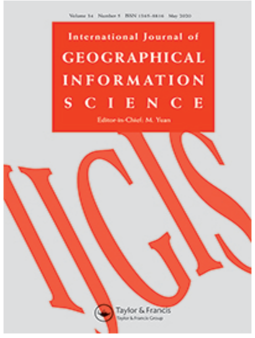

Taylor & Francis
Taylor & Francis Group

International Journal of Geographical Information Science

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/tgis20

A review of assessment methods for cellular automata models of land-use change and urban growth

Xiaohua Tong & Yongjiu Feng

INTERNATIONAL JOURNAL OF GEOGRAPHICAL INFORMATION SCIENCE
2020, VOL. 34, NO. 5, 866–898
https://doi.org/10.1080/13658816.2019.1684499

Taylor & Francis
Taylor & Francis Group

REVIEW ARTICLE

A review of assessment methods for cellular automata models of land-use change and urban growth

Xiaohua Tong[a] and Yongjiu Feng[a,b]

[a]College of Surveying and Geo-Informatics, Tongji University, Shanghai, China; [b]School of Earth and Environmental Sciences, The University of Queensland, Brisbane, Australia

ABSTRACT

Cellular automata (CA) models are in growing use for land-use change simulation and future scenario prediction. It is necessary to conduct model assessment that reports the quality of simulation results and how well the models reproduce reliable spatial patterns. Here, we review 347 CA articles published during 1999–2018 identified by a Scholar Google search using 'cellular automata', 'land' and 'urban' as keywords. Our review demonstrates that, during the past two decades, 89% of the publications include model assessment related to dataset, procedure and result using more than ten different methods. Among all methods, cell-by-cell comparison and landscape analysis were most frequently applied in the CA model assessment; specifically, overall accuracy and standard Kappa coefficient respectively rank first and second among all metrics. The end-state assessment is often criticized by modelers because it cannot adequately reflect the modeling ability of CA models. We provide five suggestions to the method selection, aiming to offer a background framework for future method choices as well as urging to focus on the assessment of input data and error propagation, procedure, quantitative and spatial change, and the impact of driving factors.

ARTICLE HISTORY

Received 31 August 2018
Accepted 21 October 2019

KEYWORDS

Cellular automata; model assessment; cross-tabulation matrix; Kappa coefficients; landscape metrics; model sensitivity

1. Introduction

Spatially explicit models (SEMs) are location-based computational approaches that can reproduce the dynamics of geographical phenomena such as crime patterns, epidemic spread, environmental dynamics, land-use change, and urban growth (Verburg and Veldkamp 2001, Brown and Xie 2006, Liu *et al*. 2017). Cellular automata (CA) are the most widely applied SEMs, especially for the land-use change and urban growth simulation (Wu 2002, Al-Shalabi *et al*. 2013, Basse *et al*. 2014, Clarke *et al*. 2018). CA models are built on transition rules that describe the evolution of geographical phenomena (Myint and Wang 2006, Cao *et al*. 2016, Newland *et al*. 2018a). For land-use modeling, these rules are commonly retrieved using samples of the land-use change and its driving factors as well as based on the interactions among neighboring cells, resulting in useful prediction of alternative scenarios (Verburg and Overmars 2009, Wang *et al*. 2013, Feng and Tong 2017).

Geographical CA models are stochastic if they incorporate a random component and model runs with identical inputs produce different outputs (Vermeiren *et al*. 2016, Mustafa

CONTACT Xiaohua Tong xhtong@tongji.edu.cn; Yongjiu Feng yjfeng@tongji.edu.cn

et al. 2018b); in contrast, they are deterministic if the random component is excluded and model runs with identical inputs produce identical outputs (Omrani *et al*. 2017, Mustafa *et al*. 2018a). CA are affected by the definition of spatial scale, driving factors, neighborhood and transition probability, leading to various simulation outcomes that yield different accuracies and spatial patterns (Feng *et al*. 2011, Basse *et al*. 2014, Barreira-Gonzalez and Barros 2017). Differences exist between the simulated results and the actual results, but in reliable CA models these should be within acceptable limits. This requires accurate calibration by CA parameter adjustment and rigorous assessment that examines whether models are in line with their design objectives.

In the past two decades, several CA packages have been developed to simulate the land-use and urban dynamics. These include UrbanSim, Dinamica EGO, SLEUTH, CLUE-S, CA-Markov in IDRISI, FLUS, and UrbanCA (Verburg *et al*. 2002, Waddell 2002, Dietzel and Clarke 2007, Liu *et al*. 2017, Clarke *et al*. 2018, Feng and Tong 2019). All the packages include their calibration, validation and assessment procedures, and some have been widely applied for a long time. The model application needs rigorous calibration that adjusts the internal parameters to improve the model's performance. In practice, CA models are calibrated using an actual map at the initial time (T0), an actual map at the final time (T1), and a set of driving factors; at the end of calibration, the simulated map at time T1 is compared with the actual map at time T1 to assess the model performance (Hagen-Zanker and Lajoie 2008, Van Vliet *et al*. 2011), which is also considered model validation of verifying the truth of the modeling results (Oreskes *et al*. 1994, Rykiel 1996, Oreskes 1998). Independent model validation requires the comparison between the simulation and the actual at future time T2 (Van Vliet *et al*. 2011). Calibration is challenging due to the complexity of urban dynamics and the complex combinations of driving factors (Verburg *et al*. 2004); as a result, CA models cannot be expected to produce outcomes that match the real world perfectly (Van Vliet *et al*. 2016). However, not all publications include the model assessment procedure because CA studies differ in purpose, method and processing (Verburg and Overmars 2009, Pérez-Molina *et al*. 2017).

Model assessment methods differ from article to article of CA models that incorporate the above procedure. We categorize the most common methods into three major types according to the aspects they aim to evaluate, with each method examining several details. The three are:

(1) *Dataset assessment* that evaluates the input dataset reliability by examining the accuracy and describes the error propagation through the modeling procedure, providing useful insights for controlling the modeling errors.
(2) *Procedure assessment* that evaluates the model run efficiency, CA transition rules, transition probability map and model sensitivity, offering useful approaches to procedure controlling for accurate and robust modeling.
(3) *Result assessment* that compares both the end-state and change between the simulated and actual results by visual inspection, statistical test, spatial pattern analysis and cross-assessment, providing a comprehensive evaluation of the overall model performance.

With regard to the above aspects, we summarize here the model assessment methods used in CA publications from 1999 to 2018, a period that witnessed remarkable growth of studies on CA modeling of land-use and urban dynamics. Through *Google Scholar*, we selected publications on CA model development and application that focused on the simulation of land-use change and urban growth. We present all CA assessment methods and briefly comment on specific methods, aiming to demonstrate how the modelers assessed their CA models in early work and providing a reference to select the appropriate metrics in future work.

2. Selected publications

We searched for articles published between 1999 and 2018 from *Google Scholar* using three keywords 'cellular automata', 'land' and 'urban' simultaneously. We retrieved a total of 347 articles, excluding CA review papers that are not related to assessment. The selected publications include articles from 64 journals, 6 book chapters, 2 books, and 1 conference publication collection. Figure 1 shows that the number of CA papers per year has grown almost exponentially from 5 in 1999 to 53 in 2018. This demonstrates increasing worldwide research interest in modeling land-use change and urban growth using CA models. Among the 64 journals, *International Journal of Geographical Information Science* published the most CA articles, followed by *Computers, Environment and Urban Systems* and *Landscape and Urban Planning* (Figure 2). All top-10 journals published at least 8 CA-related papers while many other journals (36) published only a single article.

3. Model assessment methods

3.1 An overview of the methods

Figure 3 summarizes the assessment methods of CA models with respect to three aspects: 1) the input dataset, 2) the modeling procedure, and 3) the simulated results. A crisp assessment starts with the accuracy analysis of raw dataset to provide reliable inputs for CA modeling. It

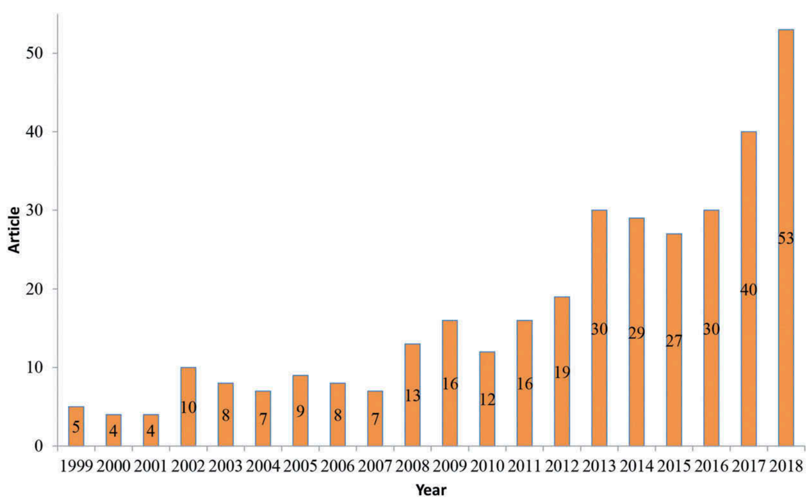

Figure 1. The proliferation of CA articles of land-use and urban modeling in the last 20 years (1999–2018).

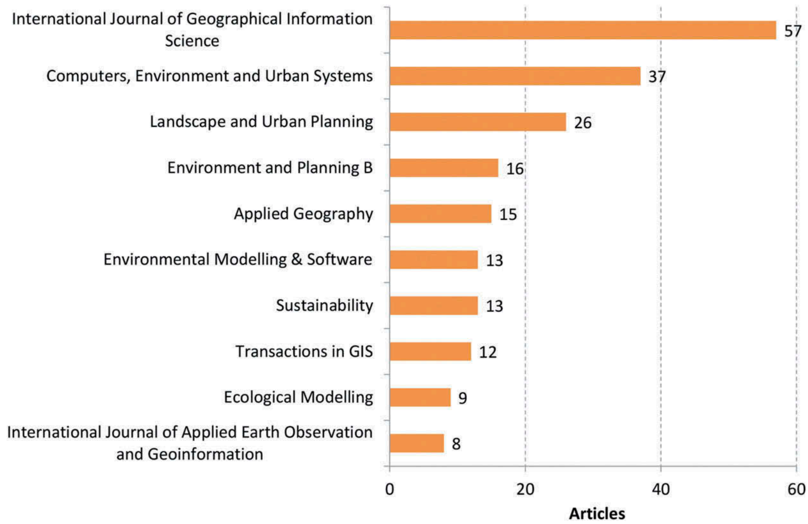

Figure 2. Top-10 journals published the most CA articles during 1999–2018.

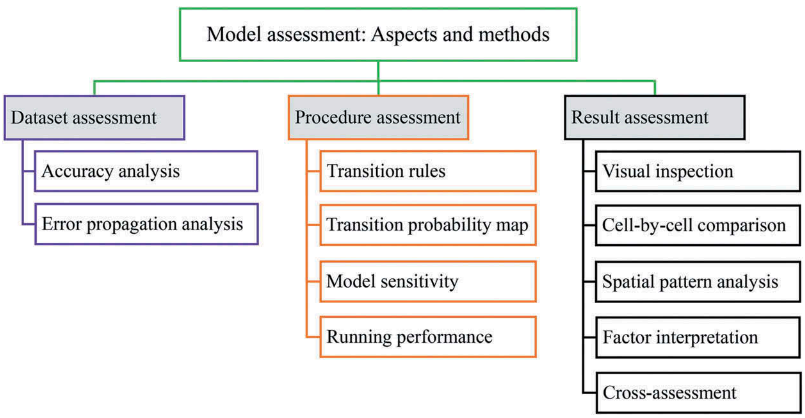

Figure 3. An overview of CA model assessment methods in respect to three aspects.

then evaluates every step of the modeling procedure to control the model's quality. Finally, it assesses the modeling results using qualitative and quantitative methods to report the model performance. In practice, the model assessment and the related metrics can be performed using the CA software mentioned in the introduction as well as a few other software packages such as ArcMap, GeoDa, Map Comparison Kit, PontiusMatrix26, and Fragstats.

3.2 Dataset assessment

The input dataset evaluation is the first step of the model assessment that aims to answer how much the data are reliable to construct accurate CA models. High dataset accuracy implies that the modeling errors may be more caused by the modeling processes instead of the raw dataset. In CA modeling, however, dataset assessment has yet to receive proper attention considering the accuracy analysis and error propagation.

3.2.1 Accuracy analysis

The reliability of simulation results is not only affected by CA model components and their relationships but also impacted by the data source accuracy (Yeh and Li 2006, Tayyebi *et al*. 2014b). Great efforts have been made to assess the influences of model components and behavior (Kocabas and Dragicevic 2006, Wu *et al*. 2019), but few have been made to address the impacts of data source error (Wu *et al*. 2012). Accurate input initial/final land maps and driving factor maps are impacted by the credibility of data source (Congalton and Green 2008), the classification of remote sensing images, and the selection of spatial, thematic, and observational scales (Foody 2002). Methods to assess the accuracy of input land-use maps are largely limited to those (e.g. cell-by-cell comparison) widely used in remote sensing classification. Meanwhile, evaluation of driving factors is definitely necessary to construct accurate CA models; unfortunately, few have discussed this issue so far because there are no references for the driving factors, making the evaluation very challenging.

3.2.2 Error propagation analysis

While the uncertainty sources in CA models are very complex, the input dataset errors may be the initial components of the uncertainties. The considerable errors in input dataset cannot be ignored (Bachmann and Allgöwer 2002), and are needed to be carefully analyzed because they can propagate through model components during the simulation processes (Yeh and Li 2006). The components and processes include scaling, sampling, neighborhood configuration, transition rules, stochastic disturbance, and threshold definition. Although these are significant factors on modeling, there is no effective method to capture the error propagation mechanism that should provide insights for the controlling of uncertainties and the construction of accurate CA models.

