Computers, Environment and Urban Systems 64 (2017) 297–308

Contents lists available at ScienceDirect

Computers, Environment and Urban Systems

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

Coupling machine learning, tree-based and statistical models with cellular automata to simulate urban growth

Hossein Shafizadeh-Moghadam [a,*], Ali Asghari [b], Amin Tayyebi [c], Mohammad Taleai [d]

[a] Department of GIS and Remote Sensing, Tarbiat Modares University, Tehran, Iran
[b] RS & GIS Center, Shahid Beheshti University, Tehran, Iran
[c] Geospatial Big Data Engineer, Monsanto, MO, United States
[d] Faculty of Geodesy & Geomatics Eng., K.N.Toosi University of Technology, Tehran, Iran

ARTICLE INFO

Article history:
Received 1 September 2016
Accepted 4 April 2017
Available online xxxx

Keywords:
Machine learning models
Tree-based models
Statistical models
Cellular automata
Error map
Accuracy assessment

ABSTRACT

This paper compares six land use change (LUC) models, including artificial neural networks (ANNs), support vector regression (SVR), random forest (RF), classification and regression trees (CART), logistic regression (LR), and multivariate adaptive regression splines (MARS). These models were used to simulate urban growth in the megacity of Tehran Metropolitan Area (TMA). These LUC models were integrated with cellular automata (CA) and validated using a variety of goodness-of-fit metrics. The results showed that the percent correct metrics (PCMs) varied between 54.6% for LR and 59.6% for MARS, while the area under curve (AUC) ranged from 67.6% for LR to 74.7% for ANNs. The results also showed a considerable difference between the spatial patterns within the error maps. The results of this comparative study will enable decision makers and scholars to better understand the performance of the models when reducing the number of misses and false alarms is a priority.
© 2017 Elsevier Ltd. All rights reserved.

1. Introduction

LUC is considered to be a local process with global consequences (Lambin & Geist, 2008), which plays a crucial role in determining biodiversity levels (e.g., Reidsma et al., 2006), as well as influencing climate change (e.g., Kalnay & Cai, 2003; Tayyebi & Jenerette, 2016), water resource availability (e.g., Spera et al., 2016) and carbon cycling levels (e.g., Houghton & Goodale, 2004). One of the most important aspects of LUC is urban expansion, given that the global urbanization rate is projected to reach 72% by 2050 (United Nations, 2012). Future urbanization is alarming for at-risk natural resources and croplands (Foley et al., 2005; Tayyebi et al., 2016a), meaning that the accurate monitoring and modelling of urban expansion is necessary for infrastructure planning, land use policy and the identification of stress, particularly where such land transformation affects environmental and ecological processes (Shafizadeh-Moghadam and Helbich, 2015). The application and understanding of rigorous and accurate LUC models are thus essential if one wants to simulate the extent, intensity and spatial pattern of urban expansion.

A variety of LUC models has been developed in land change science, which can be categorized into three groups (Pontius et al., 2008). The first group of LUC models includes statistical models (e.g., LR), which have been widely used either individually (Verburg et al., 2002; Hu & Lo, 2007; Tayyebi & Jenerette, 2016) or in combination with CA (e.g., Lin et al., 2011; Azari et al., 2016). However, the complexity of urban patterns and their underlying processes does not necessarily follow the assumptions required for some models in this group; for example, normal distribution and independency among drivers of change (Tayyebi et al., 2014a). The precise relationship between dependent and predictor variables, on the other hand, is not always known in advance, and so requires a more flexible statistical approach (Goetz et al., 2015). Such an approach is provided by MARS when used to model LUC (Tayyebi & Pijanowski, 2014; Tayyebi et al., 2014a). MARS involves an intuitive, simple and easy-to-interpret non-parametric regression algorithm, which mathematically formulates the relationship between inputs and outputs (Zhang & Goh, 2013). The second group of LUC models includes machine learning (ML) models, ANNs and SVM, all of which are becoming increasing popular among scholars (Pijanowski et al., 2002, 2010 and 2014; Lin et al., 2011; Wang & Li, 2011; Kamusoko & Gamba, 2015; Rienow & Goetzke, 2015; Shafizadeh-Moghadam et al., 2015). Despite operating in a black-box manner, they are flexible and strong when handling complex relationships. For example, Pijanowski et al. (2002) developed an ANN-based model called the land transformation model (LTM) to model LUC, which has since been applied within a number of environmental applications globally (e.g., Tayyebi & Pijanowski, 2014; Shafizadeh-Moghadam et al., 2015; Shafizadeh-Moghadam et al., 2017). Similarly, SVM and least squares SVM have been successfully applied individually or integrated with CA for modelling LUC (Rienow &

