Computers, Environment and Urban Systems xxx (xxxx) xxx–xxx

Contents lists available at ScienceDirect

Computers, Environment and Urban Systems

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

Comparison of metaheuristic cellular automata models: A case study of dynamic land use simulation in the Yangtze River Delta

Yongjiu Feng[a,b,c], Yan Liu[d], Xiaohua Tong[e,*]

[a] College of Marine Sciences, Shanghai Ocean University, Shanghai 201306, China
[b] The Key Laboratory of Sustainable Exploitation of Oceanic Fisheries Resources (Ministry of Education), Shanghai Ocean University, Shanghai 201306, China
[c] National Distant-water Fisheries Engineering Research Center, Shanghai Ocean University, Shanghai 201306, China
[d] School of Earth and Environmental Sciences, The University of Queensland, Brisbane, QLD, 4072, Australia
[e] College of Surveying and Geo-Informatics, Tongji University, Shanghai 200092, China

ARTICLE INFO

Keywords:
Cellular automata
Land use modeling
Particle swarm optimization
Generalized simulated annealing
Genetic algorithm
Yangtze River Delta

ABSTRACT

Cellular automata (CA) is a bottom-up modeling framework that has increasingly been applied to simulate land use change by capturing its dynamics. Metaheuristics such as particle swarm optimization (PSO), generalized simulated annealing (GSA) and genetic algorithm (GA) have widely been incorporated into CA modeling to generate more realistic simulation patterns. We present a comparative study of four CA models incorporating logistic regression (LR) and the three metaheuristics respectively to simulate land use change in the Yangtze River Delta from 2005 to 2015. The metaheuristic processes are guided by an objective function that represents the root-mean-square error (RMSE) of the transition rules, which can then automatically search for suboptimal CA coefficients. The three metaheuristics are substantially different in terms of the algorithm mechanism, optimization iteration, and computational time. The land conversion potentials from the metaheuristics are similar in global patterns but marginally different in local regions, which substantially differ from that calculated using LR. All three metaheuristic CA models simulated slightly less than the reference change while the LR-CA model simulated substantially more than the reference change, however all models allocated the change to similar places. Our study shows that the three metaheuristics can achieve similar outcomes in the optimization of CA transition rules and land use simulation, albeit with different sensitivities to their intrinsic control parameters. We suggest that any of the three metaheuristics could be used to construct land use CA models, if the algorithm complexity and computational time are not highly concerned.

1. Introduction

Cellular automata (CA) is a modeling framework that generates complex systems output from simple interactions at an individual cell scale. CA has been recognized as one of six categories of modeling approaches that are aimed at examining land change process and predicting future land scenarios (National Research Council, 2014). Land use CA models have commonly been constructed by incorporating geographical information systems (GIS), statistical methods and intelligent algorithms (Batty, 1998; Feng & Tong, 2018; Liu et al., 2017; O'Sullivan & Torrens, 2001). CA-based land use modeling focuses on the land conversion from non-urban (such as agriculture and forests) to urban, addressing the competition between population growth and limited urban space. The models are based on transition rules driving the change of a land cell from one state to another to reflect land use change dynamics. In the last two decades, CA-based urban models have progressed substantially with regard to cell structure, neighborhood configuration, transition rules, scale effect and model evaluation (Barreira-González & Barros, 2016; Dahal & Chow, 2015; Gonzalez, Aguilera-Benavente, & Gomez-Delgado, 2015; Moreno, Wang, & Marceau, 2009). Urban CA models such as CLUE-S, SLEUTH, CA-Markov and GeoSOS have been applied in the simulation of urban expansion and multiple land use change at local and regional scales (Arsanjani, Helbich, Kainz, & Boloorani, 2013; Chaudhuri & Clarke, 2013; Li, Lao, Liu, & Chen, 2011; Verburg et al., 2002).

To understand and capture the complex relationships between land use change and its driving factors, modelers have used a variety of methods to define or optimize CA transition rules. These include statistical methods, rough and fuzzy sets, system dynamics, neural networks and metaheuristics (Guan, Shi, Huang, & Lai, 2016; Wang, Hasbani, Wang, & Marceau, 2011).

