A spatial zoning approach to calibrate and validate urban growth models

Ali Kazemzadeh-Zow, Saeed Zanganeh Shahraki, Luca Salvati, and Najmeh Neisani Samani

A spatial zoning approach to calibrate and validate urban growth models have been largely developed using internal variables. Further investigation is required to improve model’s calibration and validation mixing internal and external variables. To reach this objective, a spatial zoning approach simulating long-term expansion of Mashhad, the second largest city of Iran, was presented in this study. Spatial zoning approaches distinguish local-scale urban dynamics in districts with different socioeconomic characteristics. Thiessen polygons were used to identify districts with different morphology and functional attributes. Urban growth was subsequently simulated for each district using a Multi-Layer Perceptron (MLP) neural network and Markov chains (MC) analysis. MLP and MC algorithms were respectively used to derive transition maps from non-urban to urban use of land and to determine spatial evolution of built-up areas at the metropolitan scale. Results of simulations based on spatial zoning were compared with outcomes of traditional urban growth models. Spatial zoning improved significantly model’s accuracy in respect to more traditional simulation modes. The approach proposed here is appropriate when simulating land-use changes under discontinuous urban expansion.

1. Introduction

To ensure urban sustainability, decision makers and planners need precise information on recent urbanization trends, especially around large cities and metropolitan regions (Jiang and Yao 2010). Recent studies (Hu and Lo 2007, Huang et al. 2009, Dubovyk et al. 2011) emphasize the importance of spatiotemporal analysis of urban expansion and the relevance of integrated approaches monitoring changes in peri-urban areas, with special regard to land take and soil sealing (Poelmans and Van-Rompaey 2009). Understanding patterns and processes of land-use change is a key issue when implementing dynamic models of urban growth (Jokar et al. 2013). Future scenarios of urban growth are
important for planners, policy makers and stakeholders interested in land-use change (Lambin and Geist 2006).

Landscape transformations involve complex processes that are influenced by dynamic, nonlinear human–nature interactions, which are difficult to be monitored through standard approaches (Vega et al. 2012, Kolb et al. 2013). Various approaches have been implemented to simulate land-use change and predict urban expansion in different socioeconomic contexts (Cao et al. 2014, Mozumder and Tripathi 2014, Olmedo et al. 2015). Urban growth models are usually classified in three categories: (i) models implementing Markov chain (MC) or general regression approaches (Bell 1974, Clark 1965); (ii) dynamic models implementing Cellular Automata (CA), agent-based models, weights of evidence, artificial neural networks (ANN), System Dynamic (SD) models and colony optimization approaches (ACO); (iii) integrated and mixed approaches (Matthews et al. 2007, Shen et al. 2007, Rafiee et al. 2009, Thapa and Murayama 2011). To improve model’s calibration and validation, methodologies integrating different algorithms have been suggested (e.g. Liu and Li 2007, Yang and Li 2007, Qiu and Chen 2008). Hybrid models have been recently proposed considering together empirical, statistical and dynamic approaches, such as ANN and MC (Thapa and Murayama 2012), CA and linear regression (Jokar et al. 2013), MC and linear regression (Geneletti 2012), CA and fuzzy logic (Wu 1998), conversion of land-use and its effect model (CLUE) and SD (Luo et al. 2010).

On the one hand, CA has demonstrated to have potential in simulating the spatio-temporal characteristics of complex systems (Li and Yeh 2002; Clancy et al. 2010). On the other hand, ANNs investigate nonlinear relationships and multiple (changing) thresholds and can be conceived as complex mathematical functions that convert input data to a desired output (Yang et al. 2008, Pijanowski et al. 2014, Eastman 2015). Although ANNs may answer questions such as ‘which pixels are changing’ based on land transformation matrices, they are less effective in simulating intensity and direction of change. Therefore, land-use changes have been better estimated using empirical and dynamic models such as MC or SD (Shen et al. 2007, Luo et al. 2010). In this sense, MC analysis is a convenient tool for modeling land-use when landscape modifications are latent or too difficult to be assessed with a traditional change detection analysis (Benito et al. 2010). MC analysis serves as a method assessing direction and magnitude of change thanks to the intrinsic capability of prediction of both linear and nonlinear trends in a multiple set of variables, such as landscape metrics.