3.3 Procedure assessment

CA's procedure denotes several key modeling processes, which include the transition rule construction, transition probability mapping, model sensitivity test, and model running. Each process is crucial to control the quality of CA models. The procedure assessment provides statistics about how well transition rules fit the sampling data and how efficient are the models in generating simulation results. End-users may focus on the quality of the simulation results while the modelers must conduct the procedure assessment if they want to improve the model's quality.

3.3.1 Transition rule

Transition rules are at the core of CA models, and well-fitted rules could consequently result in good models and accurate simulation results. The fitting performance of the transition rules can be reflected by goodness-of-fit (R^2), relative quality, and/or residual distribution (Table 1).

The goodness-of-fit reflects how well the simulation approximates the observation in building CA transition rules (Pijanowski *et al*. 2002, Jantz and Goetz 2005, Grekousis *et al*.

Table 1. Statistics to evaluate the fitting performance of CA transition rules.

No.	Metric	Aspect	Selected publications
1	R^2	Goodness-of-fit	(Verburg *et al.* 1999, Pijanowski *et al.* 2002, Jantz and Goetz 2005, Grekousis *et al.* 2013, Mustafa *et al.* 2018b)
2	AIC	Relative quality	(Feng and Tong 2017, Feng *et al.* 2018b)
3	RMSE	Residual	(Thapa and Murayama 2012, Al-Ahmadi *et al.* 2016, Crols *et al.* 2017, Feng 2017, Rimal *et al.* 2018)
4	Standard error	Residual	(Jafari *et al.* 2016, Feng *et al.* 2018b)
5	Mean absolute error	Residual	(Aljoufie *et al.* 2013, Al-Ahmadi *et al.* 2016, Li *et al.* 2018)
6	Fitness score	Residual	(Blecic *et al.* 2015, Whitsed and Smallbone 2017, Clarke *et al.* 2018)
7	Moran's I	Residual	(Feng and Tong 2017)

Note: These metrics are only applicable when the transition rules are built using 1) statistical regression methods and 2) heuristic methods that optimize the regression-based CA parameters.

2013, Shafizadeh-Moghadam *et al.* 2017), and delineates the fitting performance of the transition rules in cases they are defined using statistical methods such as logistic regression, spatial autoregressive model and geographically weighted regression. The Akaike information criterion (AIC) measures the relative quality of statistical models (Akaike 2011) where a smaller AIC denotes a better model. In CA modeling, AIC was used to select the model that best approximates the land-use dynamics from a set of competitive candidate CA models (Feng and Tong 2017).

The residuals of transition rules measure the deviation between the predicted land change and the actual land change based on the samples selected for model training. Early publications have applied root-mean-square error (RMSE), standard error, mean absolute error, fitness score, and Moran's I to analyze the residuals (Thapa and Murayama 2012, Crols *et al.* 2017, Feng *et al.* 2018b). In these approaches, three describe the differences between the estimated transition probability and the actual transition (Thapa and Murayama 2012, Crols *et al.* 2017), which is 0 for no-change and 1 for change.

- RMSE is the standard deviation of residuals that denote how concentrated the samples are around the best-fitted curve;
- Standard error (SE) is the approximate standard deviation between each sample and all candidate samples; and
- Mean absolute error (MAE) is the amount of differences between the predicted values and the actual values.

The fitness score is a final fitness (objective) function value in heuristic algorithms that search for the near-optimal CA parameters. The score is usually calculated as the modeling error (RMSE, SE or MAE) of CA transition rules (Feng and Tong 2019). A smaller score indicates a smaller error, hence better fitting performance of the CA transition rules.

Moran's I is a measure of the spatial pattern of the model residuals. This statistic denotes a value ranging from −1 (dispersion) to 1 (perfect correlation). Moran's I = 0 represents that the modeling residuals are randomly distributed so that the transition rules are not spatially biased while Moran's I > 0 indicates possible clustering of the modeling residuals. Geary's C is another statistic that denotes spatial autocorrelation but has yet to be used in assessing CA transition rules. In addition, some publications evaluate transition rules by comparing their parameters directly.

3.3.2 Transition probability map

Transition probability maps are spatially visualized layers produced based on transition rules. The relative quality of such maps can be assessed using ROC and TOC (Table 2). ROC is a threshold-based diagnostic method to assess remote sensing classification algorithms and spatial simulation models (Pontius and Si 2014). ROC has proven useful in assessing the quality of transition probability maps in CA models (Tayyebi and Pijanowski 2014, Shafizadeh-Moghadam *et al*. 2017). In ROC, the transition probability map of each land category is classified using multiple thresholds to generate different duplicates (include target and non-target classes), which are compared with the actual land-use patterns. The comparison generates two classes: correctly simulated target cells (CPC) and simulated target but actual non-target cells (STAN). They correspond to the percentage of CPC in the actual target cells, and the percentage of STAN in the actual non-target cells. The area under the ROC curve (AUC) is a summary statistic ranging from 0.5 to 1.0, where AUC = 0.5 suggests a random model and AUC = 1.0 suggests a perfect. To draw a ROC curve, only the percentage of CPC (y-axis) and the percentage of STAN (x-axis) under multiple thresholds are needed. AUC between 0.5–0.7 indicates a low accuracy, AUC between 0.7–0.9 indicates a moderate accuracy, and AUC > 0.9 denotes a high accuracy.

Pontius and Si (2014) modified ROC to propose TOC by providing the size of every entry in the cross-tabulation matrix for each threshold. As a result, TOC provides all of the information that ROC can reveal but also provides additional important information. To draw a TOC curve, only the CPC (y-axis) and the CPC+STAN (x-axis) under multiple thresholds are needed. TOC has now been accepted in evaluating CA-based land transition probability maps (Kamusoko and Gamba 2015, Liu *et al*. 2017).

3.3.3 Model sensitivity

Model sensitivity is a typical feature of CA models because they are highly affected by the selection of spatial scale, neighborhood configuration, and probability threshold (Wickramasuriya *et al*. 2009, Basse *et al*. 2014, Barreira-Gonzalez and Barros 2017, Xu and Brown 2017, Xia *et al*. 2019). Model sensitivity analysis examines what extent are CA models influenced by the scale, neighborhood and threshold (Jantz and Goetz 2005, Stevens and Dragićević 2007, Mondal *et al*. 2017, Wu *et al*. 2019). To address the scale sensitivity, a multiple scale approach was applied to quantify the degree of similarity among complex spatial patterns (Verburg *et al*. 2002). A moving-window method was also applied to examine the quality of CA models through the spatial variation on all scales (De Almeida *et al*. 2003). Thematic scale (land-use category aggregation) also affect both the CA modeling and the selection of assessment metrics substantially. Regarding this issue, Pontius and Malizia (2004) arranged five principles that dominate the effect of thematic scale on land-use change modeling and analysis, while Aldwaik *et al*. (2015) further developed a computer program to assist the proper land-use categorization.

Table 2. ROC and TOC for evaluating the land transition probability.

No.	Method	Aspect	Selected publications
1	ROC	Relative quality	(Tayyebi and Pijanowski 2014, Al-Sharif and Pradhan 2015, Rienow and Goetzke 2015, Sakieh *et al*. 2015b, Goodarzi *et al*. 2017, Clarke *et al*. 2018)
2	TOC	Relative quality	(Pontius and Si 2014, Kamusoko and Gamba 2015, Liu *et al*. 2017)

The neighborhood impact is particularly important to CA models because they are bottom-up techniques depend on the interactions among nearby cells (Stevens and Dragićević 2007, Liao *et al*. 2014, Barreira-Gonzalez and Barros 2017). Different neighborhood configurations (e.g. square, circular, triangular, and irregular neighbors) can lead to different accuracies, spatial patterns, and landscape structures of simulation results. Modelers therefore examined the neighborhood effects and the related model behaviors (Moreno *et al*. 2009, Basse *et al*. 2016, Pinto *et al*. 2017). Wu *et al*. (2012) noted that larger neighborhood size and planar shape contribute to higher prediction accuracy because such configuration has greater ability to resist the propagation of data source error. As compared with a conventional configuration, a decaying CA neighborhood shows higher prediction ability for urban land, in terms of overall accuracy and Kappa coefficient (Liao *et al*. 2014). Neighborhood strongly was found affecting the simulation by measuring the neighborhood effect in irregularly-spaced CA models (Barreira-Gonzalez and Barros 2017). The papers listed in Table 3 should improve our understanding of CA models and offer a justifiable selection regarding the neighborhood configuration for accurate land-use change and urban growth simulation.

Moreover, model results are significantly affected by the threshold that determines whether a cell can transform its state at the next time. For the transition probability-based CA models, this threshold is a benchmark for land transition (Tayyebi and Pijanowski 2014, Mondal *et al*. 2017); whereas for the fuzzy transition rule-based CA models, the threshold is a benchmark for factor attributes (Cao *et al*. 2016). Because defining an appropriate threshold is challenging, modelers have attempted to apply multiple thresholds, dynamic thresholds, and random thresholds to identify their optimum values (De Almeida *et al*. 2003, Arsanjani *et al*. 2013). Many studies applied multiple thresholds and analyzed their effects on the model behavior and simulation results, finally suggesting the optimum threshold. A weaker threshold sensitivity is preferred because it indicates greater robustness of CA models.

3.3.4 Running performance

CA models require intensive computing resources because they use an iterative modeling process (Guan *et al*. 2016), especially when the models are integrated with artificial intelligence algorithms. Consequently, the run time and computational efficiency of models are essential for their application because the run time clearly indicates the computing performance. This leads to the development of parallel and high-performance computing CA models in recent years (Guan *et al*. 2016, Xia *et al*. 2018).

3.4 Result assessment

Result assessment is the core of the CA model assessment because it provides modelers and users how accurate are the simulation results and what extent the models can be

Table 3. Model sensitivity related to scale, neighborhood configuration and probability threshold.

No.	Metric	Aspect	Selected publications
1	Scale sensitivity	Scale	(Verburg *et al*. 2002, Jantz and Goetz 2005, Kocabas and Dragicevic 2006, Poelmans and Van Rompaey 2009, Wickramasuriya *et al*. 2009, Feng *et al*. 2011, Rabbani *et al*. 2012, Carter 2018)
2	Neighborhood sensitivity	Neighborhood	(Verburg *et al*. 2004, Kocabas and Dragicevic 2006, Poelmans and Van Rompaey 2009, Mantelas *et al*. 2012, Liao *et al*. 2014, Basse *et al*. 2016, Barreira-Gonzalez and Barros 2017, Newland *et al*. 2018b)
3	Threshold sensitivity	Threshold	(Jantz *et al*. 2004, Moreno *et al*. 2008, Rabbani *et al*. 2012, Al-Sharif and Pradhan 2015, Feng *et al*. 2016, Yin *et al*. 2018)

used for decision making. The result assessment can be realized by several methods including visual inspection, cell-by-cell map comparison, spatial pattern analysis, factor contribution analysis, and cross-assessment. These methods delineate three aspects of the simulation results: end-state, change, and the explanatory ability of driving factors.

3.4.1 Visual inspection

Visual inspection intuitively examines the similarities and differences in overall patterns of the simulated results as compared with the actual results. It has been acknowledged that the human eye is an amazingly powerful tool to detect the differences and similarities between two or more maps across spatial scales (Straatman *et al*. 2004). In a case study conducted by Straatman *et al*. (2004), for example, visual inspection only was applied to evaluate the simulation results. This method is arguably the best approach to the simulation result evaluation; however, it is highly subjective because visual inspection may be different from people to people (Van Vliet *et al*. 2011).

3.4.2 Cell-by-cell map comparison

A cell-by-cell comparison method detects whether the state of each cell matches between the simulations and the observations, and generates a cross-tabulation matrix with a few accuracy and error statistics. This method is also called cell-to-cell comparison (Musa *et al*. 2017), pixel-by-pixel comparison (Feng *et al*. 2011), and grid-by-grid comparison (Liu *et al*. 2008).

3.4.2.1. Cross-tabulation matrix. Cell-by-cell comparison commonly produces a cross-tabulation (contingency) matrix that has also been called an error matrix or a confusing matrix (Charif *et al*. 2017). Early publications apply the cross-tabulation matrix to demonstrate the differences in each category between the simulated results and the actual results (Rabbani *et al*. 2012, Pinto *et al*. 2017), intuitively showing the errors in cells, percentages or areas. In the literature, many studies present only the cross-tabulation matrix instead of the statistics (e.g. overall accuracy) generated from the matrix (Myint and Wang 2006, Rabbani *et al*. 2012, Charif *et al*. 2017, Pinto *et al*. 2017).

The cross-tabulation matrix (Table 4) consists of J rows and C columns, a standard table for accuracy evaluation. For remote sensing image classification, the matrix compares the classified results with the actual results (Congalton and Green 2008); for land-use simulation, the matrix compares the simulated results with the actual results (Bozkaya *et al*. 2015, Feng 2017). Each column in the matrix represents an actual category, with the column sum indicating the total quantity; each row represents a predicted category, with the row sum indicating the total

Table 4. The cross-tabulation matrix derived from the cell-by-cell comparison.

	Actual (Reference)					
Simulated (Classification)	L1	L2	L3	…	LJ	Row total
L1	N_{11}	N_{12}	N_{13}	…	N_{1J}	T_{R1}
L2	N_{21}	N_{22}	N_{23}	…	N_{2J}	T_{R2}
L3	N_{31}	N_{32}	N_{33}	…	N_{3J}	T_{R3}
…	…	…	…	…	…	…
LJ	N_{J1}	N_{J2}	N_{J3}	…	N_{JJ}	T_{RJ}
Column total	T_{C1}	T_{C2}	T_{C3}	…	T_{CJ}	Sum

Note: N_{ij} means the number of cells that are Class Li in the simulated results but Class Lj in the actual results, Tri denotes the total cells of the simulated class Li, and T_{Cj} denotes the total cells of the actual class Lj.