Statistical methods make definitive estimates of the weights of the

---

* Corresponding author.
E-mail address: xhtong@tongji.edu.cn (X. Tong).

https://doi.org/10.1016/j.compenvurbsys.2018.03.003
Received 8 December 2017; Received in revised form 1 March 2018; Accepted 7 March 2018

Please cite this article as: Feng, Y., Computers, Environment and Urban Systems (2018), https://doi.org/10.1016/j.compenvurbsys.2018.03.003

driving factors. Logistic regression (LR) is the most commonly applied statistical approach and has become the baseline for comparison by various other CA models (Li, Liu, & Yu, 2014; Liu & Feng, 2016; Wu, 2002). LR has limitations in capturing the relationships between land use change (response variables) and the driving factors (explanatory variables) because it requires factor independence. This requirement is frequently violated in CA-based land use modeling, especially when tens or hundreds of factors are being included (Wang et al., 2011). Recent work has demonstrated the effect of multicollinearity when more than five factors are used in modeling (Feng & Tong, 2017), indicating that more efforts are required to develop new methods to improve CA simulation.

Metaheuristics are intelligent optimization algorithms that modelers use to optimize CA transition rules. Metaheuristics' goal is to address a core issue in CA modeling, that is, how conversion rules can be calibrated to more accurately describe the actual land use transformation and generate more realistic simulation results. In most CA models, it is common to observe differences between the simulation results and the classifications from remote sensing images. Such differences can be minimized using metaheuristics by projecting CA transition rules into an algorithmic space (Cao, Huang, Li, & Li, 2014; Feng & Liu, 2013; Li, Lin, Chen, Liu, & Ai, 2013). The projection is realized by an objective function that guides the metaheuristic optimization to the final solution (Feng, 2017).

A wide range of metaheuristics have been used by modelers to construct CA models of land use change. These include particle swarm optimization (PSO), simulated annealing (SA), genetic algorithm (GA), generalized pattern search (GPS), ant colony optimization (ACO), artificial bee colony (ABC) and cuckoo search (CS) (Blecic, Cecchini, & Trunfio, 2013; Cao et al., 2014; Cao, Tang, Shen, & Wang, 2015; Feng & Liu, 2013). Among these, PSO, SA and GA are the most widely applied algorithms in the modeling of land use and urban growth. A PSO integrated CA model (PSO-CA) has been developed to simulate urban growth in Shanghai, yielding substantial improvement in overall simulation accuracy and reduction in allocation error (Feng, Liu, Tong, Liu, & Deng, 2011). Due to PSO's capabilities in automatically retrieving near globally optimal CA transition rules, PSO-CA has been applied or extended in urban modeling exercises elsewhere (Blecic, Cecchini, & Trunfio, 2014; Liao et al., 2014; Rabbani, Aghababaee, & Rajabi, 2012; Yao, Hao, & Zhang, 2016; Zhang et al., 2016; Zhang, Li, Du, & Ren, 2015). SA was used to optimize CA coefficients in modeling dynamic urban expansion, which resulted in improvement in locational simulation accuracy (Feng & Liu, 2013). This modeling method has been extended in the simulation of land use change, land use allocation and future urban patterns in rapidly urbanized areas (Mahiny, Asadolahi, Sabaee, Kamyab, & NasirAhmadi, 2014; Mohammadi, Nastaran, & Sahebgharani, 2016). A GA integrated CA model was developed in the 1990s to perform a globally-coordinated computation with local interactions (Das, Mitchell, & Crutchfield, 1994). GA has also been integrated with a CA-Markov model to simulate land use change (Jenerette & Wu, 2001) and used to search for the suboptimal combination of CA coefficients to produce more compact urban forms for planning (Li et al., 2013). Different versions of GA have also been used to improve the performance of CA models of land use change and urban growth (García, Santé, Boullón, & Crecente, 2013; Qiang & Lam, 2016; Shan, Alkheder, & Wang, 2008; Zhang, Zeng, & Bian, 2010).

