reviewed paper

Of Cells and Cities: a Comparative Econometric and Cellular Automata Approach to Urban Growth Modelling

Tamas Krisztin, Eric de Noronha Vaz, Matthias Koch

(Department of Computer Science, University of Algarve, Vienna University of Economics and Business)

1 ABSTRACT

This paper presents a comparative assessment of two distinct urban growth modeling approaches. The first urban model uses a traditional Cellular Automata methodology, based on Markov transition chains to prospect probabilities of future urban change. Drawing forth from non-linear cell dynamics, a multi-criteria evaluation of known variables prospects the weights of variables related to urban planning (road networks, slope and proximity to urban areas). The latter model, frames a novel approach to urban growth modeling using a linear Logit model (LLM) which can account for region specific variables and path dependency of urban growth. Hence, the drivers and constraints for both models are used similarly and the same study area is assessed. The comparative approach of both these model introduces the region of the Algarve, the farthest south region of Continental Portugal, which has been since the sixties victim to excessive urban sprawl, bringing specific vulnerability to the coastal region. Within the European context, the CORINE Land Cover project is used for assessment in the multi-temporal spatial landcover layers for 1990, 2000 and 2006. Both models are projected in the segment of Faro-Olh~ao for 2006 and a comparative assessment to ground truth is held. The calculation of Cohen's Kappa for both projections in 2006 allows for an assessment of both models. This instrumental approach illuminates the differences between the traditional model and the new type of urban growth model which is used. Both models behave quite differently: While the Markov Cellular Automata model brings an over classification of urban growth, the LLM responds in the underestimation of urban sprawl for the same period. Both excelled with a Kappa calculation of over 89%, and showed to have fairly good estimations for the study area. One may conclude that the Markov CA Model permits a riper understanding of urban growth, but fails to analyze urban sprawl. The LLM model shares interesting results within the possibility of identifying urban sprawl patterns, and is therefore an interesting solution for some locations. Another advantage of the LLM is directly linked to the possibility of establishing probability for urban growth. Thus, while the traditional methodology shared better results, LLM can be also an interesting estimate for urban patterns from an econometric perspective. Hence further research is needed in exploring the utility of spatial econometric approaches to urban growth.

2 INTRODUCTION

The objective of this paper is to compare the application of two distinct urban land modelling approaches. Using high resolution land use inventories from the CORINE Land Cover inventory, part of the region of the Algarve coast will be assessed (Figure 1). The region of the Algarve has been since the sixties a region of immense pressure regarding urban growth [11] and is therefore a region of high interest in understanding urban dynamics. The first urban growth model represents a linear Logit model (LLM), which is basically an econometric approach to urban growth that is capable of taking spatial interaction into account. The second modelling approach takes advantage of a cellular automata (CA) urban land-cover change model [3]. CA models are based on the dynamics of cellular automata adapted to the spatial transition of urban areas. Predominantly, this paper focuses on the comparison of the differences in both modelling approaches, answering to the issue of better and more stable prediction of urban land use and which criteria may be more relevantly assessed and tested. While the utility of CA models are largely documented ([13], [2]), they have shown great capacities to model regional and local urban changes [1] as environmental and socioeconomical variables may be integrated to show urban change [12]. On the other hand there have only been a few attempts to use econometric modelling approaches to predict urban growth ([7], Chapter 10). Whereas these approaches model the spatial aspects in the probabilities, the LLM approach models the spatial realizations in the variables. Hence the paper motivates in section 3.2. the use of an econometric approach. It is represented by one Logit model incorporating the explanatory variables in a linear fashion. The quality of the models' forecast is measured via Cohen's Kappa. The development of land-use maps with higher accuracy and multi-temporal layers has supported the development of understanding the patterns of urban change. In Europe, several initiatives have been carried, showing the concern of rapid urban sprawl witnessed in the last decades. CORINE Land Cover (CLC) has been a major initiative to report the land-use inventory for entire