Table 5. Cross-tabulation matrix-based decomposition of quantity and allocation ability of CA models (Pontius 2000). Each row denotes the proportion of correctly classified cells in simulations with different (perfect, medium or no) allocation ability.

	Quantity		
Allocation	No quantity (NQ)	Medium quantity (MQ)	Perfect quantity (PQ)
Perfect allocation (PA)	$\sum_{i=1}^{J}\left(\frac{1}{J}, T_{Ci}\right)$	$\sum_{i=1}^{J}(T_{Ci}, T_{Ri})$	1
Medium allocation (MA)	$\frac{1}{J} + \text{Klocation}(\text{NQPA})$	OA	$PQNA + \text{Klocation}(1 - \text{PQNA})$
No allocation (NA)	$\frac{1}{J}$	$\sum_{i=1}^{J}(T_{Ci} \times T_{Ri})$	$\sum_{i=1}^{J}(T_{Ci})^2$

Note: NQPA denotes the percentage of cells with no quantity agreement and perfect allocation agreement, PQNA denotes the percentage of cells with perfect quantity agreement and no allocation agreement, OA denotes the overall accuracy, and Klocation denotes the Kappa for location.

quantity. The values on the major diagonal indicate the number of cells where the predicted (classified) class exactly matches the actual class.

In addition, the simulation ability of CA models can be explained by both the quantity and allocation aspects that respectively reflect the quantitative differences and allocation differences between the simulated results and the actual results. According to Pontius (2000), in Table 5, the PA row denotes the proportion of correctly classified cells in simulations with perfect allocation ability, which means all cells of category A in the simulated maps are cells of same category A in the actual maps; the MA row denotes the proportion of correctly classified cells in simulations with medium allocation ability, which means more cells of category A, than would be expected by chance, in the simulated maps are cells of same category A in the actual maps; the NA row denotes the proportion of correctly classified cells in simulations with no allocation ability, which means no more cells of category A, than would be expected by chance, in the simulated maps are cells of same category A in the actual maps. Similarly, the PQ row denotes the simulations with perfect quantity ability, which means the cell quantity of category A in the simulated maps is the same as in the actual maps; the MQ row denotes the simulations with medium quantity ability, which means the cell quantity of category A in the simulated maps is not the same as in the actual maps; the NQ row denotes simulations with no quantity ability, which means the cell quantity of category A in the simulated maps is zero in the actual maps.

3.4.2.2. Kappa coefficient. Kappa coefficients evaluate how well land classification or modeling performs excluding chance agreement (Van Vliet *et al*. 2011). This statistic ranges from −1 (significantly worse than random) to 1 (perfect), but it typically lies between 0 and 1. Gwet (2014) noted that Kappa agreement can be categorized into five different levels: 0.0 ~ 0.20 (slight agreement), 0.21 ~ 0.40 (fair agreement), 0.41 ~ 0.60 (moderate agreement), 0.61 ~ 0.80 (substantial agreement), and 0.8 ~ 1.0 (almost perfect agreement). Five variants of Kappa coefficients were defined by Pontius and Millones (2011) and have been widely applied in model assessment: standard Kappa (Kstandard), Kappa for histogram (Khisto), Klocation, Kappa for quantity (Kquantity), and Kappa for no ability (Kno):

$$\begin{cases} \text{Kstandard} = \frac{OA-NAMQ}{1-NAMQ} \\ \text{Khisto} = \frac{1-QD-NAMQ}{1-NAMQ} \\ \text{Kquantity} = \frac{OA-MANQ}{MAPQ-MANQ} \\ \text{Klocation} = \frac{OA-NAMQ}{PAMQ-NAMQ} \\ \text{Kno} = \frac{OA-NANQ}{1-NANQ} \end{cases} \quad (1)$$

Kstandard is a typical statistic that tests inter-rater reliability and is calculated as the actual proportion correct to the expected proportion due to change (Pontius and Millones 2011). Khisto is a function of the histogram of the matrix's proportions of the categories, aiming to measure the quantitative similarity of two maps in comparison (Hagen-Zanker 2002). Similarly, Kquantity measures the model's ability to designate quantity divided by the perfect ability to designate quantity (Pontius 2000). For example, Kquantity = 1 if the predicted quantity for each category is the same as the actual quantity. Focusing on the location assessment, Klocation measures the model's ability to correctly allocate cells divided by the perfect ability to allocate cells. Kno denotes the proportion predicted correctly relative to the expected proportion predicted correctly by a Null model (Pontius 2000). Van Vliet *et al.* (2009) noted that these Kappa statistics consider slight displacements as errors, but such displacements can be considered approximately correct from a modeler's perspective. The fuzzy Kappa, i.e. Kfuzzy, calculates the fuzzy similarity between two categorical maps by ignoring the slight displacements (Hagen-Zanker 2003). Since 2007, many studies have applied Kfuzzy to validate CA models and assess the simulation results (Ménard and Marceau 2007, Petrov *et al.* 2009). Van Vliet *et al.* (2011) then developed a Fuzzy Kappa simulation (Ksimulation) to evaluate the agreement between the simulated and actual land-use change. This metric is particularly useful in counting the cells simulated correctly from those with state change during the modeling period (Lauf *et al.* 2012). Kfuzzy and Ksimulation can be calculated using Map Comparison Kit 3 (http://mck.riks.nl). All seven Kappa statistics have been widely applied in the assessment of geographical CA models (Table 6).

3.4.2.3. Accuracy and error. In addition to the Kappa coefficients, eleven more metrics can be derived from the cross-tabulation matrix to delineate the modeling success or failure. Of these, the success includes overall accuracy (OA), producer's accuracy (PAC), user's accuracy (UA), chance agreement (CHA), allocation agreement (AA), and quantity agreement (QA), while the failure includes total error (TE), omission error (OE), commission error (CE), quantity disagreement (QD), and allocation disagreement (AD). The calculation methods are given by (Pontius 2000, Congalton and Green 2008):

$$\begin{cases} \text{Overall accuracy: } OA = \left(\frac{N_{11}}{Sum} + \frac{N_{22}}{Sum} + \frac{N_{33}}{Sum} + \ldots + \frac{N_{JJ}}{Sum}\right) \times 100\% \\ \text{Producer's accuracy: } PAC = \frac{N_{JJ}}{T_{CJ}} \times 100\% \\ \text{User's accuracy: } UA = \frac{N_{JJ}}{T_{RJ}} \times 100\% \\ \text{Chance agreement: } CHA = MIN\left(\frac{1}{J}, OA, NAMQ\right) \times 100\% \\ \text{Allocation agreement: } AA = MAX((OA - NAMQ), 0) \times 100\% \\ \text{Quantity agreement: } QA = (OA - CHA - AA) \times 100\% \\ \text{Total error: } TE = (1 - OA) \times 100\% \\ \text{Omission error: } OE = (1 - PA) \times 100\% \\ \text{Commission error: } CE = (1 - UA) \times 100\% \\ \text{Quantity disagreement: QD} = (PAPQ - PAMQ) \times 100\% \\ \text{Allocation disagreement: AD} = (TE - \text{QD}) \times 100\% \end{cases} \quad (2)$$

The diagonal cells in Table 4 are those simulated correctly and the off-diagonal cells are those simulated erroneously. OA is calculated by dividing the number of correctly

Table 6. Seven Kappa coefficients used to validate CA models and assess the simulation results.

No.	Metric	Aspect	Selected publications
1	Kstandard	Integrated	(Barredo and Demicheli 2003, Myint and Wang 2006, Kocabas and Dragicevic 2007, Liu *et al*. 2008, Petrov *et al*. 2009, Lauf *et al*. 2012, Mantelas *et al*. 2012, Rabbani *et al*. 2012, Chaudhuri and Clarke 2013, Dahal and Chow 2015, Rienow and Goetzke 2015, Sakieh *et al*. 2015b, Feng *et al*. 2016, Gharbia *et al*. 2016, Ku 2016, Dezhkam *et al*. 2017, Hyandye and Martz 2017, Xie *et al*. 2018)
2	Khisto	Quantity	(Hagen-Zanker 2002, Petrov *et al*. 2009, Lauf *et al*. 2012, Wang *et al*. 2013, Chaudhuri and Clarke 2014, Rienow and Goetzke 2015, Clarke *et al*. 2018)
3	Klocation	Location	(Batisani and Yarnal 2009, Petrov *et al*. 2009, Lauf *et al*. 2012, Ahmed *et al*. 2013, Chaudhuri and Clarke 2013, Chaudhuri and Clarke 2014, Rienow and Goetzke 2015, Ku 2016, Al-Ageili *et al*. 2017, Hyandye and Martz 2017, Clarke *et al*. 2018)
4	Kquantity	Quantity	(Batisani and Yarnal 2009, Ahmed *et al*. 2013)
5	Kno	No change	(Batisani and Yarnal 2009, Ahmed *et al*. 2013, Ku 2016, Al-Ageili *et al*. 2017, Hyandye and Martz 2017, Rimal *et al*. 2018)
6	Kfuzzy	Integrated	(Ménard and Marceau 2007, Petrov *et al*. 2009, Wickramasuriya *et al*. 2009, Lauf *et al*. 2012, Ahmed *et al*. 2013, Newland *et al*. 2018b)
7	Ksimulation	Change	(Van Vliet *et al*. 2011, Mantelas *et al*. 2012, Van Vliet *et al*. 2013b, Chaudhuri and Clarke 2014, Blecic *et al*. 2015, Kamusoko and Gamba 2015, Van Vliet *et al*. 2016, Wang *et al*. 2018)

predicted cells (on the major diagonal) by the total of cells (Congalton and Green 2008), indicating the overall performance of a CA model or the goodness of the result. PAC is calculated by dividing the number of cells correctly classified in each category (on the principal diagonal) by the number of reference cells in that category. Therefore, PAC represents how well the reference cells in each class are predicted (Congalton and Green 2008). UA is computed by dividing the number of cells correctly predicted in each category by the total number of pixels that are predicted in that category. Therefore, UA represents the probability that a cell predicted into a given category actually represents that category in reality (Congalton and Green 2008). A few publications use different names for UA: goodness-of-fit (Liu *et al*. 2008) and accuracy (Ke *et al*. 2016). CHA is the percentage agreement (between the simulated results and the actual results) that would be expected by chance (Pontius 2000). AA and QA are attributed to the model's correct allocation and correct quantity prediction, respectively. QA is also called the non-spatial metric in some publications (Bradley *et al*. 2016). In Table 7, we present the eleven metrics derived from the cross-tabulation matrix, and selected many publications that apply these metrics to validate CA models.

3.4.2.4. Map overlay. The spatial overlay between the simulated results and the actual results could reveal differences in structural conformity (Wu 2002). By evaluating land change, a three-map overlaying method derives the quantity and allocation components of the simulation successes and errors (Pontius *et al*. 2008, 2011). The three maps are: [1] the actual map of the start year, [2] the actual map of the end year, and [3] the predicted map of the end year. The overlay generates four metrics, including hit, false alarm, miss, and correct rejection, which are respectively known in the contingency matrix as true positive, false positive, false negative and true negative values (Puertas *et al*. 2014). The overlay also generates other three metrics including the figure of merit (FOM), precision, and recall (Charif *et al*. 2017). These seven metrics are given by (Pontius *et al*. 2008):

Table 7. Success and error metrics that focus on the end-state assessment.

No.	Metric	Aspect	Selected publications
1	Overall accuracy	Successes	(Myint and Wang 2006, Moreno *et al*. 2008, Ahmed and Ahmed 2012, Rabbani *et al*. 2012, Thapa and Murayama 2012, Al-Shalabi *et al*. 2013, Feng *et al*. 2016, Gharbia *et al*. 2016, Dezhkam *et al*. 2017, Pinto *et al*. 2017, Rimal *et al*. 2018)
2	Producer's accuracy	Successes	(Jantz *et al*. 2004, Myint and Wang 2006, Mundia and Murayama 2010, Ahmed and Ahmed 2012, Al-Shalabi *et al*. 2013, Feng *et al*. 2016, Gharbia *et al*. 2016, Al-Ageili *et al*. 2017, Yin *et al*. 2018)
3	User's accuracy	Successes	(Jantz *et al*. 2004, Myint and Wang 2006, Liu *et al*. 2010, Mundia and Murayama 2010, Ahmed and Ahmed 2012, Al-Shalabi *et al*. 2013, Feng *et al*. 2016, Gharbia *et al*. 2016, Al-Ageili *et al*. 2017, Rimal *et al*. 2018)
4	Quantity agreement	Successes	(Batisani and Yarnal 2009, Feng *et al*. 2011, Zhang *et al*. 2011, Hyandye and Martz 2017, Mondal *et al*. 2017, Newland *et al*. 2018a)
5	Allocation agreement	Successes	(Batisani and Yarnal 2009, Hyandye and Martz 2017, Mondal *et al*. 2017, Newland *et al*. 2018a)
6	Chance agreement	Successes	(Batisani and Yarnal 2009, Feng *et al*. 2011, Ahmed *et al*. 2013, Hyandye and Martz 2017, Mondal *et al*. 2017, Newland *et al*. 2018a)
7	Total error	Error	(Straatman *et al*. 2004, Myint and Wang 2006, Ahmed and Ahmed 2012, Mantelas *et al*. 2012, Meentemeyer *et al*. 2013, Chaudhuri and Clarke 2014, Basse *et al*. 2016, Feng *et al*. 2016, Charif *et al*. 2017)
8	Commission error	Error	(Jantz *et al*. 2004, Dahal and Chow 2014, Feng *et al*. 2016, Gharbia *et al*. 2016, Liu and Feng 2016)
9	Omission error	Error	(Jantz *et al*. 2004, Dahal and Chow 2014, Feng *et al*. 2016, Gharbia *et al*. 2016, Liu and Feng 2016)
10	Quantity disagreement	Error	(Batisani and Yarnal 2009, Meentemeyer *et al*. 2013, Chaudhuri and Clarke 2014, Hyandye and Martz 2017, Mondal *et al*. 2017, Xie *et al*. 2018)
11	Allocation disagreement	Error	(Batisani and Yarnal 2009, Ahmed *et al*. 2013, Meentemeyer *et al*. 2013, Chaudhuri and Clarke 2014, Feng 2017, Feng and Tong 2017, Hyandye and Martz 2017, Mondal *et al*. 2017, Xie *et al*. 2018)

$$\begin{cases} \text{Hit: The percentage of change predicted correctly (\%)} \\ \text{False alarm: The percentage of persistence predicted as change (\%)} \\ \text{Miss: The percentage of change predicted as persistence (\%)} \\ \text{Correct rejection: The percentage of persistence predicted correctly (\%)} \\ \text{Figure of merit (FOM)} = \frac{\text{Hit}}{\text{Hit+Miss+False alarm+ACIC}} \times 100\% \\ \text{Precision} = \frac{\text{Hit}}{\text{Hit+False alarm}} \times 100\% \\ \text{Recall} = \frac{\text{Hit}}{\text{Hit+Miss}} \times 100\% \end{cases} \tag{3}$$

where ACIC indicates the actual change for a land category predicted as an incorrect land category.

