Applied Geography 69 (2016) 1–9

Contents lists available at ScienceDirect

Applied Geography

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

Incorporating spatial regression model into cellular automata for simulating land use change

Chia-An Ku

Department of Geography, University of Cambridge, Downing Place, Cambridgeshire CB2 3EN, UK

A R T I C L E I N F O

Article history:
Received 8 July 2015
Received in revised form
16 February 2016
Accepted 16 February 2016
Available online 27 February 2016

Keywords:
Land use simulation
Spatial regression
Cellular automata (CA)
Integrated simulation framework

A B S T R A C T

Cellular automata (CA) model has become one of the most widely used spatial explicit methods for simulating land use change in recent decades. There are many CA models that apply ordinary least squares (OLS) regression model to estimate the weights of land-use driving factors. In these models, however, the dependent variable, which is normally the amount of change in land use, may be correlated with each other spatially. This would result in spatial dependency in residuals causing higher standard deviations of estimations that might accordingly influence the overall performance of the model. In order to solve the problem, this paper applies a spatial regression model to explain the amount of land use change as a function of land-use factors in the CA model. The results are further compared with that of an OLS-based CA model. The Local Indicators of Spatial Association (LISA) analysis showed that there was significant spatial autocorrelation ($P < 0.05$) in residuals of the OLS estimation, while the spatial autocorrelation in residuals had been significantly reduced when spatial regression was applied. Based on the coefficients of each land use driving factor in both models, weights of factors were determined and the transition potential maps were generated. Cramer's V indicators suggested that map of transition potentials created using spatial regression had better explanatory power. Finally, both maps were implemented into CA simulations, and the validation of each model indicated that the CA simulation based on spatial regression was more accurate according to Kappa statistics. The results show that the integration of spatial regression into the CA model improved the performance of land-use modeling for this region and thus could be further applied in related studies.

1. Introduction

The rapid growth of population in urban areas has changed urban land use significantly, which has accordingly influenced the balance between human and natural environments. To support the decision-making process of spatial planning for mitigating the negative impacts of land-use change on the environment, it is necessary for planners and researchers to further understand the causes and consequences of land-use change process. Therefore, many efforts have been made to study the causes, processes, consequences and impacts of land use and land cover change (LUCC) from the viewpoint of sustainable living (Lambin et al., 2001; Lawler et al., 2014; Meyfroidt, Lambin, Erb, & Hertel). In order to explore the changes in urban land use patterns from past to present, and even to the future, researchers have developed many tools and techniques to enhance the ability in understanding such mechanisms (Martinez, 1992; Veldkamp & Fresco, 1996; Verberg et al., 2012; Waddell, 2002). Given the wide variety of methods, selections of the most appropriate tools became undoubtedly important for achieving the research goals. One of the most powerful tools among them is the computational modeling approach (Dietzel & Clarke, 2006). The modeling techniques of LUCC have been intensively developed due to advances in computer and information technology in recent decades (Agarwal, Green, Grove , Evans, & Schweik, 2002; Dietzel & Clarke, 2006; Jones, 2005; Syphard, Clarke, & Franklin, 2005; Zhao & Chung, 2006). Among these models, cellular automata (CA) started to be widely applied due to its relatively simple assumption and its computational efficiency, when it comes to spatial modeling for land-use change (He, Okada, Zhang, Shi, & Li, 2008). Therefore, the concept of using grid-based systems in CA models, which aim at simulating land-use pattern based on local interactions between neighboring cells, has been widely used in urban land-use research and applications (Dragicevic, Marceau, & Marois, 2001; Du, Ge, Lakhan, Sun, & Cao, 2012; Hakan, Klein, & Srinivasan, 2007; Zhang, Li, & Chen, 2007;). In terms of the assumptions and theories

E-mail address: kuchiaan@gmail.com.

http://dx.doi.org/10.1016/j.apgeog.2016.02.005
0143-6228/

involved, there is a variety of different algorithms used in CA models to determine the transition probability and the logic of spatial allocation of land use (Clarke, Hoppen, & Gaydos, 1996; He et al., 2008; Kamuroko et al., 2009). The decision on which algorithm to use might also differ according to the local characteristics of study areas. As a result, CA model has many evolved versions with different algorithms to determine the transition probability and spatial allocation of land use. These include Slope, Land use map, Excluded area, Urban area, Transportation map, Hillshade area (SEUTH) model (Clarke & Gaydos, 1998; Clarke et al., 1996; Hakan et al, 2007), the CA-Markov model, neural network CA model, and system-dynamic CA model (Berger, 2001; He et al., 2008; Kamusoko, Aniya, Adi, & Manjoro, 2009). Among them Markov chain mathematical process is often considered as one of the most effective methods to estimate the transition of each kind of land use due to its convenient and efficient algorithm using transition probability matrix (Kamusoko et al., 2009; Muller & Middleton, 1994). The assumption behind Markov chain is that the future transition possibility of land uses is based on the transition of the past two periods (Guan, Gao, Watari, & Fukahori, 2008). Meanwhile, there have also been many studies that proved the effectiveness of applying Markov chain in CA models (Kamusoko et al., 2009; Li & Reynolds, 1997; Silvertown, Holtier, Johnson, & Dale, 1992).