Although integrated models may accurately simulate land-use changes, they often consider relevant variables as space-invariant. In fact, urban growth patterns vary largely in a given landscape according to the spatial variation of certain drivers of change. In this sense, our study introduces a methodology based on spatial zoning with the aim of improving calibration and validation of urban growth models. Spatial zoning operates partitioning a given space into homogeneous units according to specific criteria (Moreno-Regidor et al. 2012). The idea underlying the notion of spatial zoning is that each district is homogeneous for urban patterns, landscape composition and rate of change in the use of land along a defined time interval. These place-specific attributes influence model’s capability in providing accurate scenarios of future urbanization. Appropriate discrimination of urban zones based on homogeneous criteria contributes to stable and reliable urban models. Various techniques were applied to zoning natural
or urban areas including automated zoning procedure (Williams 2002), mathematical programming models (Shirabe 2005), graph partitioning models (Assunção et al. 2006, Tavares-Pereira et al. 2007), cluster analysis (Ochoa et al. 2009), artificial intelligence such as genetic algorithm (Demetriou et al. 2013), artificial immune systems (Liu et al. 2011), ant colony optimization (Li et al. 2011), simulated annealing (Santé et al. 2016), multi-objective hybrid metaheuristic approaches (Wei and Chai 2004) and spatial multi-criteria analysis (Habtemariam and Fang 2016). Many of the mentioned models are efficient only when applied to problems with a low number of zones (M) and/or basic spatial units (N), or with a set of less restrictive spatial conditions, as in cases where the centers of a given regions are predetermined (Moreno-Regidor et al. 2012). Thiessen (Voronoi) polygon is a geometric method that has been largely used to partitioning geographical space into different zones thanks to the joint use of spatial information and computational geometry: Casaer et al. (1999) applied this method with the aim to estimate wildlife space-use patterns; Morillo et al. (2016) used Voronoi polygons to zoning Madrid’s districts in order to estimate housing prices; Moreno-Regidor et al. (2012) designed agricultural districts using adaptive additively weighted Voronoi diagrams. This method is particularly effective in partitioning a heterogeneous space into homogeneous urban zones or districts.

Zoning techniques have been marginally considered in land change modeling. To investigate urban growth at the county level in the United States, Jantz et al. (2010) have implemented a regional model of land-use change using the SLEUTH CA approach. Tayyebi et al. (2011) have modeled expansion of Tehran, Iran, partitioning the study area in three regions and developing separate models of growth.

Our study proposes an original framework integrating a spatial zoning approach (Thiessen polygons) into a growth model based on MC and ANN: land-use changes were analyzed implementing a transition matrix based on a MC approach; transition rules were based on a Multi-Layer Perceptron Neural Network (MLP-NN). This approach was used to simulate long-term expansion of Mashhad, the second largest city in Iran. Mashhad is a representative example of compact cities expanding under intense demographic pressure until early 1990s, and then shifting to a progressively dispersed morphology. The paper is organized as follows: Section 2 gives a brief description of the study area and the applied methodology, introducing the proposed model. Section 3 illustrates model’s outcomes for Mashhad expansion providing a scenario for future land-use changes covering 12 years (2013–2025). Section 4 offers a comparative discussion of the results gathered in this study and provides conclusions and indications for future research.

2. Materials and methods

2.1. Study area

To test the model proposed here, medium-term (2013–2025) urban growth has been simulated for Mashhad, the second largest city in Iran, situated in the northeast part of the country (Figure 1). Resident population in Mashhad amounted to 2.4 million inhabitants in 2006 and increased to 2.7 million inhabitants in 2011. Internal migration alimented urban growth producing uneven population gaps with other big cities like
Isfahan, Tabriz and Shiraz. Mashhad is a tourism-specialized city with nearly 27 million visitors every year, 90 percent domestic. More recently, about 0.7 million refugees from Afghanistan settled in Mashhad creating new neighborhood units.