The above calculations are easily performed using GIS-based software programs such as (e.g. ArcMap). Compared with OA, the Hit metric considers only the change and excludes Null (no-change) successes (Pontius *et al*. 2008). The False alarm, Miss, and Correct rejection (Null success) metrics are easily understood as their calculations presented in Equation (3). The total simulation errors are the sum of allocation disagreement and quantity disagreement, and are the sum of False and Miss. FOM focuses on the change in land-use or urban pattern rather than the continuity of land classes. To calculate FOM, the numerator is the intersection of the actual change and the predicted change, and the denominator is the union of the actual change and the predicted change. In a case study, Pontius *et al*. (2008) noted that land models have satisfactory predictive power when FOM is greater than 21%. Precision denotes how many changed cells are correctly predicted through all predicted changed cells, while Recall denotes how many changed cells are correctly predicted through all actual changed cells (Charif *et al*. 2017).

Table 8. Three-map overlaying comparison metrics that focus on the change assessment.

No.	Metric	Aspect	Selected publications
1	Hit	Success	(Thapa and Murayama 2011, García *et al*. 2012, Meentemeyer *et al*. 2013, Kamusoko and Gamba 2015, Liu and Feng 2016, Feng 2017, Shafizadeh-Moghadam *et al*. 2017, Gounaridis *et al*. 2018)
2	Null success	Success	(Thapa and Murayama 2011, García *et al*. 2012, Ahmed *et al*. 2013, Kamusoko and Gamba 2015, Liu and Feng 2016, Shafizadeh-Moghadam *et al*. 2017)
3	FOM	Success	(Thapa and Murayama 2011, García *et al*. 2012, Wang *et al*. 2013, Chen *et al*. 2014, Kamusoko and Gamba 2015, Liu *et al*. 2017, Shafizadeh-Moghadam *et al*. 2017, Liu *et al*. 2018)
4	Precision	Success	(Charif *et al*. 2017, Omrani *et al*. 2017)
5	Recall	Success	(Charif *et al*. 2017, Omrani *et al*. 2017)
6	Miss	Error	(Thapa and Murayama 2011, García *et al*. 2012, Blecic *et al*. 2015, Kamusoko and Gamba 2015, Feng 2017, Gounaridis *et al*. 2018)
7	False alarm	Error	(Thapa and Murayama 2011, García *et al*. 2012, Ahmed *et al*. 2013, Chen *et al*. 2014, Blecic *et al*. 2015, Kamusoko and Gamba 2015, Feng 2017, Shafizadeh-Moghadam *et al*. 2017, Gounaridis *et al*. 2018)

Table 8 lists a few representative CA publications that use the three-map overlaying metrics in the model assessment.

3.4.3 Spatial pattern analysis

While the cross-tabulation matrix can measure the agreement between the simulated maps and the actual maps, it does not capture their similarities in spatial patterns. For land maps, landscape metrics can be used to characterize the landscape structures and spatial autocorrelation statistics that measure the degree to which one feature is similar to nearby features. As a result, the combined use of landscape metrics and spatial autocorrelation statistics improve our understanding of the overall spatial patterns of simulation results.

3.4.3.1. Spatial autocorrelation statistic. The spatial interdependence of the simulation results is another essential aspect of model assessment. Spatial autocorrelation refers to the potential interdependence of observations in the same region (Getis 2010). Spatial autocorrelation statistics measure the degree of interdependence between a data point and its adjacent neighbors. Three statistics measuring global spatial autocorrelation are in common use, including Moran's I, Geary's C and Getis-Ord general G (Table 9). Among these, Moran's I has been most frequently applied to the comparison of spatial clustering between the simulations and observations (Wu 2002, Aljoufie *et al*. 2013, Dahal and Chow 2015, Ku 2016). Moran's I explores the spatial distribution of land-use and urban patterns, and whether they are clustered, dispersed or randomly distributed. This statistic ranges from −1 (dispersion) to 1 (perfect correlation), with zero denoting a random spatial pattern (Getis and Aldstadt 2010). Moran's I usually generates a z-score and a p-value to evaluate whether the spatial autocorrelation is statistically significant (Bradley *et al*. 2016). Similar to Moran's I, Geary's

Table 9. Spatial autocorrelation statistics for validating CA models.

No.	Metric	Aspect	Selected publications
1	Moran's I	Spatial autocorrelation	(Wu 2002, Aljoufie *et al*. 2013, Dahal and Chow 2015, Bradley *et al*. 2016, Ku 2016, Du *et al*. 2018)
2	Geary's C	Spatial autocorrelation	(Bradley *et al*. 2016)

C measures spatial autocorrelation by analyzing the interdependence among geographic observations based on a spatial weight matrix (Anselin 2019). Geary's C has a positive value ranging from 0 to 2, with C = 1 denoting no spatial autocorrelation (random distribution). Moreover, C < 1 denotes a positive spatial autocorrelation (nearby observations are similar), and C > 1 denotes a negative spatial autocorrelation (nearby observations are dissimilar). To the best of our knowledge, Getis-Ord general G has not been used in the CA model evaluation. In practice, each category of the simulated and actual maps is assigned an integer where the spatial autocorrelation statistics can be readily computed using software such as ArcMap and GeoDa.

3.4.3.2. Landscape metric. To quantify model performance across landscapes and land-use categories (Brown *et al*. 2005), modelers have applied a set of landscape metrics in the result assessment (Whitsed and Smallbone 2017). These measures have usually been applied to validate CA models regarding the aggregate of patches, classes and landscapes (Soares-Filho *et al*. 2002, Herold *et al*. 2005, Chaudhuri and Clarke 2014, Sakieh and Salmanmahiny 2016). Landscape metrics are categorized as area-edge, shape, aggregation, or diversity. The most frequently applied metrics are presented in Table 10, with a few selected publications. According to Mcgarigal (2014), area-edge metrics reflect the size and edge, shape metrics reflect the complexity, aggregation metrics refer to the spatial aggregation, and diversity metrics quantify the land cover diversity. Detailed explanations and calculations of these landscape metrics can be found in the Fragstats 4.2 manual (www.umass.edu/landeco).

Many other landscape metrics were applied to assess various aspects of the simulation results, with each metric used in at least one article. For example, publications have applied landscape metrics such as patch size standard deviation (Mitsova *et al*. 2011), Shannon's evenness index (Wang and Li 2011), mean patch edge and area-weighted mean shape index (Mitsova *et al*. 2011), mean radius of gyration (Aguilera *et al*. 2011), interspersion and juxtaposition index (Zhou *et al*. 2012), landscape division index (Liu and Feng 2012), percentage of like adjacencies (Chowdhury and Maithani 2014), mean parameter-area ratio (Li *et al*. 2015), average-perimeter-to-area-ratio (Bradley *et al*. 2016), landscape cohesion index (Musa *et al*. 2017), and coefficient of variation (Kocabas and Dragicevic 2007). Modelers have also applied a few other indices modified based on landscape metrics to assess the compactness and shape complexity of the simulated results. These include urban compactness, (radial) fractal dimension (differs from but relates to PAFRAC), entropy, and cluster size (Soares-Filho *et al*. 2002, Van Vliet *et al*. 2009, Chowdhury and Maithani 2014, Hewitt and Diaz-Pacheco 2017). Typically, the modified compactness index (CI) to delineate the urban pattern can be given by (Li *et al*. 2012):

$$CI = \frac{\sqrt{S}}{P} \times 10000 \tag{4}$$

where S is the total urban area (in square meters) and P is the perimeter (in meters) of all urban patches. CI ranges from 0 to 2,821 with a value related to a stronger compact shape, where 2,821 denotes the most compact shape (i.e. circle).

Table 10. Landscape metrics used to evaluate the spatial patterns of simulations.

No.	Metric	Aspect	Selected publications
1	Percentage of landscape (PLAND)	Area-edge	(Herold *et al.* 2005, Ménard and Marceau 2007, Moreno *et al.* 2008, Mundia and Murayama 2010, Aguilera *et al.* 2011, Chaudhuri and Clarke 2013, Sakieh *et al.* 2015b, Dezhkam *et al.* 2017, Goodarzi *et al.* 2017)
2	Edge density (ED)	Area-edge	(Liu *et al.* 2010, Mitsova *et al.* 2011, Chaudhuri and Clarke 2013, Sakieh *et al.* 2015a, Sakieh and Salmanmahiny 2016, Dezhkam *et al.* 2017)
3	Total edge (TE)	Area-edge	(Barredo and Demicheli 2003, Barredo *et al.* 2004, Kocabas and Dragicevic 2007, Ménard and Marceau 2007, Mitsova *et al.* 2011, Whitsed and Smallbone 2017)
4	Mean patch area (AREA_MN)	Area-edge	(Barredo and Demicheli 2003, Kocabas and Dragicevic 2007, Aguilera *et al.* 2011, Mitsova *et al.* 2011, Meentemeyer *et al.* 2013, Dahal and Chow 2015, Li *et al.* 2015, Sakieh *et al.* 2015b, Musa *et al.* 2017, Alaei moghadam *et al.* 2018)
5	Perimeter-area fractal dimension (PAFRAC)	Shape	(Herold *et al.* 2005, Liu *et al.* 2010, Chaudhuri and Clarke 2013, Van Vliet *et al.* 2013b, Li *et al.* 2015, Sakieh *et al.* 2015a, Whitsed and Smallbone 2017)
6	Mean patch shape index (SHPAE_MN)	Shape	(Li *et al.* 2008, Aguilera *et al.* 2011, Lin *et al.* 2011, Dahal and Chow 2014)
7	Number of patches (NP)	Aggregation	(Herold *et al.* 2005, Kocabas and Dragicevic 2007, Moreno *et al.* 2008, Liu *et al.* 2010, Aguilera *et al.* 2011, Mitsova *et al.* 2011, García *et al.* 2012, Chaudhuri and Clarke 2013, Dahal and Chow 2015, Sakieh *et al.* 2015b, Bradley *et al.* 2016, Dezhkam *et al.* 2017, Alaei moghadam *et al.* 2018)
8	Patch density (PD)	Aggregation	(Wang and Li 2011, Liu and Feng 2012, Feng and Tong 2017)
9	Largest patch index (LPI)	Aggregation	(Herold *et al.* 2005, Liu *et al.* 2008, García *et al.* 2012, Li *et al.* 2015, Sakieh and Salmanmahiny 2016, Padmanaban *et al.* 2017)
10	Landscape shape index (LSI)	Aggregation	(Barredo and Demicheli 2003, Liu *et al.* 2010, Wang and Li 2011, Chaudhuri and Clarke 2013, Chowdhury and Maithani 2014, Sakieh and Salmanmahiny 2016, Dezhkam *et al.* 2017)
11	Aggregation index (AI)	Aggregation	(Li *et al.* 2008, Liu *et al.* 2010, Wang and Li 2011, Feng and Tong 2017)
12	Mean Euclidean Nearest Neighbor Distance (ENN_MN)	Aggregation	(Ménard and Marceau 2007, Li *et al.* 2008, Aguilera *et al.* 2011, García *et al.* 2012, Sakieh and Salmanmahiny 2016, Dezhkam *et al.* 2017, Goodarzi *et al.* 2017, Alaei moghadam *et al.* 2018)
13	Contagion index (CONTAG)	Aggregation	(Soares-Filho *et al.* 2002, Herold *et al.* 2005, Liu *et al.* 2008, Chaudhuri and Clarke 2013)
14	Splitting index (SPLIT)	Aggregation	(Barredo and Demicheli 2003, Barredo *et al.* 2004, Sakieh and Salmanmahiny 2016)
15	Clumpiness index (CLUMPY)	Aggregation	(Hewitt and Diaz-Pacheco 2017, Padmanaban *et al.* 2017, Feng *et al.* 2018a)
16	Diversity index (SHDI and SIDI)	Diversity	(Barredo and Demicheli 2003, Mitsova *et al.* 2011, Feng *et al.* 2018a)

To quantify the similarity of the compared maps considering the combined effects of multiple landscape metrics, a relative error index (REI) was proposed as (Sakieh and Salmanmahiny 2016):

$$REI = \left[\frac{M_a - M_s}{M_a}\right] \times 100 \tag{5}$$

where M_a is a landscape metric of the actual map and M_s is the same landscape metric of the simulated map.

Each metric above reports a single value that indicates the overall landscape pattern of each study area. For the spatial assessment, however, moving window-based analysis of spatial variation in landscape patterns is particularly helpful to evaluate the simulation results produced by CA models (Hagen-Zanker 2006, 2016, Hagen-Zanker and Martens 2008).