Europe, integrating the temporal dynamics since the first project, carried out in 1986. Although the CLC is a major contribution to understand the changes in land use in Europe, most of the data comprehend a low spatial resolution to accurately monitor urban change in smaller cities and regions. While the etymology of the word growth presupposes an organized process of construction, the word sprawl converges into a random and unorganized pattern on evolution. It is as a consequence of this clear difference in word- ing that one might define the axiomatic difference between urban growth and urban sprawl. While the latter entails an organized process leading to development of cities, and considered nowadays manageable, the second fails ecosystems and questions the vulnerability of land use. Whichever the case, sprawl or growth, it results from a multi-faceted phenomenon, adjusted by geographical, economical and demographical factors. While socioeconomi- cal factors are generally hard to manage, the supporting infrastructures for city growth are manageable, and depend on the decision making of planners and stakeholders alike. Concerning urban sprawl, lessons of the past are im- portant for future planning, as development of cities largely depends on the creation (and legal constraining) of infrastructures. Satellite imagery and urban growth models are as such, important tools to envision future out- comes within continuing trends, and help decision makers at regional level to integrate better actions [9].

3 STUDY AREA

The Concelhos of Faro and Loule in the region of the Algarve have played an increasingly important role in the development of the Algarve at regional level. The district's capital of the Algarve is Faro, and although with only 41.307 inhabitants, it yields an international Airport, offing the necessary infrastructures for Tourism. At national level, the Algarve is the most Touristic region of Portugal with 41.4% of tourism industry in the region. One of the major consequences of economic growth in the Algarve brought by Tourism industry has been felt in the population dynamics throughout the region. The figure below, shows the evolution of the population in the Algarve since 1990 to 2006, based on the CORINE Land Cover data. While population density has been steadily increasing until the sixties, since the sixties, a rapid and unprecedented population growth has been felt, contributed by the mass tourism industry in the Algarve (Figure 1).

4 METHODS

4.1 Notation

Let $y_{i;t} \in Y_t$ the typical element out of Y_t. $y_{i;t}$ can either be zero or one (urban and non-urban). Hence $y_{i;t}$ follows a Bernoulli distribution. The probability for the realization of $y_{i;t} = 1$ is denoted with $p_{i;t}$. If Z_t is a n by k_z matrix then $z_{i;t}$ is i-th row of Z_t.The region specific variables for time t are the columns of the matrix Xt. All metric or "countable" variables of X_t like slope, the mean of the neighbouring slopes are labeled X^m_t and all categorial variables as X^d_t . Hence X_t can be written as $[X^d_t ; X^m_t]$. Note that X_t has to have full rank. The rank of a matrix X is given by rank(X).

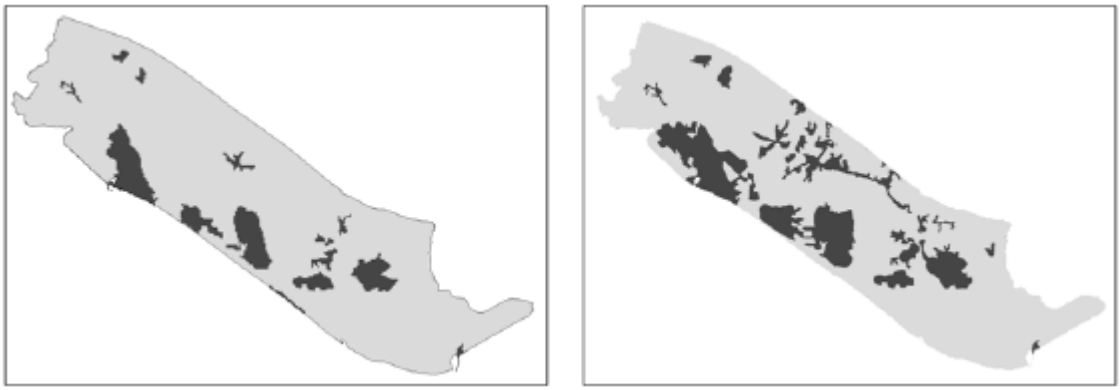

Fig. 1: CORINE Land Cover from the Algarve from 1990 and 2006.

