Environmental Modelling & Software 75 (2016) 163–175

Contents lists available at ScienceDirect

Environmental Modelling & Software

journal homepage: www.elsevier.com/locate/envsoft

Incorporation of extended neighborhood mechanisms and its impact on urban land-use cellular automata simulations

Jiangfu Liao [a, b], Lina Tang [a, *], Guofan Shao [a, c], Xiaodan Su [a], Dingkai Chen [a], Tong Xu [a]

[a] Key Lab of Urban Environment and Health, Institute of Urban Environment, Chinese Academy of Sciences, Xiamen, 361021, China
[b] Computer Engineering College, Jimei University, Xiamen, 361021, China
[c] Department of Forestry and Natural Resources, Purdue University, West Lafayette, IN 47907, USA

ARTICLE INFO

Article history:
Received 5 January 2015
Received in revised form
15 October 2015
Accepted 17 October 2015
Available online 11 November 2015

Keywords:
Urban expansion
Cellular automata
Extended enrichment factors
Neighborhood effects
Neighborhood rules

ABSTRACT

Urban cellular automata (CA) models are broadly used in quantitative analyses and predictions of urban land-use dynamics. However, most urban CA developed with neighborhood rules consider only a small neighborhood scope under a specific spatial resolution. Here, we quantify neighborhood effects in a relatively large cellular space and analyze their role in the performance of an urban land use model. The extracted neighborhood rules were integrated into a commonly used logistic regression urban CA model (Logistic-CA), resulting in a large neighborhood urban land use model (Logistic-LNCA). Land-use simulations with both models were evaluated with urban expansion data in Xiamen City, China. Simulations with the Logistic-LNCA model raised the accuracies of built-up land by 3.0%–3.9% in two simulation periods compared with the Logistic-CA model with a 3 × 3 kernel. Parameter sensitivity analysis indicated that there was an optimal large window size in cellular space and a corresponding optimal parameter configuration.

1. Introduction

Land-use dynamics constitute an open and complex spatio-temporal evolution process that involves multi-element composite effects from natural, social, and economic factors (Arsanjani et al., 2013; Fuglsang et al., 2013; Hewitt et al., 2014). Environmental modeling can support scientific decision-making processes, and thus contribute to sustainable development associated with land-use changes. Spatial simulations and quantitative analyses of urban land-use dynamics are effective ways to improve the understanding of the evolution of urban landscapes. Cellular automata (CA) have drawn increasingly more attention in the field of land-use and land-cover analysis and simulation. The 'bottom–up' approach of CA fully reflects the concept that complex global patterns emerge from interactions governed by local rules. In addition, CA are ideal for simulating and predicting complex geographic phenomena (Liu et al., 2008a).

Based on the pioneering work by Tobler (1979) and Couclelis (1988), many researchers have developed urban land-use CA models over the last three decades, resulting in significant achievements (Batty and Xie, 1994; Clarke et al., 1997; Li and Yeh, 2000; Liu et al., 2007; Stevens et al., 2007; Takeyama and Couclelis, 1997; Verburg et al., 2004b; Wu, 2002). These models generally included a combination of drivers and spatiotemporal interactions among land uses in neighborhoods.

Identifying transition rules is a key issue in urban CA. Typically, a variety of biophysical and socioeconomic factors are included in transition rules as driving forces of urban development. Researchers have proposed various methods to determine the contributions of different spatial variables and to calibrate urban CA models (Al-Ahmadi et al., 2009; Dai et al., 2005; Feng and Liu, 2013; Feng et al., 2011; Kocabas and Dragicevic, 2007; Li and Yeh, 2002, 2004; Liao et al., 2014; Liu et al., 2008a; Verstegen et al., 2014; Wang et al., 2013; Wu, 2002; Wu and Webster, 1998; Yang et al., 2008). The binary logistic regression method developed by Wu (2002) has been widely used in urban land-use modeling because of its strict theoretical basis of statistical learning and empirical characteristics, and it has become a classic calibration method for urban CA (Cheng and Masser, 2003; Dendoncker et al., 2007; Hu and Lo, 2007; Verburg et al., 2004a). More recently, new socio-economic factors such as per-capita gross domestic product (GDP), land price, employment potential, and population density have been incorporated into the driving forces of urban CA models and integrated with logistic regression and Markov chain analysis to

* Corresponding author.
E-mail address: lntang@iue.ac.cn (L. Tang).

http://dx.doi.org/10.1016/j.envsoft.2015.10.014

predict future scenarios of urban development (Arsanjani et al., 2013; Guan et al., 2011; Mas et al., 2014). However, these models mainly considered a 3 × 3 kernel, which is a relatively small neighborhood, though studies have noted that the logistic regression urban CA model is sensitive to scale (Pan et al., 2010).

Neighborhood interaction rules are an important subset of transition rules and play a key role in the calculation of cellular conversion probabilities. To quantify and analyze the neighborhood effects generated by surrounding cells at different distances from a central cell, Verburg et al. (2004b) defined an enrichment factor formula for measuring the over- and under-representation of specific land uses in cellular space. More recently, other studies have achieved better simulation results by applying this enrichment factor to determine the neighborhood parameters of urban cellular models or as empirical data for calibrating neighborhood interaction rules (Hansen, 2008; Van Vliet et al., 2013). However, these models generally considered a small neighborhood scope with a relatively short distance from the central cell under a specific resolution during the application. For example, Van Vliet et al. (2013) used a neighborhood radius covering 0–4 unit distances (the discrete ring of a cell with a width of 500 m) to simulate urban land-use dynamics at a country scale in Germany.