3.4.4 Factor contribution analysis

The cell-by-cell comparison and spatial analysis focus on the assessment of the end-state and change, where they have limitations because the assessment does not consider the relationships between the simulation results and their driving factors. From the factor perspective, however, we need to be clear about how much the factors contribute to the simulations. On this regard, Feng *et al*. (2019) proposed a novel method by using a generalized additive model (GAM) to quantify the factors' contribution to simulation results. A standard GAM can be given by (Feng *et al*. 2019):

$$Chg(u) = a_0 + SF_1(x_1) + \ \ldots \ + SF_p(x_p) \tag{6}$$

where $Chg(u)$ represents the land change between the initial map and the simulated map, a_0 is a constant, $SF_i(x_i)$ is a smooth function that links $Chg(u)$ and a selected factor (x_i), and p is the number of factors.

Using GAM, the contribution of each factor can be explained by the explained deviance (ED) and AIC, where a higher ED or a lower AIC denotes stronger ability of the factor to explain the simulated results, hence the better performance of the CA models. This method not only can examine the past land patterns but also can evaluate the future land scenarios.

3.4.5 Cross-assessment

Cross-assessment compares a proposed model with other known models by assessing their simulation results under the same conditions. On this regard, the assessment methods described above serve as the criteria for selecting a better model. Among the known CA models, the logistic regression-based model has been selected as the benchmark in numerous publications to compare with other models (Lin *et al*. 2011, Feng 2017). Here, we summarize benchmark models in the absence of similar bottom-up processes like CA (Pontius and Malanson 2005, Hagen-Zanker and Lajoie 2008), and these models are considered as the references for the cross-assessment.

3.4.3.1. A null model. The first type of benchmark models is the Null model that predicts no land change. It uses the actual map at the start time as the simulation map at the end time, then compares the accuracy and error between the Null model and the proposed CA models (Pontius and Malanson 2005). Pontius *et al*. (2008) defined a Null resolution as when the accuracy of a CA model equals that of the Null model, but is less [more] accurate at a finer [coarser] resolution. The authors then applied the Null resolution and FOM to compare 13 different CA models, facilitating the communication among modelers. The Null model has been used as a benchmark to assess CA models in many applications of land-use change and urban growth simulation (Lauf *et al*. 2012, Thapa and Murayama 2012, Van Vliet *et al*. 2013b, Rienow and Goetzke 2015, Bradley *et al*. 2016, Newland *et al*. 2018b).

3.4.3.2. Neutral models. While there are diverse metrics, the model performance they indicated cannot be mutually comparable between different CA models in different study areas; moreover, it is not clear about what extent is the model performance ascribed to the modeling processes (Hagen-Zanker and Lajoie 2008). Also, it has been acknowledged that CA models yielding high overall accuracy may not be good enough if the land state at the end time of simulation is mostly identical to the initial actual state (Hagen-Zanker and Lajoie 2008, Pontius *et al*. 2011). This has been confirmed by Feng *et al*. (2018a) in a most recent study of urban growth where a higher overall simulation accuracy may be related to a lower FOM because the study area has largely maintained its original ecological landscape. To provide the reference for cross-assessment, Hagen-Zanker and Lajoie (2008) proposed Neutral models that can generate land patterns without considering specific land change processes. These Neutral models use the same boundary conditions and constraints that are used in the evaluated CA models. Similar to the Null model, the Neutral models tend to maintain the initial land state; however, differences exist because the Neutral models are stochastic to modify the initial state to meet the predefined constraints. The Neutral models as benchmarks have been applied to the assessment of CA-based simulations and projections (Soares-Filho *et al*. 2013, Wang and Marceau 2013, Hewitt and Diaz-Pacheco 2017, Williams *et al*. 2017), and have been programmed in software for semi-automatic calibration of CA models (He *et al*. 2018, Newland *et al*. 2018b).

3.5 Frequency analysis

3.5.1 Change in methodology over time

While the assessment is crucial in CA modeling, not all publications include this procedure in simulating land-use change and urban growth. Among all 347 articles we reviewed, 37 (~11%) do not include the model assessment (Dietzel and Clarke 2004, Stevens and Dragićević 2007, Verburg and Overmars 2009, Arsanjani *et al*. 2013). Some publications centralize the development of new modeling methods and prototype models (Barreira-GonzaLez *et al*. 2015, De Noronha Vaz and Nijkamp 2015), while others centralize the future scenario prediction under different development strategies where the model assessment is not their focus (Pérez-Molina *et al*. 2017). Although most CA models are built on a gridded space, there are a few CA models based on vector data where assessment is technically challenging (Stevens and Dragićević 2007, Stevens *et al*. 2007).

Table 11 shows that over the past 20 years, the methods used for CA model assessment were increasingly diverse, with more specific metrics used in each paper; however, the assessment of data and running performance yet to receive more attention. The 31 articles published from 1999 to 2003 applied 26 metrics in seven categories, with an average of 0.8 metrics for each article. The three-map overlaying method was not proposed until the 2000s, such that it could not appear in articles published earlier. From 2004 to 2008, the 44 publications used 81 specific metrics in eight categories, with an average of 1.8 metrics for each article. The last decade (2008–2017) witnessed a proliferation of CA papers and corresponding assessment metrics, with about 3 assessment metrics per article. This indicates that model assessment is increasingly a concern of modelers, and that further research of model assessment could have a higher scientific and practical benefit.

Table 11. Application frequency of assessment methods in CA modeling in the past two decades.

Category	Period 1999–2003	2004–2008	2009–2013	2014–2018	Total frequency of use
Procedure assessment					
Transition rule assessment	3	4	9	43	**59**
Transition map assessment (ROC & TOC)	0	2	3	34	**39**
Model sensitivity	2	7	18	29	**56**
Result assessment					
Cross-tabulation matrix	2	18	78	163	**261**
Three-map overlaying	0	0	14	93	**107**
Kappa agreement	3	11	60	114	**188**
Spatial autocorrelation analysis	1	2	1	12	**16**
Landscape analysis	13	35	67	124	**239**
Factor contribution analysis	0	0	0	1	**1**
Null/neutral model	1	2	1	13	**17**
Total frequency of use	**25**	**81**	**251**	**626**	**983**
Total articles	**31**	**44**	**93**	**179**	**347**
Metric per article	**0.8**	**1.8**	**2.7**	**3.5**	

3.5.2 The top-20 most common metrics

Figure 4 shows the top-20 metrics most frequently applied in the CA model assessment. Of these, seven are cell-by-cell comparisons, five are landscape metrics, four are three-map overlays, and the remaining four are ROC, neighborhood sensitivity, Null/neutral model, and Moran's I. This illustrates that the cell-by-cell comparison and landscape analysis are the two most applied categories for CA model assessment. Meanwhile, overall accuracy and Kstandard are the most frequently applied metrics, and half of the papers that use these two metrics report the UA, PAC, AD, and QD metrics. This shows that reporting the total error is not necessary when OA is presented. NP related to landscape pattern ranks third and is the most attractive landscape metric to assess the simulation results. Hit is the most popular among the three-map overlays, suggesting that when reporting Hit, not all papers simultaneously used FOM, False and Miss in the same category.

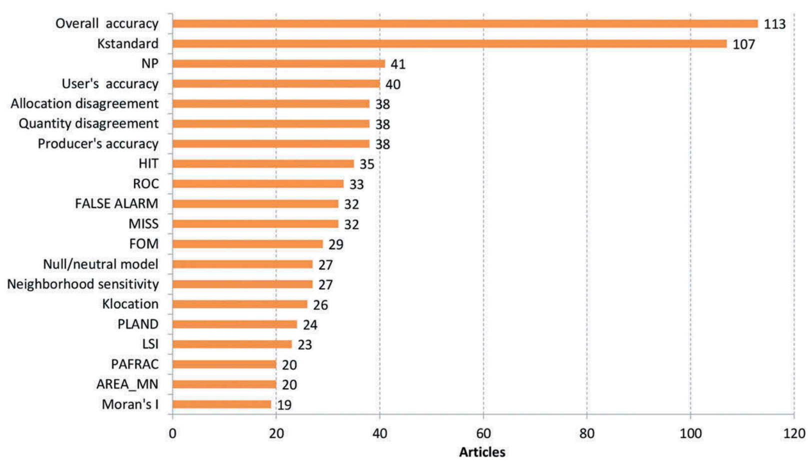

Figure 4. The top-20 metrics most frequently used in CA model assessment.

4. Discussion

4.1 End-state versus change

CA models are aimed at the accurate modeling of land-use and urban patterns by capturing their changes over time. However, the model assessment may likely show contradictions between the end-state and the change. Pontius *et al*. (2008) acknowledged that model performance strongly correlates to the magnitude of change across the simulation period, and Hagen-Zanker and Lajoie (2008) gave an example to show that a high end-state accuracy may be related to a very low change accuracy. This has been further proved by Feng *et al*. (2018a) in a case study of urban growth in a region within one-hour high-speed rail distance from Shanghai. The authors also showed that the same CA model may likely yield varying accuracies and errors of urban growth in different regions; as a result, there are no benchmark of metric values to assess the model performance. However, we suggest change assessment instead of end-state assessment to report the quality of the evaluated CA models. The commonly applied change metrics include Hit, Miss, False and FOM (Pontius *et al*. 2008).

4.2 Further evaluation of Kappa

While Kstandard has been widely applied in the assessment of CA modeling as demonstrated by our frequency analysis, this metric fails to distinguish between quantity error and location error (Pontius 2000). As such, Kappa variants such as Kquantity and Klocation were proposed to overcome the shortcomings. In the literature, these metrics would be used with very different names; for example, the error rate metric (Yang *et al*. 2012) to assess the simulation ability of area match is substantially has similar performance as Kquantity. While Pontius and Millones (2011) noted that all these Kappa coefficients are not useful and suggested using QD and AD in land model assessment, and we found that QD is perfectly related to Kquantity because their sum in percentage is equal to 100%.

Most recent work from McGarigal *et al*. (2018) shows that CA models mainly provide useful comparisons among alternative scenarios, thus high accurate overall patterns instead of exact match in cells can provide useful decision references for formulating urban planning regulations and urban macroeconomic development strategies. To compare the similarity of categorical maps, a Fuzzy approach (Kfuzzy) is strongly recommended to correct the cell-average similarity (Hagen-Zanker 2003). These authors then proposed an improved Kfuzzy that accounts for spatial autocorrelation (Hagen-Zanker 2009) and a Ksimulation approach to assesses land-use transitions by examining their nuance (Van Vliet *et al*. 2013a). In terms of change assessment, the Ksimulation is an appropriate choice except the Hit, Miss, False and FOM metrics.

4.3 Other points of assessment metrics

Except for the transition rule assessment, the goodness-of-fit (R^2) have been widely used to other aspects of CA modeling. These include the evaluation of model sensitivity and fitting performance between the simulations and the observations. For example, modelers usually apply optimum SLEUTH metrics (OSMs) to assess the results generated by the SLEUTH models (Jantz *et al*. 2004, Dietzel and Clarke 2007, Onsted and Clarke 2012, Al-Shalabi *et al*. 2013, Chaudhuri and Clarke 2014). OSMs compare the simulated and actual maps in terms of pattern, cluster, shape, edge, quantity and location. In these metrics, modelers use the least squares regression (LSR) score to denote the goodness-of-fit when fitting two items (e.g. urban area) between the simulated results and the actual results (Dietzel and Clarke 2007). Clearly, the metrics in OSMs are directly related to other measures such as landscape metrics to examine the similar aspects. For instance, OSMs use '% urban' to examine the percentage of the simulated and actual urban area, which can also be measured using class-level PLAND (percentage of landscape). The simulated urban percent is directly divided by actual urban percent, leading to a new metric called percent correct match (PCM) (Al-Sharif and Pradhan 2015, Shafizadeh-Moghadam *et al*. 2017). The annual growth rate between the simulated and actual results are also compared by a division operation (Wang and Murayama 2017). In other studies, cross-correlation (CC) uses a way similar to goodness-of-fit to assess the correlation between the simulated results and the actual results (Guan *et al*. 2005, Kocabas and Dragicevic 2007, Barreira-GonzaLez *et al*. 2015). These metrics enrich the assessment techniques of CA models, but they depict features that can be measured using previously mentioned methods (Tayyebi *et al*. 2014a). On this regard, metrics such as OSMs, PCM and CC can be substituted by metrics derived from cell-by-cell comparison and spatial pattern analysis.

A few other metrics have multiple roles in the CA model assessment. An example of this kind is the spatial autocorrelation statistics that include Moran's I, Geary's C and General G. These statistics can be applied to examine clusters of the residuals and compare the clusters between the simulated results and the actual results. In addition, many landscape metrics in Fragstats are loosely correlated in describing the spatial clustering (Peng *et al*. 2010), and these may be correlated to spatial autocorrelation statistics as well (Fan and Myint 2014).

Assessment methods are substantially affected by the generalization of reality and the spatial scales that CA models rely on (Brown *et al*. 2005). In map comparison, Costanza (1989) noted that the assessment at one scale is not sufficient to delineate the similarity of complex landscape patterns. For example, the scale selection could affect the goodness-of-fit analysis, spatial autocorrelation, spatial clustering, Null/Neutral model-based cross-assessment, and so on (Hagen-Zanker and Lajoie 2008, Pontius *et al*. 2008). Multiscale analysis or the test of the scale effect is needed for the reliable assessment of CA models.

4.4 Suggestions to method selection

Our review provides a reference frame for choosing appropriate metrics for future work, and we have a few suggestions as follows.

(1) We strongly recommend a proper evaluation of the input dataset used for model calibration because data source errors and their propagation through the CA modeling can substantially affect the simulation results (Wu *et al*. 2012, Grinblat *et al*. 2016).
(2) The procedure assessment is of particularly significant to CA modeling, and modelers should pay more attention to this issue because it substantially influences the model performance hence the simulation results. However, for example, it is

challenging for assessing the transition rules derived using black-box methods such as neural networks and deep learning (Omrani *et al*. 2017, Qiao *et al*. 2017), which do not produce detailed fitting reports. When using black-box methods to build CA models, we suggest using ROC and TOC (c.f. 3.3.2 Transition probability map) to evaluate the transition probability maps.