* Corresponding author.
E-mail addresses: h.shafizadeh@modares.ac.ir (H. Shafizadeh-Moghadam), taleai@kntu.ac.ir (M. Taleai).

http://dx.doi.org/10.1016/j.compenvurbsys.2017.04.002
0198-9715/© 2017 Elsevier Ltd. All rights reserved.

Goetzke, 2015; Feng et al., 2015). The last group of LUC models incorporates tree-based models, CART and RF, all of which have a proven ability for modelling LUC (Tayyebi & Pijanowski, 2014; Kamusoko & Gamba, 2015). CART is easy to understand and interpret, which means they can straightforwardly handle qualitative and quantitative predictors (James et al., 2014). RF, on the other hand, represents a powerful method, which makes use of a large number of CARTs to achieve a higher performance. For example, Tayyebi and Pijanowski (2014), who employed a CART model to characterize multiple land use classes, confirmed a CART's ability to model multiple land use classes. Kamusoko and Gamba (2015) also used RF-CA to model LUC and reported an improved performance over the SVM and LR methods.

The above three LUC model groups explore the influence of socio-economic and environmental factors on LUC. All generate a transition suitability map, indicating the likelihood of a cell changing from non-built-up to a built-up class. The next step involves allocating the cells that are going to change using a spatially explicit technique (Azari et al., 2016; Shafizadeh-Moghadam et al., 2017). Thus, built on a transition suitability map, CA have been integrated with these six LUC models for the cell allocation phase within the LUC simulation process. As a result, spatial effects, local influences and neighbourhood interactions can all be explicitly captured using CA (White & Engelen, 1993; Li & Yeh, 2002). To evaluate the performance of the above-mentioned models, we used the figure of merit (FOM) (Pontius et al., 2008), overall accuracy (Pontius et al., 2008), PCMs (Pijanowski et al., 2002, 2010 and 2014) and total operating characteristic (TOC) methods here. Pontius and Si (2014b) recently introduced the TOC method to rectify the limitations of the relative operating characteristic (ROC) method, given that the ROC method fails in cases where some types of error are more important than others (Dodd and Pepe, 2003). In addition, the ROC method also fails to reveal the size of each entry in the contingency table for each threshold (Pontius and Parmentier, 2014a).

There were a number of reasons why we chose these LUC models. First, models in land change science usually fall into three categories (Tayyebi et al., 2014a) including: 1) ML models; 2) tree-based models and 3) statistical models. In this study, we chose two well-known models from each category, which have been widely used for modelling LUC in the course of the last 30 years. The criterion for model selection was based on their usage, importance and popularity in the field. For example, LR from a statistical group has been used to build the Conversion of Land Use and its Effects (CLUE) model, which is one of the most famous LUC models (Verburg et al., 2002). An ANN from an ML group has been used to build the LTM, which is another famous LUC model (Pijanowski et al., 2010 and 2014). Second, very few studies have compared LUC models that belong to these three categories (Pontius et al., 2008). Even Pontius et al. (2008) compared LUC models, which applied these models, to separate case studies, such that their conclusion showed a bias from a comparison perspective. Here, we applied all these LUC models to one study area and for the same phenomenon, which was urban growth. This study can be used as a reference for LUC modellers in order to better understand the performance of available LUC models. Finally, we used the most recent calibration metric called the TOC method, which was introduced by Pontius and Si (2014b). This method is the best available metric for calibration and comparison among LUC models. However, other existing metrics may have flaws and provide biased results.

