Can Statistical Methods on Land-Use Change Patterns Calibrate Agent-Based Models?

Li An[1*], Daniel Brown[1], William Rand[2], Scott Page[3]

1. School of Natural Resources and Environment
Center for the Study of Complex Systems
The University of Michigan

2. Department of Electrical Engineering and Computer Science
Center for the Study of Complex Systems
The University of Michigan

3. Department of Electrical Political Science
Center for the Study of Complex Systems
The University of Michigan

*Correspondence author:
School of Natural Resources and Environment
2004 Dana Natural Resources Building
The University of Michigan
Ann Arbor, MI 48109-1115
email: lian@umich.edu

BIOGRAPHY

Li An is a research fellow in the School of Natural Resources and Environment, the University of Michigan. His research interests include modeling Complexity in Human-environment interactions, exploring the methodology of quantitative landscape modeling, integrating social sciences with GIS/remote sensing and spatial analysis, addressing population, development, and environmental issues in developing countries (e.g., China), and researching qualitative methods in human geography. He has publications in prestigious journals such as *Annals of the Association of American Geographers* and *Science*.

INTRODUCTION

Human-environment systems are characterized by heterogeneity, nonlinear relationships, and hierarchical structures that give rise to difficulties in understanding system behavior in response to exogenous factors. Furthermore, validating models that focus on such systems is a challenge. To characterize such complex systems, *bottom-up models* (such as agent-based models) start from a general understanding of the low-level processes and elements, and generate aggregate system behavior by simulating the individual entities in the system. On the other hand, *top-down models* (such as regression) represent system relationships between aggregate variables in the system (e.g., inductively fitting empirical data with regression models). Because top-down models are useful for revealing relationships in data, it is tempting to use them to calibrate and validate agent-level rules and behaviors in bottom-up models, which can then be used for evaluating the implications of relationships at the system level. While this complementarity of approaches should be investigated, important questions remain unanswered: (1) by what metrics should top-down and bottom-up methods be compared (e.g., system level outcomes, individual agent characteristics)? and (2) in what ways and in what situations might statistical methods produce significant bias in the estimates of agent preferences and behaviors?

CURRENT DEVELOPMENTS

This research aims to answer the above questions using experiments based on an agent based model (bottom-up) and an analysis of model output using survival analysis (top-down). A spatially explicit agent-based model was developed to understand how the individual decision-making processes drive the emergent land use patterns over space and time (Brown et al. 2004; Brown et al. 2005). In previous work, we have demonstrated that survival analysis, when drawing on emergent data (i.e., data regarding what the system appears to be given many invisible or undetectable low-level mechanisms and elements), can reveal some key relationships regarding the drivers, location, and timing of land-use change, as well as their relationships (An and Brown in review). It remains to be tested whether these relationships can be used for validating agent behaviors in an agent-based model.

In the agent-based model, aesthetic quality (A) and distance to services (D) affect utility value of a parcel in an additive (multiplicative alternatively) way when a resident evaluates the parcel:

$$u_{x,y} = \frac{\sum_{j=1}^{J}[\alpha_i(\beta_i - n_i)]}{J} \qquad \text{(i for different residents)} \qquad (1)$$

Where the utility of a parcel at (x, y) is jointly determined by J (J = 2 here) landscape factors, which include aesthetic quality (A) and distance to services (D). β_j is the preferred value for one factor, and n_j is the actual factor score of the j^{th} factor, and α's are weights of these two factors that range from 1 (very important) to 0 (not important), all normalized to range from 0 to 1. Our hypothetical landscape involves all the information related to these factors.

In survival analysis, the hazard $h_i(t)$ (instantaneous risk that a parcel would be developed, which is equal to probability if multiplied by the time interval) that a parcel of undeveloped land would be developed into a specific type can be regressed against a set of exogenous variables that may take changing values over both space and time, and the coefficients (betas) can also change over time:

$$\text{Log } h_i(t) = \alpha(t) + \gamma_1(t)X_{i1}(t) + \ldots + \gamma_k(t)X_{ik}(t) \qquad \text{(i for different parcels)} \qquad (2)$$

Building upon these two models, we ran the agent-based model with the two landscape factors on a hypothetical landscape under several experimental settings: (1) *Examination of the effects of individual factors*. We varied the weights (α's in equation 1) that residents gave to the factors, from all the weight on A, through equal weight on D and A, to all the weight on D, but keep the preferences constant (i.e., set *β's* in equation 1 at 1.0); and (2) *Examination of the effects of heterogeneity in preferences*. Under the various scenarios in (1), we drew values from uniform and normal distributions for preferences (*β's* in equation 1) for each resident. We (1) ran the model 150 steps and record the spatial patterns at steps 30, 60, 90, 120, and 150; (2) output these data to ArcGIS as raster data and create 5 images that represent land-use status at five times; (3) randomly sample 10% of the total available parcels in the image of last time, i.e., step 30; (4) record its development time and type for each of the sampled parcels in step 3, along with its values for D and A; (5) conduct multinomial logistic analysis using the development type as response variable and the five landscape factors and time as explanatory variables; and (6) conduct survival analysis using the development time and type as response variables and the five landscape factors as explanatory variables.

This research is significant for several reasons. First, we can use the sign and standard deviations of the coefficients in the survival model (*γ's* in equation 2) as the metrics to evaluate if the survival analysis gives relationships that reflect the "true" mechanisms we have assumed (in relation to the α's and *β's* in equation 1). Second, the effectiveness of many top-down methods can be evaluated using bottom-up models involving some assumptive "true" mechanisms. For instance, when multiple time steps and land-use change types are involved in any top-down analysis, survival analysis gives more consistent results (compared to the "true" mechanism we have assumed) and is thus more reliable. Last, we have explored a new way to integrate bottom-up and top-down approaches by combining their strengths. In the future, we will further explore the strengths and weaknesses of these two approaches. For instance, we will use the survival analysis to evaluate the robustness of the agent-based model (e.g., some rules are only based on expert opinion/experience and subject to further justification), especially when the survival analysis is based on empirical data.

References

An, L., and D.G. Brown (in review). Exploring temporal complexity in land-use and land-cover transitions: integrating survival analysis with GIS and remote sensing. Submitted to *Landscape Ecology.*

Brown, D.G., Page, S.E., Riolo, R.L., and Rand, W. 2004. Agent based and analytical modeling to evaluate the effectiveness of greenbelts. *Environmental Modelling and Software*. 19(12): 1097-1109.

Brown, D.G., Page, S.E., Riolo, R., Zellner, M., and Rand, W. 2005. Path dependence and the validation of agent-based spatial models of land use. *International Journal of Geographical Information Science,* 19(1).