4.2 A Motivation for econometric Modelling in urban growth

Econometric models treat the $y_{i;t}$ as the realizations of a Bernoulli distributed variable. This is similar to tossing an unfair coin where the realization head represents the property of the cell being treated as urban

and tail as non-urban. The aim of the econometric modelling is to predict the probabilities for the Bernoulli distributed variable $y_{i;t}$. One of the simplest econometric modelling approaches is to treat the process underlying the probabilities as linear, like equation (1) where ß$_1$ and ß$_2$ are merely some constants and the error term $\varepsilon_{i,t+1}$ is assumed to independently and identically distributed with zero mean and finite variance. Using ordinary least squares for equation (1) produces estimators for ß$_1$ and ß$_2$. To predict the probabilities of model (1) the estimators $\widehat{\beta_1}$ and $\widehat{\beta_2}$ are used instead of ß$_1$ and ß$_2$ in equation (1), thus resulting in equation (2).

$$y_{i,t+1} = \beta_1 + y_{i,t}\beta_2 + \varepsilon_{i,t+1} \quad (1)$$

$$\widehat{y}_{i,t+1} = \widehat{\beta}_1 + y_{i,t}\widehat{\beta}_2 \quad (2)$$

$\widehat{y}_{i,t+1}$ is the estimation for the probability $p_{i;t}$. Given the probability $p_{i;t}$ one can construct a Transition Matrix M for each land cell (this is the same for every $y_{i;t}$). The appendix shows that the linear econometric model (2) yields the following intuitive Transition matrix M for urban and non-urban cells if T = 1:

$$MT = \begin{pmatrix} \widehat{\widehat{b}}_1 + \widehat{\widehat{b}}_2 & 1 - \widehat{\widehat{b}}_1 - \widehat{\widehat{b}}_2 \\ \widehat{\widehat{b}}_2 & 1 - \widehat{\widehat{b}}_2 \end{pmatrix} = \begin{pmatrix} \frac{\overline{u}}{u_t} & 1 - \frac{\overline{u}}{u_t} \\ 1 - \frac{\overline{l_t}}{l_t} & \frac{\overline{l_t}}{l_t} \end{pmatrix} \quad (3)$$

In (3) $\overline{u}$ represents the number entries that equal one in both vectors Y_t and Y_{t+1}. Hence u is the number of entries that were urban in t and are still urban in t+1. The variable u_t is equal to the total number of urban regions in t. The variables l (the land use) follows the same notational logic as u_t. So far this subsection showed that using a linear econometric model yields an intuitive Transition matrix M. In order to account for more complex transition patterns (e.g. that the different y_i have different Transition-matrices) region specific variables like the coressponding slope, proximity to the road-network or the realizations of the neigbouring cells[1] can be incorporated into (1). These additional variables require another specification to estimate $\widehat{\mathbf{Y}}_t$, since the linear model in (1) could predict probabilities that are greater than one or smaller zero. To avoid this problem another specification, namely a logit model[2] (for details see [6] page 671 or [5] page 189) are used for modelling urban growth. Before the next section will introduce the Logit model, some additional variables will be discussed. To account for the spatial spillover the LLM will also include the variables $\mathbf{W}_{j,1\leq j\leq 5}\mathbf{Y}_{t-10}$ and $\mathbf{W}_{j,1\leq j\leq 5}\mathbf{X}_t^m$ where the W_j matrices reflect a queen neighbouring patterns of order j. The entries $w_{il;j}$ of W_j are set one if two cells are considered to be j th order neighbors and zero otherwise. The matrices containing all the spatially lagged variables are called $\mathbf{W}_{t-10,\mathbf{Y}}$ and $\mathbf{W}_{t,\mathbf{X}}$. Per definition, the diagonal entries $w_{il;j}$ are set to zero. The basic idea behind the incoporation of a time-lagged spatial lag is that a random change from a non-urban cell to an urban cell in a non-urban area will, even if everything else is set constant, change in the next time period the neighboring cells' probabilities to be urban as well. Hence this model specification allows for some path dependency of urban sprawl. Additionally, to the path dependency, this specification proves to be computational convenient. We will basically incorporate the variables if the in sample Akaike Information Criteria is reduced.