2.2. Data and input variables

Land change modeling used in this study was illustrated in Figure 2. To investigate land-use changes in Mashhad, three remote-sensed images (28 April 1986, 19 April 1999 and 22 April 2013) derived from Thematic Mapper (Landsat 5) and OLI/TIRS (Landsat 8) sensors were considered. The maximum likelihood supervised classification algorithm has been used to extract and classify seven land-use classes in the area: (i) urban land-use (buildings and streets), (ii) vegetation (garden and agriculture lands), (iii) urban green space, (iv) barren land, (v) mountainous and rocky land, (vi) water surfaces and (vii) sedimentary surfaces. ‘Overall accuracy’ and ‘Kappa coefficient’ were used to evaluate classification accuracy. Kappa coefficient is a measure of agreement between model simulation and reality; it determines if the values contained in an error matrix may represent a spatial pattern significantly better than a random matrix (Wang et al. 2012). Results were compared with sample areas extracted from historical land-use maps dated 1986 and 1999 and obtained from Mashhad’s municipality. Google Earth high resolution imagery was used to assess accuracy of 2013 classification. Kappa coefficients for classification results were respectively 0.89, 0.88 and 0.91 for 1986, 1999 and 2013. Overall accuracy was 0.903, 0.896 and 0.930, respectively, for 1986, 1999 and 2013. Additional variables influencing urban growth were selected based on research.
2.3. Spatial zoning

A problematic issue in urban growth models is model’s calibration and validation is a problematic issue in urban growth simulations (Silva and Clarke 2002, Jantz et al. 2010, Wang et al. 2011, 2012). To improve accuracy in model’s calibration and validation, the present study integrates a spatial zoning approach with a traditional land-use change model. Using a spatial zoning approach is justified from the fact that Mashhad urban

Figure 2. Process of land change modeling.

literature and expert knowledge and partitioned in four groups: (1) distance-based variables, (2) elevation, (3) density-based variables and (4) class-based transformation rates (Supplementary Materials, Table 1).
districts display distinct growth patterns based on their socioeconomic characteristics, territorial potentials and biophysical constraints. Spatial zoning is intended to improve model calibration and validation and to obtain higher accuracy in the simulation of future urban growth. Spatial zoning implies that the study area can be segmented into homogeneous districts. For this purpose, administrative divisions or geographic units defined with specific criteria can be used. Earlier studies have simulated small-scale urban expansion at enumeration district, census track or cadastral parcel spatial levels (Pinto and Antunes 2010). In some cases, simulating urban expansion at detailed spatial scales has provided more realistic urban scenarios compared with those assured by standard aspatial models (Barreira-González et al. 2015b). However, administrative districts are often heterogeneous as far as urban growth patterns are concerned.

Thiessen polygons were used in this study to classify the study area into homogeneous districts as far as settlement density and dispersion are concerned. This methodology was first suggested by Thiessen (1911) as a way of spatial interpolation of biophysical variables. Thiessen polygons have been used in spatial partitioning problems across several disciplines, including geospatial science (Moreno-Regidor et al. 2012). This approach was also applied in economic analysis and urban planning, e.g. in market zone analysis (Okabe et al. 2000), retail trade zone design (Boots and South 1997) and service location problems (Okabe and Suzuki 1997). Thiessen spatial partitions make use of geometric distance to determine neighborhoods, producing a polygon around each target location that identifies all places matching with that target.

To classify the study area into homogeneous districts, pixels of each land-use class at time 1 (T1) and 2 (T2) were determined (30 m grid) with changes over time being calculated between T1 and T2. The required points for creation of Thiessen polygons were determined using the ArcGIS spatial analyst tool and visual analysis, considering density and dispersion of the changed pixels. To facilitate point selection, density of changed areas was calculated using a kernel density function. Locations with high and low density of settlements were selected as candidate district’s centers. District boundaries were thus determined using a Thiessen polygons function based on selected points. To ensure the appropriateness of Thiessen polygons, visual analysis was applied to correct few nodes not representative of Mashhad’s urban structure. Thiessen polygons separated dense urban zones from districts with dispersed settlements and discontinuous urban expansion. The spatial distribution of changed pixels was assessed using Moran’s I index of spatial autocorrelation. In each zone, clustered or dispersed patterns indicate appropriate discrimination from neighboring polygons. Adjacent zones with similar land-use distribution and urban growth rates were finally merged.