(3) While the end-state assessment is frequently criticized in CA modeling, metrics such as OA and Kstandard should not be discarded because they can provide an overview of the overall agreement between the simulated results and the actual results. The same CA model may produce a lower end-state accuracy for a region but a higher end-state accuracy for another region just because the former region experienced fewer changes than the latter instead of the predictive ability the model itself. As such, change assessment metrics such as Hit, Miss, False alarms and Ksimulation are highly recommended because these more properly describe the simulation performance of a model for a specific region. While Ksimulation has been applied in many publications (Hagen-Zanker and Martens 2008, Van Vliet *et al*. 2011), it should be received more attention from the community of CA modeling (c.f. the subsection of Kappa coefficient).

(4) The spatial autocorrelation and landscape analysis to date have been majorly applied to the end-state assessment, but they can also be used to evaluate the modeling change. This should be able to provide more insights into the pattern-level change other than the cell-level change.

(5) The result evaluation using driving factors (Feng *et al*. 2019) is a promising method that provides a new perspective for model calibration and validation, and can be used for credibility assessment of predicted future scenarios of land-use and urban patterns.

5. Conclusions

We summarized 69 specific metrics in four aspects that have been applied in the CA model assessment during the past 20 years. Our review shows that 89% of the selected publications (347) used an assessment procedure to evaluate the model performance, but no publication applied all these metrics because some of these are highly correlated. In addition, it is not possible for a single work to comprehensively evaluate all aspects of the models and the simulation results. We found that, during the past two decades, more scientists are working on the model development and application as demonstrated by the increasing CA publications, and the methods for CA model assessment are in growing diversity. Among these, the cell-by-cell comparison and landscape analysis are the most frequently applied methods for model assessment. While some of the metrics are widely used, they are criticized by modelers and should be not as useful as a few other metrics that have been proposed in the last decade. We provided five suggestions to method selection in this study, urging to focus on the assessment of input dataset and error propagation, modeling procedure, quantitative and spatial change, and the impact of driving factors.

While existing methods and metrics have their abilities to evaluate model performance, there are some other aspects (i.e. future research directions) that need to be more fully explored. These are: (1) to develop proper metrics for assessing the accuracy of driving factors because the data source accuracy significantly affects the model performance and credibility; 2) to explore how the CA models lead to uncertainties and how these uncertainties propagate during modeling; 3) to develop new indicators to assess the applicability of simulation results; and 4) to propose suitable metrics to asses large-scale (e.g. global and continental) land-use change simulations for which existing methods such as cell-by-cell comparison and landscape analysis are not most applicable.

Acknowledgments

We thank the Editor and five anonymous reviewers for their feedback and constructive comments to help improve the original manuscript.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This study was supported by the National Natural Science Foundation of China [41631178 and 41771414] and the National Key R&D Program of China [2018YFB0505400 and 2018YFB0505000].

Notes on contributors

Xiaohua Tong is Professor at the College of Surveying and Geo-Informatics, Tongji University, Shanghai, China. He received Ph.D. degree from Tongji University in 1999. From 2001 to 2003, he was a Post-Doctoral Researcher with the State Key Laboratory of Information Engineering in Surveying, Mapping, and Remote Sensing, Wuhan University, Wuhan, China. He was a Research Fellow with Hong Kong Polytechnic University, Hong Kong, in 2006, and a Visiting Scholar with the University of California, Santa Barbara, CA, USA, from 2008 to 2009. His research interests include photogrammetry and remote sensing, trust in spatial data, and image processing for high-resolution satellite images.

Yongjiu Feng is Professor at the College of Surveying and Geo-Informatics, Tongji University, Shanghai, China and Honorary Associate Professor at the University of Queensland, Brisbane, Australia. He received Ph.D. degree from Tongji University in 2009. From 2015 to 2016, he was a Visiting Academic at the University of Queensland. His research interests include land use change modeling, cellular automata, spatial analysis and remote sensing image processing.

ORCID

Yongjiu Feng http://orcid.org/0000-0001-8772-7218

References

Aguilera, F., Valenzuela, L.M., and Botequilha-Leit, O.A., 2011. Landscape metrics in the analysis of urban land use patterns: a case study in a Spanish metropolitan area. *Landscape and Urban Planning*, 99 (3–4), 226–238. doi:10.1016/j.landurbplan.2010.10.004

Ahmed, B. and Ahmed, R., 2012. Modeling urban land cover growth dynamics using multi-temporal satellite images: a case study of Dhaka, Bangladesh. *ISPRS International Journal of Geo-Information*, 1 (1), 3–31. doi:10.3390/ijgi1010003

Ahmed, B., Ahmed, R., and Zhu, X., 2013. Evaluation of model validation techniques in land cover dynamics. *ISPRS International Journal of Geo-Information*, 2 (3), 577–597. doi:10.3390/ijgi2030577

Akaike, H., 2011. Akaike's information criterion. *In*: Lovric, Miodrag, eds. *International encyclopedia of statistical science*. Heidelberg: Springer, 25.

Alaei moghadam, S., Karimi, M., and Habibi, K., 2018. Simulating urban growth in a megalopolitan area using a patch-based cellular automata. *Transactions in GIS*, 22 (1), 249–268.

Al-Ageili, M., Mouhoub, M., and Piwowar, J., 2017. Remote sensing, GIS and cellular automata for urban growth simulation. *Computer and Information Science*, 10 (4), 38.

Al-Ahmadi, K., ALAHMADI, M., and Alahmadi, S., 2016. Spatial optimization of urban cellular automata model. In: Ming Hung *Applications of Spatial Statistics*, InTech. London

Aldwaik, S.Z., Onsted, J.A., and Pontius, R.G., JR, 2015. Behavior-based aggregation of land categories for temporal change analysis. *International Journal of Applied Earth Observation and Geoinformation*, 35, 229–238.

Aljoufie, M., *et al*. 2013. A cellular automata-based land use and transport interaction model applied to Jeddah, Saudi Arabia. *Landscape and Urban Planning*, 112, 89–99.

Al-Shalabi, M., *et al*., 2013. Modelling urban growth evolution and land-use changes using GIS based cellular automata and SLEUTH models: the case of Sana'a metropolitan city, Yemen. *Environmental Earth Sciences*, 70 (1), 425–437.

Al-Sharif, A.A. and Pradhan, B., 2015. A novel approach for predicting the spatial patterns of urban expansion by combining the chi-squared automatic integration detection decision tree, Markov chain and cellular automata models in GIS. *Geocarto International*, 30 (8), 858–881.

Anselin, L., 2019. A local indicator of multivariate spatial association: extending Geary's c. *Geographical Analysis*, 51 (2), 133–150.

Arsanjani, J.J., *et al*., 2013. Integration of logistic regression, Markov chain and cellular automata; models to simulate urban expansion. *International Journal of Applied Earth Observations & Geoinformation*, 21 (1), 265–275.

Bachmann, A. and Allgöwer, B., 2002. Uncertainty propagation in wildland fire behaviour modelling. *International Journal of Geographical Information Science*, 16 (2), 115–127.

Barredo, J.I., *et al*., 2004. Modelling future urban scenarios in developing countries: an application case study in Lagos, Nigeria. *Environment and Planning B: Planning and Design*, 31 (1), 65–84.

Barredo, J.I. and Demicheli, L., 2003. Urban sustainability in developing countries' megacities: modelling and predicting future urban growth in Lagos. *Cities*, 20 (5), 297–310.

Barreira-Gonzalez, P. and Barros, J., 2017. Configuring the neighbourhood effect in irregular cellular automata based models. *International Journal of Geographical Information Science*, 31 (3), 617–636.

Barreira-GonzaLez, P., Gomez-Delgado, M., and Aguilera-Benavente, F., 2015. From raster to vector cellular automata models: a new approach to simulate urban growth with the help of graph theory. *Computers, Environment and Urban Systems*, 54, 119–131.

Basse, R.M., *et al*. 2014. Land use changes modelling using advanced methods: cellular automata and artificial neural networks. The spatial and explicit representation of land cover dynamics at the cross-border region scale. *Applied Geography*, 53, 160–171.

Basse, R.M., Charif, O., and Bodis, K., 2016. Spatial and temporal dimensions of land use change in cross border region of Luxembourg. Development of a hybrid approach integrating GIS, cellular automata and decision learning tree models. *Applied Geography*, 67, 94–108.

Batisani, N. and Yarnal, B., 2009. Urban expansion in centre county, Pennsylvania: spatial dynamics and landscape transformations. *Applied Geography*, 29 (2), 235–249.

Blecic, I., Cecchini, A., and Trunfio, G.A., 2015. How much past to see the future: a computational study in calibrating urban cellular automata. *International Journal of Geographical Information Science*, 29 (3), 349–374.

Bozkaya, A.G., *et al*., 2015. Forecasting land-cover growth using remotely sensed data: a case study of the Igneada protection area in Turkey. *Environmental Monitoring and Assessment*, 187 (3), 59.

Bradley, A.V., *et al*. 2016. SimiVal, a multi-criteria map comparison tool for land-change model projections. *Environmental Modelling & Software*, 82, 229–240.

Brown, D.G., *et al*., 2005. Path dependence and the validation of agent-based spatial models of land use. *International Journal of Geographical Information Science*, 19 (2), 153–174.
Brown, D.G. and Xie, Y., 2006. Spatial agent-based modelling. *International Journal of Geographical Information Science*, 20 (9), 941–943.
Cao, M., *et al*., 2016. A bat-inspired approach to define transition rules for a cellular automaton model used to simulate urban expansion. *International Journal of Geographical Information Science*, 30 (10), 1961–1979.
Carter, J.G., 2018. Urban climate change adaptation: exploring the implications of future land cover scenarios. *Cities*, 77, 73–80.
Charif, O., *et al*., 2017. A multi-label cellular automata model for land change simulation. *Transactions in GIS*, 21 (6), 1298–1320.
Chaudhuri, G. and Clarke, K.C., 2013. How does land use policy modify urban growth? A case study of the Italo-Slovenian border. *Journal of Land Use Science*, 8 (4), 443–465.
Chaudhuri, G. and Clarke, K.C., 2014. Temporal accuracy in urban growth forecasting: a study using the SLEUTH model. *Transactions in GIS*, 18 (2), 302–320.
Chen, Y., *et al*., 2014. Modeling urban land-use dynamics in a fast developing city using the modified logistic cellular automaton with a patch-based simulation strategy. *International Journal of Geographical Information Science*, 28 (2), 234–255.
Chowdhury, P.R. and Maithani, S., 2014. Modelling urban growth in the Indo-Gangetic plain using nighttime OLS data and cellular automata. *International Journal of Applied Earth Observation and Geoinformation*, 33, 155–165.
Clarke, K.C., *et al*. 2018. Land use change modeling with SLEUTH: improving calibration with a genetic algorithm. *In*: M.T. Camacho Olmedo, ed. *Geomatic approaches for modeling land change scenarios*. Cham: Springer International Publishing, 139–161.
Congalton, R.G. and Green, K., 2008. *Assessing the accuracy of remotely sensed data: principles and practices*. Boca Raton: CRC press.
Costanza, R., 1989. Model goodness of fit: A multiple resolution procedure. *Ecological Modelling*, 47 (3), 199–215.
Crols, T., *et al*., High-resolution simulations of population-density change with an activity-based cellular automata land-use model. ed. *Proceedings of GeoComputation,* London *2017*, 2017.
Dahal, K.R. and Chow, T.E., 2014. An agent-integrated irregular automata model of urban land-use dynamics. *International Journal of Geographical Information Science*, 28 (11), 2281–2303.
Dahal, K.R. and Chow, T.E., 2015. Characterization of neighborhood sensitivity of an irregular cellular automata model of urban growth. *International Journal of Geographical Information Science*, 29 (3), 475–497.
De Almeida, C.M., *et al*., 2003. Stochastic cellular automata modeling of urban land use dynamics: empirical development and estimation. *Computers, Environment and Urban Systems*, 27 (5), 481–509.
De Noronha Vaz, E. and Nijkamp, P., 2015. Gravitational forces in the spatial impacts of urban sprawl. *Habitat International*, 45 (2), 99–105.
Dezhkam, S., *et al*., 2017. Performance evaluation of land change simulation models using landscape metrics. *Geocarto International*, 32 (6), 655–677.
Dietzel, C. and Clarke, K.C., 2004. Spatial differences in multi-resolution urban automata modeling. *Transactions in GIS*, 8 (4), 479–492.
Dietzel, C. and Clarke, K.C., 2007. Toward optimal calibration of the SLEUTH land use change model. *Transactions in GIS*, 11 (1), 29–45.
Du, G., *et al*., 2018. A comparative approach to modelling multiple urban land use changes using tree-based methods and cellular automata: the case of greater Tokyo Area. *International Journal of Geographical Information Science*, 32 (4), 757–782.
Fan, C. and Myint, S., 2014. A comparison of spatial autocorrelation indices and landscape metrics in measuring urban landscape fragmentation. *Landscape and Urban Planning*, 121, 117–128.
Feng, Y., *et al*., 2011. Modeling dynamic urban growth using cellular automata and particle swarm optimization rules. *Landscape and Urban Planning*, 102 (3), 188–196.