In Markov chain based CA models, factors that drive land-use change (hereafter referred to as "driving factors") are often used as the reference for estimating future development potential (Kamusoko et al., 2009; Li & Reynolds, 1997). Quantitative methods are often considered more objective when determining the weight of each factor compared to subjective judgments based on researcher's own opinion (Jankowski & Richard, 1994). Among these quantitative methods, statistical analysis is widely used to understand the relationship between dependent and independent variables, which has also been applied for determination of weights of factors in different CA models (Li & Reynolds, 1997). For example, ordinary least squares (hereafter referred as "OLS") regression models can be used to estimate the relative importance of each driving factor, which is accordingly used as the weight in the Multi-Criteria Evaluation (MCE) process to create land-use transition potential maps. However, the dependent variable, which is normally the amount of change in land use, may be correlated with each other spatially that would consequently result in spatial dependency in residuals. As a result, higher standard deviations of estimations that might influence the overall performance of a model might be generated. In this sense, spatial regression models, in which spatial dependency could be explained by adding extra spatial terms into the model, could be a good alternative to solve this problem (Anselin, 1988). Chang, Lin, and Su (2008) applied a geographically weighted regression (GWR) model to modify an OLS model for estimating flood-damage where spatial dependency exists in residuals. The case study in Taiwan showed that the modified OLS model had better explanatory power. Wang, Fang, Ma, Wang, and Qin (2014) analyzed the spatial differences and multi-mechanism of carbon footprint in China using GWR model. They concluded that the application of the spatial model enabled them to more clearly detect the spatial differences. Although there are some studies that have stressed the importance of applying spatial models for better understanding spatial phenomenon, the potentials of integrating spatial regression models into a complex land-use simulation model have not been fully explored. Therefore, this paper tries to develop a framework for improving the performance of a CA land-use model by applying spatial regression model for better explanations of spatial dependency in land-use change process. Simulations based on development potential maps created using both OLS and spatial regression models were carried out and the results are compared by both spatial autocorrelation analysis and quantification of the explanatory power. Finally, the results from two CA models' simulations using different development potential maps are validated in order to understand the level of improvement by reducing the spatial dependency in residuals using the spatial regression model.

2. Data and methods

2.1. Study area and data

An empirical study was carried out in Tainan City, Taiwan (Latitude: 22°91–23°10′N; Longitude: 120°06–120°18′E) (Fig. 1). It is one of the historical cities in Taiwan while still has been growing in population in the last decade, because of its economic prosperity. Owing to the unique cultural heritages, the number of tourists in this city has also significantly increased in recent decades, which stimulates the growth of commercial activities especially those related to tourism. There were 7 main administrative districts during the focused period in this research (1995–2007) before the city was later united with Taiwan Country in 2012. The total area of Tainan City is approximately 183 km^2. By the end of 2013 the population of Tainan City was around 0.8×106, with a density of 4375.58 persons/km^2.

In order to gain greater insights into the actual distribution of different land uses during our focused periods, land-use maps in 1995 and 2007 derived from Land Use Investigation in Taiwan by National Land Surveying and Mapping Centre, M.O.I. were used to construct land-use patterns, due to higher resolution compared to satellite images. The land-use investigation map consists of 9 major groups and 103 sub-groups which provide highly detailed information on land-use characteristics. In this paper, as the focus is on the transition between developed and undeveloped areas, the maps were further re-classified into either developed or undeveloped land. Apart from land-use data, factors that influence land-use change should also be considered and analyzed. According to past research on land-use change in Tainan City (Tsou & Chang, 2004), 4 major factors are chosen to partially explain the processes of land-use change, including proximity to schools, the population density of each village, proximity to roads, and proximity to existing settlements. The reason why slope, which is one of the most-commonly used factors, is excluded in this research is that the whole city is located in Chianan Plain where urban development is mostly in a relatively flat area. This lowers the importance of slope

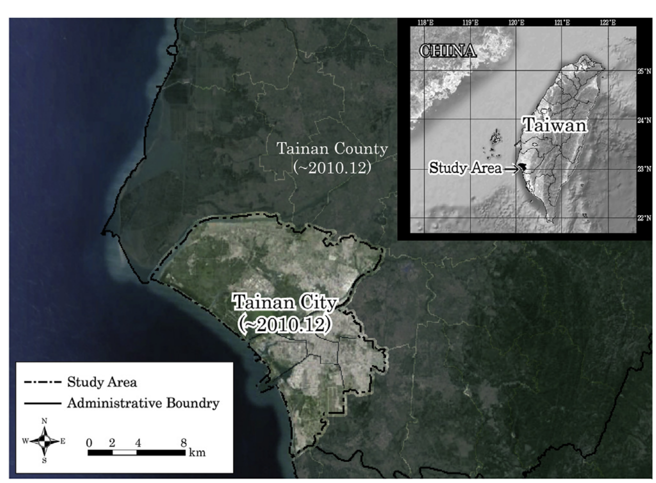

Fig. 1. Study area.
Source: Google Earth, 2013.