2.4. Model calibration and simulation

Urban growth was simulated separately for each district at T3 (Time 3: 2013) based on land-use data at T1 (1986) and T2 (1999). To validate model accuracy, urban growth over 1999–2013 was simulated based on 1986–1999 data. Urban growth was subsequently simulated for 2025 (T4) based on 1999–2013 data. Urban expansion was supposed to advance proportionally to the growth rate observed in the previous time period at each district. Local-scale growth drivers were identified using Cramer’s V coefficient analysis. Cramer’s coefficients were used to quantify the association between variables and land-
use classes. A high value for Cramér’s $V$ indicates a strong relationship between two variables. Variables with a low Cramér’s $V$ can be removed from subsequent analysis (Mozumder and Tripathi 2014); in this way, variables poorly associated with urban growth were excluded from the model. This would facilitate model’s calibration and improves the accuracy of the final model.

Simulation was done using the land change modeler (LCM) provided by IDRISI (TerrSet). Urban growth in T3 (2013) was assessed using two variables: the transition potential map (TPM) and the amount of land-use change. Taken as one of the most effective stochastic processes to assess changing-state probability (Jokar et al. 2013), a MC process was used to quantify the amount of change for each land-use class and time interval. MC models are considered effective tools for modeling land-use transformations, providing indicators for direction and magnitude of change (Benito et al. 2010, Eastman 2015). MC analysis elaborates a transition probability matrix (A):

$$A = \begin{bmatrix}
A_{11} & A_{12} & A_{13} \\
A_{21} & A_{22} & A_{23} \\
\vdots & \vdots & \vdots \\
A_{n1} & A_{n2} & A_{n3}
\end{bmatrix},$$

where $A_{ij}$ (with both $i$ and $j = 1, 2, \ldots, n$) is the area changing from the $i$-th state to the $j$-th state during the investigated time interval, and $n$ is the number of land-use classes.

### 2.5. Multi-Layer perceptron (MLP) neural network

A Multi-Layer Perceptron (MLP) artificial neural network was used to determine the potential of land to transition and spatial location of transitioning cells. An artificial neural network consists of a connected network of processing units that are modeled on basic properties of neurons in the human brain. Neural networks are nonlinear and can be conceived as a complex mathematical function that converts input data to a desired output. The MLP approach using Back-Propagation (BP) learning algorithms is one of the most widely used neural network models. A typical MLP network contains one input layer, one or more hidden layers and one output layer (Eastman 2015). These layers are used for data input, data processing and data output, respectively (Hu and Weng 2009).

In our study input layers represent areas experiencing land-use changes and the related criteria. The hidden layer was used to identify the relations between changed pixels and criteria. The output layer is based on a TPM.

Functioning of a MLP neural network involves two major steps, forward and backward propagation, to accomplish its modification of the neuron connection weights. During training, each sample is fed into the input layer and the receiving node sums the weighted signals from all nodes to which it is connected in the preceding layer. The following algorithm was used to calculate the input that a single MLP node $j$ received:

$$\text{net}_j = \sum_{i=1}^{m} W_{ij} \varphi_i,$$
where \( net_j \) refers to the input that a single node \( j \) receives; \( w_{ij} \) represents the weight between node \( i \) and node \( j \) and \( O_i \) is the output from node \( i \). The output from a node \( i \) was calculated as follows:

\[
O_j = f(net_j)
\] (3)

The function \( f \) is usually a nonlinear sigmoidal function that is applied to the weighted sum of inputs before the signal passes to the next layer. Once the forward signal is finished, the activities of the output nodes are compared with the related expected activities. Except unusual circumstances, the actual output will differ from the desired outcome, the difference being associated with errors in the network. Errors are then propagated backward with weights for relevant connections corrected via a relation known as the delta rule:

\[
\Delta w_{ji}(t+1) = \eta \delta_i O_i + \alpha \Delta w_{ji}(t),
\] (4)

where \( \eta \) is learning rate, \( \alpha \) is momentum factor and \( \delta \) is computed error. The forward and backward passes continue until the network has learned the characteristics of all land-use classes. The purpose of network training is to get the proper weights for the connections between input and hidden layers, and between hidden and output layers. The number of the hidden layer nodes is estimated by the following equation:

\[
N_h = INT\left( \sqrt{N_i \times N_o} \right),
\] (5)

where \( N_h, N_i \) and \( N_o \) are the number of the hidden, input and output layer nodes respectively (Eastman 2015).