Feng, Y., *et al.*, 2016. Simulation of dynamic urban growth with partial least squares regression-based cellular automata in a GIS environment. *ISPRS International Journal of Geo-Information*, 5 (12), 243.
Feng, Y., 2017. Modeling dynamic urban land-use change with geographical cellular automata and generalized pattern search-optimized rules. *International Journal of Geographical Information Science*, 31 (6), 1198–1219.
Feng, Y., *et al.*, 2018a. The effect of observation scale on urban growth simulation using particle swarm optimization-based CA models. *Sustainability*, 10 (11), 4002.
Feng, Y., *et al.*, 2018b. Modelling coastal land use change by incorporating spatial autocorrelation into cellular automata models. *Geocarto International*, 33 (5), 470–488.
Feng, Y., *et al*. 2019. How much can temporally stationary factors explain cellular automata-based simulations of past and future urban growth? *Computers, Environment and Urban Systems*, 76, 150–162.
Feng, Y. and Tong, X., 2017. Using exploratory regression to identify optimal driving factors for cellular automaton modeling of land use change. *Environmental Monitoring and Assessment*, 189 (10), 515.
Feng, Y. and Tong, X., 2019. A new cellular automata framework of urban growth modeling by incorporating statistical and heuristic methods. *International Journal of Geographical Information Science*, 1–24. https://doi.org/10.1080/13658816.2019.1648813
Foody, G.M., 2002. Status of land cover classification accuracy assessment. *Remote Sensing of Environment*, 80 (1), 185–201.
García, A.M., *et al*., 2012. A comparative analysis of cellular automata models for simulation of small urban areas in Galicia, NW Spain. *Computers, Environment and Urban Systems*, 36 (4), 291–301.
Getis, A., 2010. Spatial autocorrelation. *In*: M.M. Fischer and A. GETIS, eds. *Handbook of applied spatial analysis: software tools, methods and applications*. Berlin, Heidelberg: Springer Berlin Heidelberg, 255–278.
Getis, A. and Aldstadt, J., 2010. Constructing the spatial weights matrix using a local statistic. *In*: Luc Anselin; Sergio J. Rey, *Perspectives on spatial data analysis*. New York: Springer, 147–163.
Gharbia, S.S., *et al*., 2016. Land use scenarios and projections simulation using an integrated GIS cellular automata algorithms. *Modeling Earth Systems and Environment*, 2 (3), 151.
Goodarzi, M.S., Sakieh, Y., and Navardi, S., 2017. Scenario-based urban growth allocation in a rapidly developing area: a modeling approach for sustainability analysis of an urban-coastal coupled system. *Environment, Development and Sustainability*, 19 (3), 1103–1126.
Gounaridis, D., Chorianopoulos, I., and Koukoulas, S., 2018. Exploring prospective urban growth trends under different economic outlooks and land-use planning scenarios: the case of Athens. *Applied Geography*, 90, 134–144.
Grekousis, G., Manetos, P., and Photis, Y.N., 2013. Modeling urban evolution using neural networks, fuzzy logic and GIS: the case of the Athens metropolitan area. *Cities*, 30 (1), 193–203.
Grinblat, Y., Gilichinsky, M., and Benenson, I., 2016. Cellular automata modeling of land-use/land-cover dynamics: questioning the reliability of data sources and classification methods. *Annals of the American Association of Geographers*, 106 (6), 1299–1320.
Guan, Q., *et al*., 2016. A hybrid parallel cellular automata model for urban growth simulation over GPU/CPU heterogeneous architectures. *International Journal of Geographical Information Science*, 30 (3), 494–514.
Guan, Q., Wang, L., and CLARKE, K.C., 2005. An artificial-neural-network-based, constrained CA model for simulating urban growth. *Cartography and Geographic Information Science*, 32 (4), 369–380.
Gwet, K.L., 2014. *Handbook of inter-rater reliability: the definitive guide to measuring the extent of agreement among raters*. Piedmont: Advanced Analytics, LLC.
Hagen-Zanker, A., Multi-method assessment of map similarity. ed. *Proceedings of the fifth AGILE conference on geographic information science*, 2002, Palma, Spain, 171–182.
Hagen-Zanker, A., 2003. Fuzzy set approach to assessing similarity of categorical maps. *International Journal of Geographical Information Science*, 17 (3), 235–249.

Hagen-Zanker, A., 2006. Map comparison methods that simultaneously address overlap and structure. *Journal of Geographical Systems*, 8 (2), 165–185.
Hagen-Zanker, A., 2009. An improved fuzzy Kappa statistic that accounts for spatial autocorrelation. *International Journal of Geographical Information Science*, 23 (1), 61–73.
Hagen-Zanker, A., 2016. A computational framework for generalized moving windows and its application to landscape pattern analysis. *International Journal of Applied Earth Observation and Geoinformation*, 44, 205–216.
Hagen-Zanker, A. and Lajoie, G., 2008. Neutral models of landscape change as benchmarks in the assessment of model performance. *Landscape and Urban Planning*, 86 (3), 284–296.
Hagen-Zanker, A. and Martens, P., 2008. Map comparison methods for comprehensive assessment of geosimulation models. *In*: O. Gervasi, *et al*., ed.. Computational Science and Its Applications: Berlin, Heidelberg. 194–209.
He, J., *et al*., 2018. Mining transition rules of cellular automata for simulating urban expansion by using the deep learning techniques. *International Journal of Geographical Information Science*, 32 (10), 2076–2097.
Herold, M., Couclelis, H., and Clarke, K.C., 2005. The role of spatial metrics in the analysis and modeling of urban land use change. *Computers, Environment and Urban Systems*, 29 (4), 369–399.
Hewitt, R. and Diaz-Pacheco, J., 2017. Stable models for metastable systems? Lessons from sensitivity analysis of a cellular automata urban land use model. *Computers, Environment and Urban Systems*, 62, 113–124.
Hyandye, C. and Martz, L.W., 2017. A Markovian and cellular automata land-use change predictive model of the Usangu catchment. *International Journal of Remote Sensing*, 38 (1), 64–81.
Jafari, M., *et al*., 2016. Dynamic simulation of urban expansion based on cellular automata and logistic regression model: case study of the Hyrcanian region of Iran. *Sustainability*, 8 (8), 810.
Jantz, C.A. and Goetz, S.J., 2005. Analysis of scale dependencies in an urban land-use-change model. *International Journal of Geographical Information Science*, 19 (2), 217–241.
Jantz, C.A., Goetz, S.J., and Shelley, M.K., 2004. Using the Sleuth urban growth model to simulate the impacts of future policy scenarios on urban land use in the Baltimore-Washington metropolitan area. *Environment and Planning B: Planning and Design*, 31 (2), 251–271.
Kamusoko, C. and Gamba, J., 2015. Simulating urban growth using a random forest-cellular automata (RF-CA) model. *ISPRS International Journal of Geo-Information*, 4 (2), 447–470.
Ke, X., Qi, L., and Zeng, C., 2016. A partitioned and asynchronous cellular automata model for urban growth simulation. *International Journal of Geographical Information Science*, 30 (4), 637–659.
Kocabas, V. and Dragicevic, S., 2006. Assessing cellular automata model behaviour using a sensitivity analysis approach. *Computers, Environment and Urban Systems*, 30 (6), 921–953.
Kocabas, V. and Dragicevic, S., 2007. Enhancing a GIS cellular automata model of land use change: Bayesian networks, influence diagrams and causality. *Transactions in GIS*, 11 (5), 681–702.
Ku, C.-A., 2016. Incorporating spatial regression model into cellular automata for simulating land use change. *Applied Geography*, 69, 1–9.
Lauf, S., *et al*. 2012. Uncovering land-use dynamics driven by human decision-making–A combined model approach using cellular automata and system dynamics. *Environmental Modelling & Software*, 27, 71–82.
Li, H., *et al*., 2018. Simulating urban cooperative expansion in a single-core metropolitan region based on improved CA model integrated information flow: case study of Wuhan urban agglomeration in China. *Journal of Urban Planning and Development*, 144 (2), 05018002.
Li, X., *et al*., 2012. Assimilating process context information of cellular automata into change detection for monitoring land use changes. *International Journal of Geographical Information Science*, 26 (9), 1667–1687.
Li, X., Liu, X., and Gong, P., 2015. Integrating ensemble-urban cellular automata model with an uncertainty map to improve the performance of a single model. *International Journal of Geographical Information Science*, 29 (5), 762–785.
Li, X., Yang, Q., and Liu, X., 2008. Discovering and evaluating urban signatures for simulating compact development using cellular automata. *Landscape and Urban Planning*, 86 (2), 177–186.

Liao, J., *et al*., 2014. A neighbor decay cellular automata approach for simulating urban expansion based on particle swarm intelligence. *International Journal of Geographical Information Science*, 28 (4), 720–738.
Lin, Y.-P., *et al*., 2011. Predictive ability of logistic regression, auto-logistic regression and neural network models in empirical land-use change modeling–a case study. *International Journal of Geographical Information Science*, 25 (1), 65–87.
Liu, D., *et al*. 2018. Interoperable scenario simulation of land-use policy for Beijing–Tianjin–Hebei region, China. *Land Use Policy*, 75, 155–165.
Liu, X., *et al*., 2008. Simulating complex urban development using kernel-based non-linear cellular automata. *Ecological Modelling*, 211 (1–2), 169–181.
Liu, X., *et al*., 2010. Simulating land-use dynamics under planning policies by integrating artificial immune systems with cellular automata. *International Journal of Geographical Information Science*, 24 (5), 783–802.
Liu, X., *et al*. 2017. A future land use simulation model (FLUS) for simulating multiple land use scenarios by coupling human and natural effects. *Landscape and Urban Planning*, 168, 94–116.
Liu, Y. and Feng, Y., 2012. A logistic based cellular automata model for continuous urban growth simulation: a case study of the Gold coast city, Australia. *In*: Alison J. Heppenstall, Andrew T. Crooks, Linda M. See and Michael Batty, *Agent-based models of geographical systems*. Springer, 643–662.
Liu, Y. and Feng, Y., 2016. Simulating the impact of economic and environmental strategies on future urban growth scenarios in Ningbo, China. *Sustainability*, 8 (10), 1045.
Mantelas, L., *et al*., 2012. Using fuzzy cellular automata to access and simulate urban growth. *GeoJournal*, 77 (1), 13–28.
Mcgarigal, K. 2014. *Fragstats v4: spatial pattern analysis program for categorical and continuous maps-help manual*. Amherst: University of Massachusetts. Available fr http://www.umass.edu/landeco/research/fragstats/fragstats.html
McGarigal, K., *et al*. 2018. Modeling non-stationary urban growth: the SPRAWL model and the ecological impacts of development. *Landscape and Urban Planning*, 177, 178–190.
Meentemeyer, R.K., *et al*., 2013. FUTURES: multilevel simulations of emerging urban–rural landscape structure using a stochastic patch-growing algorithm. *Annals of the Association of American Geographers*, 103 (4), 785–807.
Ménard, A. and Marceau, D.J., 2007. Simulating the impact of forest management scenarios in an agricultural landscape of southern Quebec, Canada, using a geographic cellular automata. *Landscape and Urban Planning*, 79 (3), 253–265.
Mitsova, D., Shuster, W., and Wang, X., 2011. A cellular automata model of land cover change to integrate urban growth with open space conservation. *Landscape and Urban Planning*, 99 (2), 141–153.
Mondal, B., Das, D.N., and Bhatta, B., 2017. Integrating cellular automata and Markov techniques to generate urban development potential surface: a study on Kolkata agglomeration. *Geocarto International*, 32 (4), 401–419.
Moreno, N., Ménard, A., and Marceau, D.J., 2008. VecGCA: a vector-based geographic cellular automata model allowing geometric transformations of objects. *Environment and Planning B: Planning and Design*, 35 (4), 647–665.
Moreno, N., Wang, F., and Marceau, D.J., 2009. Implementation of a dynamic neighborhood in a land-use vector-based cellular automata model. *Computers, Environment and Urban Systems*, 33 (1), 44–54.
Mundia, C.N. and Murayama, Y., 2010. Modeling spatial processes of urban growth in African cities: a case study of Nairobi city. *Urban Geography*, 31 (2), 259–272.
Musa, S.I., Hashim, M., and Reba, M.N.M., 2017. A review of geospatial-based urban growth models and modelling initiatives. *Geocarto International*, 32 (8), 813–833.
Mustafa, A., *et al*., 2018a. Comparing support vector machines with logistic regression for calibrating cellular automata land use change models. *European Journal of Remote Sensing*, 51 (1), 391–401.
Mustafa, A., *et al*., 2018b. A time Monte Carlo method for addressing uncertainty in land-use change models. *International Journal of Geographical Information Science*, 32 (11), 2317–2333.