4.3 The Linear Logitmodel

One possibility to estimate the expected probabilities for Bernoulli distributed variables given region specific variables zi;t is Maximum Likelihood with a logit- specification. The Likelihood (L) for what the literature refers to as Logit model is given by ()

$$L(\widehat{y}_{1,t} = y_{1,t}, \ldots, \widehat{y}_{n,t} = y_{n,t} | \mathbf{Z}, t = 2000) = \quad (4)$$

$$\prod_{t=1990}^{1990} \prod_{i=1}^{n} F(z_{i,t}\theta)^{y_{i,t+10}} \left(1 - F(z_{i,t}\theta)\right)^{1-y_{i,t+10}} \quad (5)$$

[1] These variables are basically the same for the CA, the LLM and the NLM.

[2] Note that althoug the estimation procedure for the Logit Models is the same as in [6] or [5], the inference for the estimators is diff erent since the explenatory variables contain lagged y.

where $F(x) = \frac{e^x}{1+e^x}$ and Z is a nT by k_z matrix given by Z = $[\iota_n, \mathbf{Y}_t, \mathbf{W}_{t-10,\mathbf{Y}}, \mathbf{W}_{t,\mathbf{X}}, \mathbf{X}_t, \mathbf{X}_t^d]$. Equation (4) will be referred to as the Linear-Logit-Model. Note that we do not need any restriction for the variables $W_{t;YX}$ since this model is stationary[3] for any $\theta \in \mathbb{R}^{k_z}$. To derive consistent estimators for (4) we maximize log(L). The Appendix shows that maximizing log(L) results the same $\widehat{\theta}$ as the nonlinear least squares problem stated in (4)

$$\widehat{\theta} = \min_{\theta \in \mathbb{R}^{k_z}} \frac{1}{2} \sum_{t \in \{1990\}} \sum_{i=1}^{n} \left(y_{i,t+10} - F(z_{i,t}\theta)\right)^2 \tag{6}$$

Since Matlab provides efficient nonlinear least squares estimation routines optimizing (6) is no problem even for huge data. Our forecast is based on the metric variables slope and binary variables road 250, road 500 and road 750 and calculated via $\widehat{y}_i = F(z_{i,2000}\widehat{\theta})$. Note that $F(z_{i,t}\theta)$ is a nonlinear function and hence the paprameter es-timates can no longer be interpreted like in (1). This is one of the main reasons why (6) might result in unsreasonable forecasts. Therefore it seems staightforward to model the possible nonlinearty in F() what we will do in our future work.

4.4 The CA application for urban growth

While Cellular Automata had their origins in the fifties in an attempt to compute the relationship of computing machines and human nervous sys-tems [8], they later developed into a more complex attempt of understanding spatial interactions between agents such as Conway's `Game of Life' [4]. De- velopment of planning, and the integration of spatial explicit dynamics [?], were a further background to understand the growth of cities with Cellular Automata [12]. A new era for Cellular Automata had emerged, taking advantage of geo-information as tools for regional decision support systems [9] and o ering an insight on the discrete future dynamics of cities [1]. The discrete functions are mostly carried out through five dimension with which the cellular automata interacts [?]: (i) lattice of cells (agglomeration of indi-vidual cell neighborhood), (ii) state of cells (e.g. urban or non-urban), (iii) time interval (spatiotemporal datasets of urban measurement), (iv) transi-tion rules (defining the capacity of a cell to change from one state to another given a set of rules) and (v) neighborhood.

The Cellular Automata Model started by a comparison of the weighted decisions of CORINE Land Cover in 90 and 2000. The proxy comparison of the Urban class in 1990 registered more significant changes, influencing directly the changes to urban, agriculture and forest classes. Incorporation of the proximity factors from urban areas was used as a key variable to make the Euclidean distance assessment from 1990 to 2000. This was achieved by calculating the average distance weight from one moment to the other. Furthermore, within decision criteria of future change, road networks were incorporated as a weighting factor. Within the same criteria, besides the road network, information on the slope of the region was also added. A weighting system of Multi-criteria Evaluation (MCE) was arranged for the input variables with equal weights to assess the distribution for the known year of 2006. Finally, a Markov transition matrix (MTM) was generated based on the input results of transi-tion within the di erent classes. One of the main advantages of this UGM is related to the possibility of assessing several classes simultaneously, allowing to compare the generated results with ground truth for future moment. This allows then to calibrate the prediction and verify the accuracy of the initial estimate. For the maximum likelihood classification of the ten year image difference a proportional error of 15% was considered. The resulting estimate brings a transition of probabilities of the conditioned land classes, which al-low to assess the possibility of change in the next time frame for 2006. This becomes a Markov transitional by the inputs of the most probable registered changes. In this case, the class of urban, had a probability of maintaining itself in urban of 83.69%, while changing within the classes of agriculture, forest and wetland with the rest of the total percentage. It is important to consider, that Boolean generated maps allowed also to create constraints for future urban growth within the study area, limiting hence any wrong inter- pretation of the possibilities of future urban expansion. Finally, a Cellular Automata was generated to project the results up for 2006, arriving to the results of a possible projection for 2006.