### 3. Results

#### 3.1. Land-use change analysis

Land-use maps for the study area are shown in Figure 3. These contain seven classes: (i) urban land-use; (ii) vegetation; (iii) urban green space; (iv) open land; (v) mountainous and rocky land; (vi) water bodies and (vii) sedimentary land. Maps indicate that Mashhad has expanded more in the northwest direction than in other directions. Class area and changes over time are reported in Table 1. Built-up areas increased respectively by 7039 ha (70.8%), 7246 ha (42.7%) and 14286 ha (143.7%) in the time intervals 1986–1999, 1999–2013 and 1986–2013. Other classes except open land (1999–2013) have experienced negative rates of change.

#### 3.2. Spatial zoning

To partition the study area into homogeneous districts, land-use maps for 1986 and 1999 were overlaid computing change-detection maps for conversion from non-urban classes to built-up areas; spatial density of changed areas were estimated using a kernel density function. District centers were selected using visual analysis and were used as input for creation of Thiessen polygons. To improve simulation accuracy, urban growth was predicted separately for each Thiessen polygon. Point selection and creation of Thiessen polygons was iterated several times until the appropriate zoning was obtained.
Figure 4(a) shows the density of urbanized areas for 1986 and 1999 along with the selected points and Thiessen polygons.

The spatial distribution of transformed pixels is expected to differ between urban districts. For this purpose, we computed the Moran’s I coefficient of spatial autocorrelation in the use of land based on the polygons created above (Figure 4(b)). Some adjacent districts have similar coefficients and were merged to reduce simulation runtime, obtaining a total of 18 districts with Moran’s I coefficients illustrated in Figure 4(c) and Supplementary Materials, Table 2. Peripheral districts showed higher Moran’s I coefficients compared with inner districts, outlining relevant differences in growth pattern and amount of change among urban zones.

Figure 3. Land maps of Mashhad for 1986 (a), 1999 (b) and 2013 (c).

Table 1. Land-use class area and changes over time (hectare) in Mashhad, 1986–2013.

| Class               | 1986  | 1999  | 2013  | 1986–1999 | 1999–2013 | 1986–2013 |
|---------------------|-------|-------|-------|-----------|-----------|-----------|
| Vegetation          | 39,203| 36,604| 31,262| −6.63     | −14.59    | −20.26    |
| Mountainous land    | 32,834| 32,337| 29,662| −1.51     | −8.27     | −9.66     |
| Water bodies        | 71    | 26    | 18    | −63.38    | −30.77    | −74.65    |
| Open land           | 37,411| 34,140| 35,922| −8.74     | 5.22      | −3.98     |
| Sedimentary land    | 1153  | 698   | 170   | −39.46    | −75.64    | −85.26    |
| Urban land use      | 9942  | 16982 | 24229 | 70.81     | 42.67     | 143.70    |
| Urban green space   | 1341  | 1167  | 685   | −12.98    | −41.30    | −48.92    |
3.3. Simulating urban expansion for 2013

Urban expansion was simulated separately for all districts using MC analysis and a MLP neural network (Figure 5(b)). Table 2 provides a summary of the Markov probability matrix used for simulating the transition between initial and final cell state. The highest probability of conversion to built-up areas was observed for open land (4270 ha), vegetation (1685 ha), mountainous and rocky land (503 ha) and urban green space (412 ha).