Myint, S.W. and Wang, L., 2006. Multicriteria decision approach for land use land cover change using Markov chain analysis and a cellular automata approach. *Canadian Journal of Remote Sensing*, 32 (6), 390–404.
Newland, C.P., *et al*. 2018a. Multi-objective optimisation framework for calibration of cellular automata land-use models. *Environmental Modelling & Software*, 100, 175–200.
Newland, C.P., *et al*. 2018b. Empirically derived method and software for semi-automatic calibration of cellular automata land-use models. *Environmental Modelling & Software*, 108, 208–239.
Omrani, H., Tayyebi, A., and Pijanowski, B., 2017. Integrating the multi-label land-use concept and cellular automata with the artificial neural network-based land transformation model: an integrated ML-CA-LTM modeling framework. *GIScience & Remote Sensing*, 54 (3), 283–304.
Onsted, J. and Clarke, K.C., 2012. The inclusion of differentially assessed lands in urban growth model calibration: a comparison of two approaches using SLEUTH. *International Journal of Geographical Information Science*, 26 (5), 881–898.
Oreskes, N., 1998. Evaluation (not validation) of quantitative models. *Environmental Health Perspectives*, 106 (suppl 6), 1453–1460.
Oreskes, N., Shrader-Frechette, K., and Belitz, K., 1994. Verification, validation, and confirmation of numerical models in the earth sciences. *Science*, 263 (5147), 641–646.
Padmanaban, R., *et al*., 2017. Modelling urban sprawl using remotely sensed data: a case study of Chennai city, Tamilnadu. *Entropy*, 19 (4), 163.
Peng, J., *et al*., 2010. Evaluating the effectiveness of landscape metrics in quantifying spatial patterns. *Ecological Indicators*, 10 (2), 217–223.
Pérez-Molina, E., *et al*. 2017. Developing a cellular automata model of urban growth to inform spatial policy for flood mitigation: a case study in Kampala, Uganda. *Computers, Environment and Urban Systems*, 65, 53–65.
Petrov, L.O., Lavalle, C., and Kasanko, M., 2009. Urban land use scenarios for a tourist region in Europe: applying the Moland model to Algarve, Portugal. *Landscape and Urban Planning*, 92 (1), 10–23.
Pijanowski, B.C., *et al*., 2002. Using neural networks and GIS to forecast land use changes: a land transformation model. *Computers, Environment and Urban Systems*, 26 (6), 553–575.
Pinto, N., Antunes, A.P., and Roca, J., 2017. Applicability and calibration of an irregular cellular automata model for land use change. *Computers, Environment and Urban Systems*, 65, 93–102.
Poelmans, L. and Van Rompaey, A., 2009. Detecting and modelling spatial patterns of urban sprawl in highly fragmented areas: a case study in the Flanders–Brussels region. *Landscape and Urban Planning*, 93 (1), 10–19.
Pontius, R.G., 2000. Quantification error versus location error in comparison of categorical maps. *Photogrammetric Engineering and Remote Sensing*, 66 (8), 1011–1016.
Pontius, R.G., *et al*., 2008. Comparing the input, output, and validation maps for several models of land change. *The Annals of Regional Science*, 42 (1), 11–37.
Pontius, R.G. and Malanson, J., 2005. Comparison of the structure and accuracy of two land change models. *International Journal of Geographical Information Science*, 19 (2), 243–265.
Pontius, R.G. and Malizia, N.R., 2004. Effect of category aggregation on map comparison. *In*: M. J. Egenhofer, C. Freksa, and H.J. Miller, eds. *Geographic Information Science*. Berlin, Heidelberg: Geographic Information Science, 251–268.
Pontius, R.G. and Millones, M., 2011. Death to Kappa: birth of quantity disagreement and allocation disagreement for accuracy assessment. *International Journal of Remote Sensing*, 32 (15), 4407–4429.
Pontius, R.G., Peethambaram, S., and Castella, J.-C., 2011. Comparison of three maps at multiple resolutions: a case study of land change simulation in Cho Don district, Vietnam. *Annals of the Association of American Geographers*, 101 (1), 45–62.
Pontius, R.G. and Si, K., 2014. The total operating characteristic to measure diagnostic ability for multiple thresholds. *International Journal of Geographical Information Science*, 28 (3), 570–583.
Puertas, O.L., Henriquez, C., and Meza, F.J., 2014. Assessing spatial dynamics of urban growth using an integrated land use model. Application in Santiago metropolitan area, 2010–2045. *Land Use Policy*, 38, 415–425.

Qiao, W., *et al*., 2017. Evaluation of intensive urban land use based on an artificial neural network model: a case study of Nanjing city, China. *Chinese Geographical Science*, 27 (5), 735–746.
Rabbani, A., Aghababaee, H., and Rajabi, M.A., 2012. Modeling dynamic urban growth using hybrid cellular automata and particle swarm optimization. *Journal of Applied Remote Sensing*, 6 (1), 063582.
Rienow, A. and Goetzke, R., 2015. Supporting SLEUTH–enhancing a cellular automaton with support vector machines for urban growth modeling. *Computers, Environment and Urban Systems*, 49, 66–81.
Rimal, B., *et al*., 2018. Land use/land cover dynamics and modeling of urban land expansion by the integration of cellular automata and Markov Chain. *ISPRS International Journal of Geo-Information*, 7 (4), 154.
Rykiel, E.J., 1996. Testing ecological models: the meaning of validation. *Ecological Modelling*, 90 (3), 229–244.
Sakieh, Y., *et al*., 2015a. Scenario-based evaluation of urban development sustainability: an integrative modeling approach to compromise between urbanization suitability index and landscape pattern. *Environment, Development and Sustainability*, 17 (6), 1343–1365.
Sakieh, Y., *et al*., 2015b. Simulating urban expansion and scenario prediction using a cellular automata urban growth model, SLEUTH, through a case study of Karaj city, Iran. *Journal of Housing and the Built Environment*, 30 (4), 591–611.
Sakieh, Y. and Salmanmahiny, A., 2016. Performance assessment of geospatial simulation models of land-use change—a landscape metric-based approach. *Environmental Monitoring and Assessment*, 188 (3), 169.
Shafizadeh-Moghadam, H., *et al*. 2017. Coupling machine learning, tree-based and statistical models with cellular automata to simulate urban growth. *Computers, Environment and Urban Systems*, 64, 297–308.
Soares-Filho, B., Rodrigues, H., and Follador, M., 2013. A hybrid analytical-heuristic method for calibrating land-use change models. *Environmental Modelling & Software*, 43, 80–87.
Soares-Filho, B.S., Cerqueira, G.C., and Pennachin, C.L., 2002. DINAMICA—a stochastic cellular automata model designed to simulate the landscape dynamics in an Amazonian colonization frontier. *Ecological Modelling*, 154 (3), 217–235.
Stevens, D. and Dragićević, S., 2007. A GIS-based irregular cellular automata model of land-use change. *Environment and Planning B: Planning and Design*, 34 (4), 708–724.
Stevens, D., Dragicevic, S., and Rothley, K., 2007. iCity: A GIS–CA modelling tool for urban planning and decision making. *Environmental Modelling & Software*, 22 (6), 761–773.
Straatman, B., White, R., and Engelen, G., 2004. Towards an automatic calibration procedure for constrained cellular automata. *Computers, Environment and Urban Systems*, 28 (1), 149–170.
Tayyebi, A., Perry, P.C., and Tayyebi, A.H., 2014a. Predicting the expansion of an urban boundary using spatial logistic regression and hybrid raster–vector routines with remote sensing and GIS. *International Journal of Geographical Information Science*, 28 (4), 639–659.
Tayyebi, A. and Pijanowski, B.C., 2014. Modeling multiple land use changes using ANN, CART and MARS: comparing tradeoffs in goodness of fit and explanatory power of data mining tools. *International Journal of Applied Earth Observation and Geoinformation*, 28, 102–116.
Tayyebi, A.H., Tayyebi, A., and Khanna, N., 2014b. Assessing uncertainty dimensions in land-use change models: using swap and multiplicative error models for injecting attribute and positional errors in spatial data. *International Journal of Remote Sensing*, 35 (1), 149–170.
Thapa, R.B. and Murayama, Y., 2011. Urban growth modeling of Kathmandu metropolitan region, Nepal. *Computers, Environment and Urban Systems*, 35 (1), 25–34.
Thapa, R.B. and Murayama, Y., 2012. Scenario based urban growth allocation in Kathmandu valley, Nepal. *Landscape and Urban Planning*, 105 (1–2), 140–148.
Van Vliet, J., *et al*. 2013a. A fuzzy set approach to assess the predictive accuracy of land use simulations. *Ecological Modelling*, 261–262, 32–42.
Van Vliet, J., *et al*. 2013b. Measuring the neighbourhood effect to calibrate land use models. *Computers, Environment and Urban Systems*, 41, 55–64.
Van Vliet, J., *et al*. 2016. A review of current calibration and validation practices in land-change modeling. *Environmental Modelling & Software*, 82, 174–182.
Van Vliet, J., Bregt, A.K., and Hagen-Zanker, A., 2011. Revisiting Kappa to account for change in the accuracy assessment of land-use change models. *Ecological Modelling*, 222 (8), 1367–1375.

Van Vliet, J., White, R., and Dragicevic, S., 2009. Modeling urban growth using a variable grid cellular automaton. *Computers, Environment and Urban Systems*, 33 (1), 35–43.
Verburg, P., *et al*., 1999. A spatial explicit allocation procedure for modelling the pattern of land use change based upon actual land use. *Ecological Modelling*, 116 (1), 45–61.
Verburg, P.H., *et al*., 2002. Modeling the spatial dynamics of regional land use: the CLUE-S model. *Environmental Management*, 30 (3), 391–405.
Verburg, P.H., *et al*., 2004. A method to analyse neighbourhood characteristics of land use patterns. *Computers, Environment and Urban Systems*, 28 (6), 667–690.
Verburg, P.H. and Overmars, K.P., 2009. Combining top-down and bottom-up dynamics in land use modeling: exploring the future of abandoned farmlands in Europe with the Dyna-CLUE model. *Landscape Ecology*, 24 (9), 1167.
Verburg, P.H. and Veldkamp, A., 2001. The role of spatially explicit models in land-use change research: a case study for cropping patterns in China. *Agriculture, Ecosystems & Environment*, 85 (1–3), 177–190.
Vermeiren, K., *et al*., 2016. ASSURE: a model for the simulation of urban expansion and intra-urban social segregation. *International Journal of Geographical Information Science*, 30 (12), 2377–2400.
Waddell, P., 2002. UrbanSim: modeling urban development for land use, transportation, and environmental planning. *Journal of the American Planning Association*, 68 (3), 297–314.
Wang, F. and Marceau, D.J., 2013. A patch-based cellular automaton for simulating land-use changes at fine spatial resolution. *Transactions in GIS*, 17 (6), 828–846.
Wang, H., *et al*. 2013. Simulating urban expansion using a cloud-based cellular automata model: A case study of Jiangxia, Wuhan, China. *Landscape and Urban Planning*, 110, 99–112.
Wang, R., Derdouri, A., and Murayama, Y., 2018. Spatiotemporal simulation of future land use/cover change scenarios in the Tokyo metropolitan area. *Sustainability*, 10 (6), 2056.
Wang, R. and Murayama, Y., 2017. Change of land use/cover in Tianjin city based on the Markov and cellular automata models. *ISPRS International Journal of Geo-Information*, 6 (5), 150.
Wang, Y. and Li, S., 2011. Simulating multiple class urban land-use/cover changes by RBFN-based CA model. *Computers & Geosciences*, 37 (2), 111–121.
Whitsed, R. and Smallbone, L.T., 2017. A hybrid genetic algorithm with local optimiser improves calibration of a vegetation change cellular automata model. *International Journal of Geographical Information Science*, 31 (4), 717–737.
Wickramasuriya, R.C., *et al*., 2009. The dynamics of shifting cultivation captured in an extended constrained cellular automata land use model. *Ecological Modelling*, 220 (18), 2302–2309.
Williams, B., Petrov, L., and Ustaoglu, E., 2017. Scenario analysis of alternative land development patterns for the Leipzig-Halle region: implications for transport-land-use sustainability. *Urban Planning*, 2 (1), 108–129.
Wu, F., 2002. Calibration of stochastic cellular automata: the application to rural-urban land conversions. *International Journal of Geographical Information Science*, 16 (8), 795–818.
Wu, H., *et al*., 2012. Quantifying and analyzing neighborhood configuration characteristics to cellular automata for land use simulation considering data source error. *Earth Science Informatics*, 5 (2), 77–86.
Wu, H., *et al*., 2019. Examining the sensitivity of spatial scale in cellular automata Markov chain simulation of land use change. *International Journal of Geographical Information Science*, 33 (5), 1040–1061.
Xia, C., *et al*., 2018. A high-performance cellular automata model for urban simulation based on vectorization and parallel computing technology. *International Journal of Geographical Information Science*, 32 (2), 399–424.
Xia, C., *et al*., 2019. Modeling urban growth in a metropolitan area based on bidirectional flows, an improved gravitational field model, and partitioned cellular automata. *International Journal of Geographical Information Science*, 33 (5), 877–899.

Xie, W., *et al*. 2018. Projecting the impacts of urban expansion on simultaneous losses of ecosystem services: A case study in Beijing, China. *Ecological Indicators*, 84, 183–193.
Xu, H. and Brown, D.G., 2017. Sensitivity of a stochastic land-cover change model to pixel versus polygonal land units. *International Journal of Geographical Information Science*, 31 (4), 738–762.
Yang, X., Zheng, X.-Q., and Lv, L.-N., 2012. A spatiotemporal model of land use change based on ant colony optimization, Markov chain and cellular automata. *Ecological Modelling*, 233, 11–19.
Yeh, A.G.-O. and Li, X., 2006. Errors and uncertainties in urban cellular automata. *Computers, Environment and Urban Systems*, 30 (1), 10–28.
Yin, H., *et al*. 2018. Exploring zoning scenario impacts upon urban growth simulations using a dynamic spatial model. *Cities*, 81, 214–229.
Zhang, Q., *et al*., 2011. Simulation and analysis of urban growth scenarios for the greater Shanghai area, China. *Computers, Environment and Urban Systems*, 35 (2), 126–139.
Zhou, D., Lin, Z., and Liu, L., 2012. Regional land salinization assessment and simulation through cellular automaton-Markov modeling and spatial pattern analysis. *Science of the Total Environment*, 439, 260–274.

Appendix. Definition of acronyms

AA: Allocation agreement
ACIC: the actual change for a land category predicted as an incorrect land category
AD: Allocation disagreement
AIC: Akaike information criterion
AUC: Area under the ROC curve
CA: Cellular automata
CC: Cross-correlation
CE: Commission error
CHA: chance agreement
CI: Compactness index
CPC: Correctly simulated target cells
ED: Explained deviance
FLUS: Future land-use simulation
FOM: Figure of merit
GAM: Generalized additive model
Kfuzzy: fuzzy Kappa
Khisto: Kappa for histogram
Klocation: Kappa for location
Kno: Kappa for no ability
Kquantity: Kappa for quantity
Kstandard: Standard Kappa
LSR: Least squares regression
MA: Medium allocation
MAE: Mean absolute error
MQ: Medium quantity
NA: No allocation
NQ: No quantity
OA: Overall accuracy
OE: omission error
OQM: Other quantitative measures
OSM: Optimum SLEUTH metric
PA: Perfect allocation
PAC: Producer's accuracy
PCM: Percent correct match
PQ: Perfect quantity

QA: quantity agreement
QD: Quantity disagreement
REI: Relative error index
RMSE: Root-mean-square error
ROC: Relative operating characteristic
SE: Standard error
SEM: Spatially explicit model
STAN: Simulated target but actual non-target cells
TE: Total error
TOC: Total operating characteristic
UA: User's accuracy