[3] Note that $\lim_{x\to\infty} \frac{e^x}{1+e^x} = 1$ and $\lim_{x\to-\infty} \frac{e^x}{1+e^x} = 0$, hence no matter how high the value of the time lag is, the estimated probabilities are always betweeen one and zero.

5 RESULTS

5.1 Results of the CA

The results of the cellullar automata are represented in Figure 2. The projec-tion clearly shows a slight overestimation of urban areas, though the general shape and distribution of urban sprawl was sufficiently outlined.

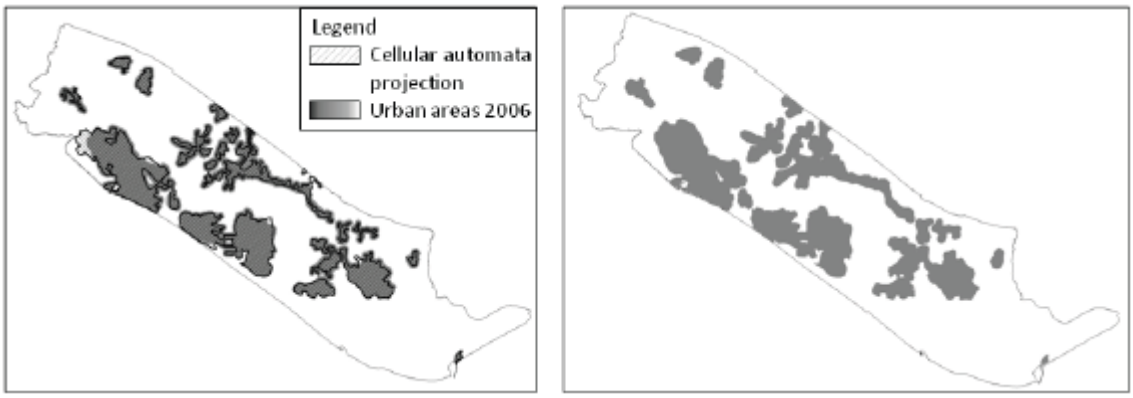

Fig. 2: Results of the Cellular Automata for the Algarve region.

5.2 Results of the linear Logit Model

Figure 3 shows the projections of the LLM model for the year 2006, overlaid over the CORINE Land Cover data for that region. The results indicate that the LLM model underestimates the number of urban regions. One advantage of the LLM model is that it can forecast areas to be urban, which were prior not close to other urban areas.