Since our aim is simulation of urban growth, changes to urban land-use were specifically considered in the analysis. Supplementary Materials Table 3 shows

### Table 2. Markov transition probability for the study area, 1999–2013.

| Class                  | Vegetation | Mountainous land | Water bodies | Open land | Sedimentary land | Urban land-use | Urban green space |
|------------------------|------------|------------------|--------------|-----------|------------------|----------------|-------------------|
| Vegetation             | 0.5983     | 0                | 0            | 0.3458    | 0.0008           | 0.046          | 0.009             |
| Mountainous land       | 0          | 0.9837           | 0            | 0         | 0.0007           | 0.0156         | 0                 |
| Water bodies           | 0.0797     | 0                | 0.2238       | 0.5286    | 0.0664           | 0.0904         | 0.011             |
| Open land              | 0.3415     | 0                | 0            | 0.5179    | 0.0072           | 0.1251         | 0.0083            |
| Sedimentary land       | 0.1752     | 0                | 0.0006       | 0.4301    | 0.2518           | 0.1109         | 0.0314            |
| Urban land-use         | 0          | 0                | 0            | 0         | 0                | 1              | 0                 |
| Urban green space      | 0.0176     | 0                | 0            | 0.292     | 0.0041           | 0.3533         | 0.3329            |

Figure 5. Simulation and validation maps. TPM in the non-zoning simulation (a1), TPM in the zoning simulation (b1), urban growth simulated map in the non-zoning simulation in 2013 (a2), urban growth simulated map in the zoning simulation in 2013 (b2), validation map in the non-zoning simulation (a3), validation map in the zoning simulation (b3).
Table 3. Cramer’s coefficient quantifying the association between variables and urban area by district, for 1999–2013 simulation (Dist: distance); bold indicates significant association at $p < 0.05$.

| District number | Elevation | Road density | Settlement density | Dist to parks | Dist to rivers | Dist to roads | Dist to rocks | Dist to sediment | Dist to urban 1986 | Dist to urban 1999 | Dist to vegetation | Evidence Likelihood | Slope |
|-----------------|-----------|--------------|-------------------|---------------|---------------|---------------|---------------|-------------------|-------------------|-------------------|-------------------|------------------|-------|
| Total           | 0.352     | 0.29         | 0.34              | 0.29          | 0.16          | 0.31          | 0.28          | 0.2               | 0.31              | 0.33              | 0.37              | 0.54             | 0.24  |
| 1               | 0.37      | 0.35         | 0.4               | 0.35          | 0.2           | 0.33          | 0.35          | 0.2              | 0.31              | 0.33              | 0.35              | 0.55             | 0.17  |
| 2               | 0.4       | 0.26         | 0.32              | 0.22          | 0.17          | 0.31          | 0.32          | 0.2              | 0.31              | 0.32              | 0.32              | 0.36             | 0.465 | 0.2   |
| 3               | 0.11      | 0.27         | 0.315             | 0.23          | 0.08          | 0.26          | 0.19          | 0.2              | 0.28              | 0.26              | 0.29              | 0.29             | 0.02  |
| 4               | 0.036     | 0.09         | 0.283             | 0.106         | 0.1           | 0.04          | 0.21          | 0.13             | 0.09              | 0.06              | 0.21             | 0.36             | 0.09  |
| 5               | 0.16      | 0.25         | 0.32              | 0.2           | 0.26          | 0.23          | 0.22          | 0.19             | 0.24              | 0.26              | 0.30             | 0.36             | 0.07  |
| 6               | 0.42      | 0.18         | 0.263             | 0.2           | 0.24          | 0.285         | 0.33          | 0.315            | 0.26              | 0.3               | 0.354            | 0.35             | 0.26  |
| 7               | 0.3       | 0.32         | 0.37              | 0.24          | 0.07          | 0.27          | 0.21          | 0.23             | 0.29              | 0.31              | 0.3              | 0.275            | 0.04  |
| 8               | 0.38      | 0.17         | 0.27              | 0.09          | 0.12          | 0.18          | 0.32          | 0.13             | 0.16              | 0.29              | 0.37             | 0.47             | 0.26  |
| 9               | 0.035     | 0.07         | 0.28              | 0.17          | 0.1           | 0.03          | 0.097         | 0.19             | 0.00              | 0.16              | 0.24             | 0.42             | 0.027 |
| 10              | 0.325     | 0.38         | 0.45              | 0.42          | 0.375         | 0.34          | 0.24          | 0.25             | 0.38              | 0.38              | 0.44             | 0.39             | 0.08  |
| 11              | 0.07      | 0.28         | 0.33              | 0.23          | 0.24          | 0.27          | 0.22          | 0.2              | 0.34              | 0.38             | 0.29             | 0.32             | 0.01  |
| 12              | 0.14      | 0.2          | 0.32              | 0.27          | 0.1           | 0.15          | 0.15          | 0.11             | 0.21              | 0.23             | 0.25             | 0.35             | 0.05  |
| 13              | 0.15      | 0.285        | 0.35              | 0.21          | 0.26          | 0.24          | 0.3           | 0.14             | 0.3               | 0.29             | 0.33             | 0.33             | 0.03  |
| 14              | 0.395     | 0.325        | 0.38              | 0.36          | 0.09          | 0.33          | 0.33          | 0.23             | 0.36              | 0.34             | 0.32             | 0.45             | 0.1   |
| 15              | 0.24      | 0.26         | 0.35              | 0.18          | 0.12          | 0.18          | 0.35          | 0.1              | 0.19              | 0.2              | 0.12             | 0.486            | 0.11  |
| 16              | 0.21      | 0.28         | 0.33              | 0.245         | 0.19          | 0.21          | 0.16          | 0.17             | 0.24              | 0.26             | 0.3              | 0.29             | 0.05  |
| 17              | 0.34      | 0.37         | 0.30              | 0.2           | 0.28          | 0.25          | 0.22          | 0.3              | 0.3               | 0.36             | 0.5              | 0.16             |       |
| 18              | 0.275     | 0.07         | 0.27              | 0.13          | 0.21          | 0.26          | 0.2           | 0.25             | 0.28              | 0.29             | 0.31             | 0.5              | 0.14  |
absolute rates of change from non-urban classes to urban land-use calculated for each district. Intensity of change in the use of land for each district was different at the end of the simulation interval (1999–2013) in respect with the rate of change in previous time intervals. For example, the highest amount of open land converted to built-up area was observed in district 5 (1198 ha). The highest amount of land converted to built-up area respectively for vegetation, mountainous land, sedimentary land and urban green was observed in district 6 (211.2 ha), 14 (338.5 ha), 16 (34 ha) and 10 (82.7 ha).