The external effects generated by concentrative and dispersive forces play an important role in urban dynamics and are seen as the organizing forces of urban patterns (Harrop, 1973; Krugman, 1999; Rodrigue, 2004). Hagoort et al. (2008) pointed out that neighborhood interaction rules specify how the combined effects of spatial externalities work over distance in cellular space. Spatial externalities are considered to represent the aggregated effects of a specific land-use type on another in the neighborhood (Hagoort et al., 2008; Hansen, 2008). Research on neighborhood effects has shown that a neighborhood scope greater than a relatively small window size (i.e., a large neighborhood window) still has a significant influence on the development of the center cell (Hagoort et al., 2008; White and Engelen, 2003). In fundamental urban CA, the decay coefficient of a small neighborhood function will eventually approach zero as the radius of the neighborhood increases (Van Vliet et al., 2013). Thus a small neighborhood function cannot effectively express the impact of spatial externalities existing in a relatively large neighborhood window on the development of the central cell.

In summary, neighborhood interactions in urban CA models have mainly been limited to a 3 × 3 kernel or relatively small moving window, partially due to the aim of simplifying the models (White and Engelen, 2000). The neighborhood rules established in this case are unsuitable for detecting complex neighborhood effects over a larger scope. This problem is not prominent when the spatial resolution of geospatial data is low. However, high-spatial resolution remote sensing data have become readily available and increasingly popular. Thus, interaction rules designed for complex neighborhood effects in urban CA models are encountering unprecedented challenges. The goal of this paper is to characterize the role of complex neighborhood effects over a relatively large scope associated with urban sprawl simulation and prediction. A modeling exercise was designed to answer the following questions: 1) do large neighborhood effects exist on urban sprawl processes? 2) if yes, how can large neighborhood rules in urban CA modeling be calibrated? and 3) what is the expected increase in locational accuracy of the urban CA when large neighborhoods are incorporated?

This study addresses extended neighborhood effects on urban dynamics by using an extended neighborhood structure that is composed with cells with various influence weights based on their distances from the central cell. We used the extended neighborhood structure and calibrated parameter values to establish a large-window neighborhood function. Based on this, we developed an extended neighborhood model of urban land-use change, Logistic-LNCA, and applied it to simulate land-use changes in Xiamen City of China from 1990 to 2000. We then validated this method by using independent data acquired between 2000 and 2010.

The methodology for this study is given in the next section, together with a concise flowchart of the Logistic-LNCA model. Simulation experiments and result evaluations are presented in section three. Results are discussed in section four, and conclusions and further research directions are provided in section five.

2. Modeling methods

2.1. Model calibration based on logistic regression

Urban models simulate urban morphology evolution under various scenarios by characterizing a series of development profiles, which include physical attributes, socioeconomic status, planning and zoning constraints, and the effects of complex neighborhood interactions. Spatiotemporal models based on CA can reveal the agglomeration effects of land use at a local scale or the level of development through the iterative calculation of local and simple rules. Thus, the two interrelated processes of urban land development—spontaneous growth and self-organized growth—can be reproduced in urban cellular lattices (Wu, 2002). However, calibrating the contributions of the various aforementioned attributes to land development is a critical step to achieve more realistic and reliable urban CA simulations. Logistic regression or the multinomial logit model can be used to estimate the relationship between urban land-use changes and corresponding locational features (Bishop, 2006; McCullagh and Nelder, 1989; McMillen, 1989). More specifically, logistic regression can be seen as a process to extract the coefficients of the empirical relationships between observed land-use changes and driving forces in the integration with urban CA simulation (Wu, 2002).

Sample size and sampling strategy are two basic issues that affect the results of logistic regressions (Hirzel and Guisan, 2002; Huang et al., 2009; Munroe et al., 2004; Xie et al., 2005). Because sample size and resultant errors have an inverse relationship, a large sample size can better represent the characteristics of the study area but requires greater computing resources.

The sampling methods used in logistic regression models generally include systematic and random sampling. Systematic sampling can reduce spatial autocorrelation but may lose detailed information on some relatively isolated cells. Random sampling may better represent the population, but cannot effectively reduce spatial autocorrelation, especially local spatial dependence (Xie et al., 2005). A reasonable sampling scheme should maintain a balance between spatial autocorrelation and effective population representation (Huang et al., 2009). Considering the multiple characteristics of urban land-use modeling, we integrated systematic and random sampling, namely, proportional random-stratified sampling and extracted adequate samples to eliminate the spatial dependence of the population (Xie et al., 2005).

Generally, urban land-use change models quantify the local transition suitability of each cell from a set of demographic, econometric, and physical factors (Arsanjani et al., 2013; Fuglsang et al., 2013; Lauf et al., 2012). Cells with higher suitability are given higher probabilities in transition rules. In urban expansion simulations, the cellular space can be classified into two types, developed cells (built-up land) and undeveloped cells (non built-up land). The local development suitability at a location can be considered a function of various independent spatial variables, including elevation, slope, distance to the city center, distance to the town center, distance to the main road, distance to the railway,