The main objectives of this study are twofold: first, to integrate CA with six models, these being ANN and SVR (ML techniques), RF and CART (tree-based models), and LR and MARS (statistical models), to simulate urban growth within the megacity of Tehran, the capital of Iran; and, second, to explore these models in terms of their spatial accuracy and predictive ability. These models use different mechanisms to map the association between LUC and its underlying driving forces. Thus, it can be assumed that these models will result in different simulated maps. To the best of our knowledge, no work has previously evaluated these techniques in the same region and used the same influencing factors to assess and compare their performance.

2. Materials and methods

To effectively model LUC, there are two essential steps. The first step is to examine and model the influence of socio-economic and environmental factors on LUC. The outcome of this step is a transition suitability map, which indicates the likelihood of land transformation taking place for the relevant land class. The second step is to allocate the number of cells that are going to be changed. We followed the below steps to compare six LUC models (Fig. 1). Landsat images were processed and land use classes extracted. A cross-tabulation of urban growth was performed between initial and subsequent times. Then the associated driving forces in initial time were identified. LUC models used driving forces in time 1 as input, and urban growth between time 1 and time 2 as output (called calibration run). The suitability map of urban growth for each LUC model was created in time 2 during the calibration run. To convert the suitability map into a simulated map in time 2, we needed to know the quantity of urban growth from time 1 to time 2. To calculate the quantity of urban growth between time 1 and time 2, the reference land use map in time 1 was compared with the reference land use map in time 2, both of which were created from Landsat images. Next, the number of cells converted from non-built-up in time 1 to built-up in time 2 were counted. We then integrated the quantity of urban growth cells with the CA model and suitability map to create a simulated map in time 2. Finally, the simulated maps were evaluated using a set of goodness-of-fit metrics.

2.1. Study area

The megacity of Tehran (Fig. 2) is the capital, as well as the largest and most populous urban area in Iran (Census Information, 2011). Tehran is located between 35.56 and 35.83 N and 51.20–51.61E. Over the last four decades, the concentration of commercial, financial, cultural and educational activities in Tehran has attracted unprecedented migration to the city (Tayyebi et al., 2011a; Pijanowski et al., 2010). According to an official census carried out in 2011, the city has a population of 12 million people (Census Information, 2011). The urban boundary of Tehran has been expanded in all directions, but mostly focused towards the west, south-west, east and north-east, while it is restricted northward where Tehran is surrounded by mountains (Tayyebi et al., 2011b). The city currently comprises 22 district municipalities covering 700 km^2 (Madanipour, 2006).

Population pressure and settlement demand have led to the conversion of invaluable crop land into residential areas, with construction even taking place on the tops of surrounding hills (e.g., in the north-east). The city is growing and changing fast, both vertically and horizontally, and is characterized by high-rise construction, a high-density population, long-lasting traffic jams and environmental problems, such as air and noise pollution (Atash, 2007). The city is creeping out along radial road networks. It is also expanding in such a way that it is integrating adjacent settlements, resulting in the emergence of new suburban areas (Madanipour, 2006). As a result, green and open space between Tehran and its neighbouring cities has all disappeared.

2.2. Data

Retrospective information regarding the urban footprint and land use classes were extracted from long-running Landsat images, covering June 1985 (TM), July 1999 (ETM$^+$) and July 2014 (Landsat 8), with a spatial resolution of 30 × 30 m. Data between 1985 and 1999 were used for model calibration, while data between 1999 and 2014 were used for model validation (Pontius et al., 2008; Tayyebi et al., 2013).

The preparation of multi-temporal data, particularly in developing countries, is often challenging. For example, for Iran, the historical

Download English Version:

https://daneshyari.com/en/article/4965157

Download Persian Version:

https://daneshyari.com/article/4965157

Daneshyari.com