Multi-Scale Modelling of Population and Land Use with a Variable Grid CA

Roger White[1], Inge Uljee[2]

[1]Department of Geography
Memorial University of Newfoundland
St. John's NL A1B 3X9 Canada
Telephone: +1 709 737 7417
Fax: + 1 709 7373119
roger@mun.ca

[2]Research Institute for Knowledge Systems
Maastricht, The Netherlands
Telephone: +31 43 388 3322
inge@riks.nl

1. Introduction

Standard cellular automata (CA) based models capture spatial dynamics and the evolution of spatial structure insofar as these depend on the local interactions between each cell and the cells in its neighbourhood. However, it is well known that long distance interactions are important in the dynamics of cities and regions, but these long distance interactions are not captured in a CA. This problem has until now been solved by linking a standard CA model with a traditional spatial interaction based model running on larger regions like census tracts, counties, or provinces. This solution creates several problems. First, it greatly increases the number of parameters, since there are now two models instead of one, and additional parameters are required to link them. Second, since the CA is constrained regionally, artefacts appear, as when a growing cluster near the edge of one region fails to grow across the boundary into the neighbouring region.

2. The Variable Grid CA

Here we present a new approach to this modelling problem, in which processes operating at all scales are represented in a single CA model, one with a variable grid. The basic feature of the variable grid CA is that the neighbourhood is defined in terms of cells which become progressively larger toward the periphery of the neighbourhood, so that the number of cells in the neighbourhood remains small even though the neighbourhood always covers the entire modelled area (Andersson *et al*., 2002). Specifically, the cells of the basic grid or raster are defined as

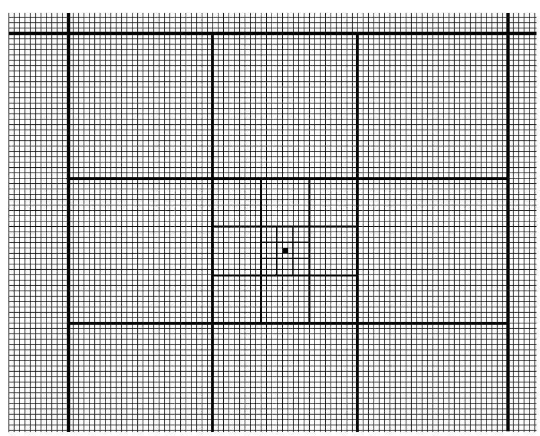

Fig. 1. A variable grid neighbourhood

level zero, or L_0. Then L_1 cells are defined as 3x3 aggregates of L_0 cells, and so on, so that each cell of level L contains 3^2 cells of level L-1, or 3^{2L} cells of L_0 (Fig. 1). As in a conventional CA, each cell is assigned a state according to its dominant land use (Engelen *et al.*). If the state (e.g. housing) is one which corresponds to a modelled activity, then the cell is also assigned an activity level (e.g. population). Cells of level $L>0$ are characterized not by states, since in general these larger cells include multiple land uses, but rather by a vector of activity levels. For each modelled land use, the activity level is the sum of the activity levels of the L_0 cells contained in the cell.

The neighbourhood effect for each activity k is calculated as the weighted sum of the activity values of all $L\geq0$ cells in the neighbourhood:

$$N_k = \Sigma_k \Sigma_j \, w_{lkj} A_{kj} \qquad (1)$$

where N_k=the neighbourhood effect for activity k; w_{lkj}=weight of the effect of activity l in $L\geq0$ cell j on the potential for activity k in the cell for which the neighbourhood effect is being calculated; and A_{kj}=the level of activity k in $L\geq0$ cell j. Since the weights are a function of distance, the neighbourhood effect captures long distance as well as local interactions (White, 2005).

The neighbourhood effect is used to calculate transition potentials as in other constrained CA models (Engelen *et al*., 2003): the neighbourhood value is weighted by measures of suitability, local accessibility to the transport network, zoning, and a measure of diseconomies of agglomeration, and the resulting value is then subjected to a random perturbation to arrive at the final value of the transition potential for the cell. For each cell, a vector of transition potentials is calculated—one for each possible transition, including a "transition" to the current state. Cells are changed to the state for which they have the highest potential, and are allocated a quantity of the corresponding activity; when all of a particular activity has been allocated, no more cells are converted to that land use. The global constraint on the CA thus operates in terms of activities rather than land use as in the standard constrained CA land use models. The amount of activity allocated to a cell at each iteration is based on the mean activity per cell, calculated over all cells with the corresponding land use and adjusted for the relative transition potential V_{ki} of the cell. Thus, since transition potentials are affected by location, the density of activity in a cell depends on the cell's location. Typically, the more accessible or more central the location of a cell, the more activity it will have. In a final step, activity levels for each L_0 cell are summed within statistical or administrative regions to give regional activity levels. This permits comparison with actual data for purposes of calibration or validation, and is also usually a requirement for end users.