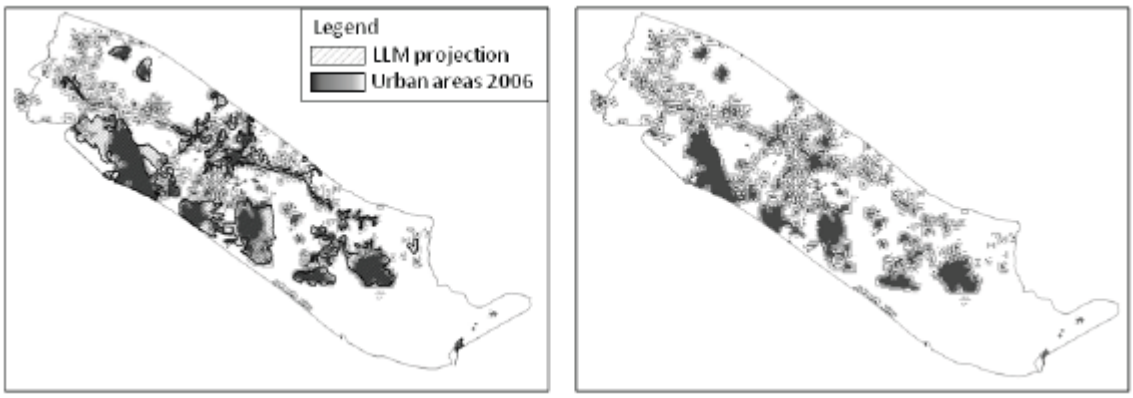

Fig. 3: Results of the LLM for the Algarve region.

6 CONCLUSION

Urban growth models have been predominantly used in a context of archi-tectural landscape planning. The relationship to urban prediction from a spatial perspective [10] has been known as a process in which spatial infor- mation and availability of data play an important role. In the context of application of the data [1] suggested an integrated methodology of approach-ing through cell based models conclusions on future urban land use. While the application urban use is not limited to a specific type of land classifier it has been largely shown that for decision making, the dimension of using spa-tial information and predictive modeling shed relevant results [9]. However, the relationship for regional planning purposes lacks a series of clear under-standing of driving forces among the knowledge of which drivers are more significant for analysis. In the specific case of the Algarve, the comparison of urban data brought from CORINE Land Cover in three distinct time series, allowed to tackle the differences in urbanization patterns over time. Two specific methodologies were used to compare and assess the results for 2006. The widely documented urban growth model follows a holistic approach of urban planning [3] and brings into account a long and traditional history for planning purposes. The econometric approach follows the traditional regional econometric modelling, outlined in [?, LeSage09] This approach is widely documented, but has been not used in this form for the

examination of urban growth. The results indicate that this approach is promising, though further work is needed to bring it to the accuracy of the more traditional Cellular Automata approach.

7 REFERENCES

[1] M. Batty. Cities and Complexity: Understanding Cities with Cellular Automata, Agent-Based Models, and Fractals. MIT Press, Cambridge, 2007.
[2] M. Batty, Y. Xie, and Z. Sun. Modeling urban dynamics through gis-based cellular automata. Computers, Environment and Urban Systems, 23(3):205-233, 1999.
[3] K. C. Clarke and S. Hoppen. A self-modifying cellular automaton model of historical urbanization in the san francisco bay area. Environment and Planning (Planning and Design), 24:247-261, 1997.
[4] H. Couclelis. Cellular worlds: a framework for modeling micro-macro dynamics. Environment and Planning A, 17(5):585-596, 1985.
[5] L. Fahrmeier, T. Kneib, and S. Lang. Regression: Modelle, Methoden und Anwendungen. Springer, Berlin/Heidelberg/New York, 2009.
[6] W. H. Greene. Econometric Analysis. Pearson Edication, Inc., Upper Saddle River, New Jersey, 2003.
[7] J. LeSage and R. Pace. An introduction to spatial econometrics. Tylor & Francis Group, Boca Raton/London/New York, 2009.
[8] V. Neumann. Theory of self-reproducing automata. UMl Reprint Uni-versity Illinois 1966 Ed, 1966.
[9] P. Nijkamp and H. Scholten. Spatial information systems: Design, mod-elling, and use in planning. International Journal of Geographical In-formation Science, 1:85-96, 1993.
[10] A. D. Syphard, K. C. Clarke, and J. Franklin. Using a cellular automa-ton model to forecast the effect of urban growth on habitat pattern in southern california. Ecological Complexity, 2:185-203, 2004.
[11] E. Vaz and P. Nijkamp. Enhancing the City: New Perspectives for Tourism and Leisure, chapter Historico-Cultural Sustainability and Ur-ban Dynamics, pages 155-177. Springer, UK, 2009.
[12] R. White, G. Engelen, and I. Uljee. The use of constrained cellular au- tomata for high-resolution modelling of urban land-use dynamics. En- vironment and Planning B, 24:232-344, 1997.
[13] Y. Xie. A generalized model for cellular urban dynamics. Geographical Analysis, 28:350-373, 1996.