The PSO, SA and GA algorithms share the ability to search for the suboptimal (near globally optimal) CA coefficients that have clear physical meanings, leading to the improvement of simulation results compared to conventional statistical methods (Feng & Liu, 2016; García et al., 2013; Li et al., 2013; Liu, Feng, & Pontius, 2014). The methods also differ algorithmically, in their local and global search capabilities, in convergence processes, in computational abilities, and in the calibration of the control parameters. They are also affected by the objective function in use, initial solution, and lower and upper bounds of the CA coefficients. Moreover, GA may have problems such as poor local search ability, premature convergence, and over-fitting. While some have reported that these methods surpass the LR-CA model, to date there has been no comparison of these algorithms by applying them to the same study area.

The integration of intelligent and metaheuristic algorithms with CA models for better outcomes is among the modeling challenges that are essential for the future research agenda (Pontius et al., 2018). This paper addresses the challenge by applying the three metaheuristics (standard PSO, generalized simulated annealing [GSA] and standard GA) to retrieve CA coefficients and optimize land transition rules. These CA models were later applied to simulate the urban land use change in China's Yangtze River Delta. By comparing the optimization capacity and outcome of these metaheuristic CA models in comparison with the baseline LR-CA model, we sought to offer suggestions for selecting an appropriate optimization algorithm in CA-based urban modeling practice.

1.1. CA models and their validation

1.1.1. A typical CA model

CA models are commonly built on land transition rules that define the state of a cell at the next time step as a function of the current state of the cell itself and its neighboring cells (Liu & Feng, 2012; Munshi, Zuidgeest, Brussel, & van Maarseveen, 2014). This approach has been widely applied to model the unidirectional conversion from non-urban to urban to address the urban growth in space and over time. A conceptual formula of the transition rule can be written as (Liu & Feng, 2016; Wu, 2002):

$$S_{i,t+1} = TranFun(S_{i,t}, P_{var}, \mathrm{CONTI}, \mathrm{CONS}, RND) \tag{1}$$

where f is a transition function that defines the state ($S_{i,\ t+1}$) of cell i at a future time $t + 1$, while the transition function consists of the state ($S_{i,\ t}$) of the land cell i at the present time t, the effects of the contiguity cells (*CONTI*), the constraints (*CONS*) that restrict the transition of cell state, and a stochastic factor (R) that simulates unknown perturbations.

An important feature of CA models is that they consider the effects of the contiguity to generate spatial patterns of complex urban systems. A commonly used neighborhood configuration is a square $m \times m$ region of cells. The conversion potential defined within a square neighborhood can be written as (Dahal & Chow, 2015; Maithani, 2010):

$$CONTI = \mathrm{NeiStat}(m \times m, Sum(Urban)/Sum(Total)) \tag{2}$$

where m represents the width of the square, and CONTI is the ratio of the number of urban cells to all neighboring cells, excluding the central cell.

Spatial constraints in CA models indicate a cell being restricted from development. Such constrained areas usually include broad water bodies, basic farmlands, ecological reserves, public parks and green spaces (Liu et al., 2017; Wu, 2002). The CONS in Eq. (1) is then assigned to 1 if a cell is available for development; otherwise, it is assigned 0.

A stochastic disturbance factor R that models uncertainties and unknown perturbations in land use change can be written as (Feng et al., 2011; White & Engelen, 1993):

$$RND = 1 + [-\ln\ RandReal()]^{\alpha} \tag{3}$$

where *RandReal* and α are controlling coefficients that adjust the effects of stochasticity on land use change. The factor *RandReal* is a pseudo-random number on [0,1], while α is an integer in the range of [0,10].

The land conversion potential (P_{var}) of cell i is affected by a set of proximity and biophysical factors, which can be written as the following equation using logistic regression (Jafari, Majedi, Monavari, Alesheikh, & Kheirkhah Zarkesh, 2016; Munshi et al., 2014; White & Engelen, 1993; Wu & Webster, 1998):