3. Results

The variable grid approach is being tested in three situations. The first is the nine-county region of Dublin, Ireland, which was previously modeled with a standard spatial interaction based regional model linked to a CA based land use model with a resolution of 200m. This thus permits comparison of the results of the variable grid approach with those of the linked model approach, as well as with a constant share baseline case. The second is Portugal and its five regions, with a basic grid size of 500m and for the period 1990-2000. This application was developed as part of the DeSurvey project under the

Framework 6 Programme of the European Commission. The third application is to the Vancouver, Canada, metropolitan area, where populations of the 22 municipalities making up the region are predicted, along with land use.

In all three applications, the model performs well, giving good, high resolution predictions of land use, as well as improved population estimates for constituent regions. In the Dublin application, errors in predicted populations for all nine counties are less than 1%, while the conventional linked CA - regional interaction model gave errors of up to 14%. In the Portugal application, with results averaged over 1000 stochastic runs of the model, the largest error in the predictions of regional populations was 0.16%, while errors using a constant share approach ranged up to 3.03%; the largest error in predicted growth, a much more sensitive measure, was 11.5% compared to 455.5% for the constant share prediction (Table 1). Vancouver results are still in process, since 2006 census results are only available from March 2007.

POPULATION			Mean of 1000 Runs			Constant Share	
region	*pop. 1990*	*pop. 2000*	*sim. 2000*	*% error*	*% error growth*	*% error*	*% error growth*
pt11	3519	3619.5	3618.706	-0.02194	-0.79004	-0.48829	-17.5857
pt12	1713.7	1742.6	1742.223	-0.02165	-1.30569	0.656214	39.5681
pt13	3308.8	3390.2	3395.678	0.161594	6.730179	-0.10386	-4.32552
pt14	528.7	525.2	524.7958	-0.07696	11.54885	3.035786	-455.544
pt15	344.8	359.1	358.7149	-0.10723	-2.69278	-1.72221	-43.2481
Root Mean Error:				2.493868		12.23089	

Table 1. Portugal: predicted regional populations, year 2000, percent errors, and percent errors in predicted growth for both the variable grid model and the constant share baseline prediction. Region pt11 includes Porto and the north, pt13 is the Lisbon area, and pt15 is the Algarve.

4. Conclusions

The variable grid approach to modelling land use dynamics together with corresponding levels of population and economic activity within a single CA framework gives significantly better regional population estimates than either constant share estimates or, in an application to the Dublin area, which had already been modelled by linked CA and regional models, the estimates generated by that model. The superior results are most likely due to the fact that the model bases activity allocations not only on spatial interaction principles such as distance decay of influences, but also on micro-scale geographical phenomena like neighbourhood land use, accessibility, zoning, and suitabilities, including physical characteristics of the land.

The fact that both land use and activity levels are modelled simultaneously with a single set of parameters makes good predictions more difficult ceteris paribus, since there are many fewer parameters to work with to produce the same output compared, with the conventional linked regional and cellular models. Consequently, the clearly superior performance of the variable grid approach in the Dublin application, where a direct comparison is possible, is gratifying.

5. Acknowledgements

This work was supported by the Research Institute for Knowledge Systems, Maastricht, The Netherlands, the Framework 6 Programme of the European Commission, and the Social Sciences and Humanities Research Council of Canada.

6. References

Andersson, C., Lindgren, K. Rasmussen, S. and White, R. "Urban Growth Simulation from First Principles". *Physical Review E* vol. 66 (2002) pp. 026024: 1-9.

Engelen, G., White, R. and De Nijs, T. "Environment Explorer: Spatial Support System for the Integrated Assessment of Socio-Economic and Environmental Policies in the Netherlands". *Integrated Assessment* vol. 4 (2003), pp. 97-105.

White, R. "Modelling Multi-scale Processes in a Cellular Automata Framework". In J. Portugali, ed.: *Complex Artificial Environments*, Springer-Verlag (2005), pp. 165-178.