7	8	14	26	43
Diversity Metric	1.5	2.33	3.75	12.71	21.5	25.57	13.38	7.95
	Santa Barbara, CA							
Number of Monte Carlo Runs	**2**	**3**	**4**	**7**	**10**	**25**	**50**	**100**
MCAWS Diversity	3	7	31	119	491	907	954	964
Diversity of Average Map	2	3	5	7	10	25	50	99
Diversity Metric	1.5	2.33	6.2	17	49.1	36.28	19.08	9.74
	Tampa, FL							
Number of Monte Carlo Runs	**2**	**3**	**4**	**7**	**10**	**25**	**50**	**100**
MCAWS Diversity	3	7	15	123	539	928	979	991
Diversity of Average Map	2	3	4	7	10	25	50	100
Diversity Metric	1.5	2.33	3.75	17.57	53.9	37.12	19.58	9.91

A graphical rendering of the Diversity Metric is below (Figure 4). From the graph, one can see that increasing the number of Monte Carlo iterations in each group helps in capturing the spatial diversity of the SLEUTH runs. However, for all cities, after 10 Monte Carlo runs, the use of additional Monte Carlo runs gives no added benefit. This is because for this spatial model additional simulation runs are merely repeating almost the same spatial pattern. From Figure 2, one can determine that adding more Monte Carlo runs does not improve the overall fit of the model. From Figure 4, it can be said that for SLEUTH, between 10 and 25 Monte Carlo runs are ideal. More than that, the model is capturing redundant information. Fewer than that, the simulation runs miss the unique inherent stochasticity of the model.

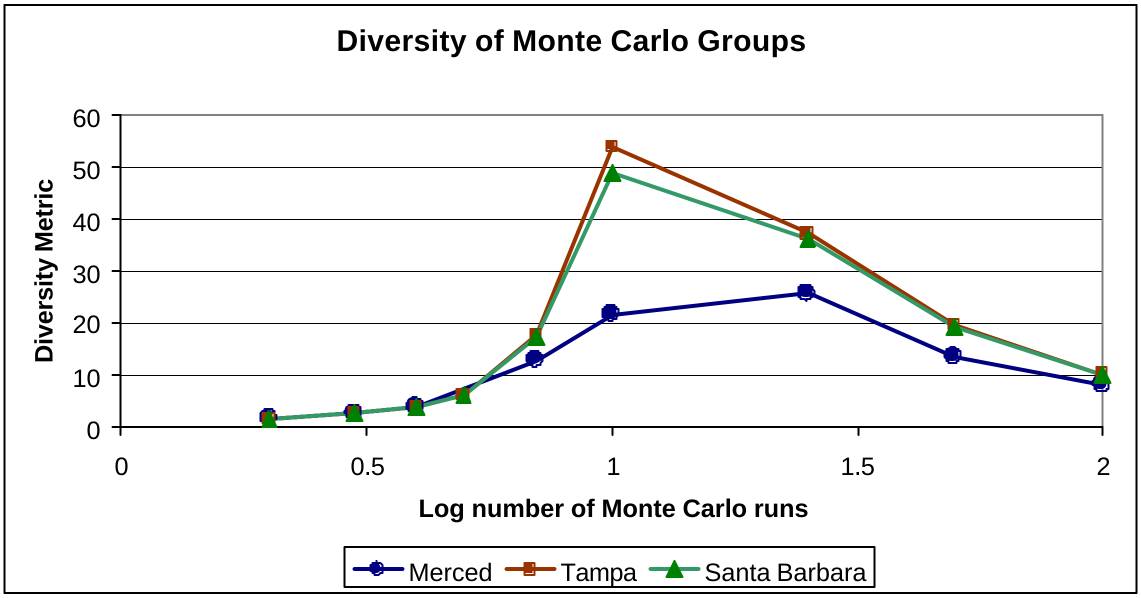

Figure 4. The Diversity Metric for each calibration using 2, 3, 4, 7, 10, 25, 50, and 100 Monte Carlo iterations. Note Log scale of number of Monte Carlo runs.

4. Conclusions

This work provides a preliminary investigation into exploring the variation in goodness of fit and spatial diversity as a function of the number of Monte Carlo iterations used in a stochastic cellular automaton. As a test case, we used one model with three data sets. For SLEUTH, we hope to gain insight into obtaining the reasonable number of Monte Carlo iterations used in an urban simulation. Since SLEUTH can run on the order of hours on contemporary computing platforms, understanding how many iterations need to run is important. This work will continue by examining scale-related issues of multiple Monte Carlo runs in SLEUTH, specifically changes in domain and spatial grain. A provocative finding of this study has been that the assumption that the more Monte Carlo iterations are performed the better is incorrect. Furthermore, MCAWS and the diversity metric provide a means to select how many iterations are best in any modeling context.

5. References

Clarke, K. C., Hoppen, S. and L. Gaydos (1997) A self-modifying cellular automaton model of historical urbanization in the San Francisco Bay area. Environment and Planning B: Planning and Design, vol. 24, pp. 247-261.

Dietzel, C., Clarke, K.C. (In Review). Determining optimal calibration metrics through the use of self-organizing maps. Transactions in GIS

Goldstein, N.C., 2004 “A Methodology for Tapping into the Spatiotemporal Diversity of Information in Simulation Models of Spatial Spread”, GIScience 2004, Adelphi, Maryland, October 2004.

Kalos, M. H. and P. A. Whitlock (1986). Monte Carlo methods. New York, J. Wiley & Sons.

NCGIA, (2003). Project Gigalopolis, NCGIA. 2003. http://www.ncgia.ucsb.edu/projects/gig/

Oreskes, N., Shrader-Frechette, K., Belitz, K., (1994) Verification, Validation, and Confirmation of Numerical Models in the Earth Sciences. Science, vol. 263, pp. 641-646.

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 ,Vol. 26, Issue 6, November 2002, Pages 525-552.