### 3.4. Model calibration and change allocation

Model calibration for each district was carried out selecting important target variables and creating TPMs. Cramer’s coefficients were used to quantify the level of association between selected variables and land-use class and to identify the most relevant factors of growth in each district. According to Mozumder and Tripathi (2014), values of Cramer’s coefficient ranging between 0.15 and 0.4 are satisfactory; significant coefficients are marked in bold (Table 3).

Distance to urban areas, settlement density and evidence likelihood (probability of change to urban land-use) were identified as key variables influencing urban growth in all districts. Distance to sedimentary land and slope were moderately associated to urban land-use and, based on low Cramer’s coefficients, these variables were excluded from the model. Settlement density was an important variable in central districts (1, 2, 3, 5, 7, 10, 17). Distance to roads was relevant in both external (6, 8, 11) and internal city districts (4, 9, 12). Urban green space was a relevant variable in districts 1, 10, 12, 14, 15, 16 and 17. Distance to river was relevant only in districts surrounded by water bodies (18, 5, 16, 13, 10 and 11). Distance to vegetation influenced urban growth in the inner city and external districts where built-up areas are generally located adjacent to patches with natural vegetation. Elevation was recognized as a relevant factor shaping urbanization in districts situated in mountainous areas south and southeast of Mashhad (1, 2, 6, 14, 15, 17 and 18).

After identifying important variables of urban growth for each district, MLP parameters were specified for extraction of TPMs. Number of input layers were chosen in accordance with effective variables for each district; number of hidden layers was considered equal to 20 and TPMs were generated as output layers; momentum factor was considered equal to 0.5. Criteria to stop training were based on a 0.01 root mean square error (RMSE), with 50,000 iterations and 100% accuracy. A sample of pixels (changed or unchanged pixels in the previous time interval, i.e. calibration period) was also selected for MLP training stage. At the end of each training, algorithm performance including RMSE and accuracy was evaluated and best training was selected. Accuracy >80% indicated a good MLP performance; values below 75% were considered not reliable (Eastman 2015). An accuracy index above 75% was obtained for all districts except 12 and 15.

Determination of a TPM for each class completes the training process. TPMs for both non-zoning and zoning simulation modes were achieved by merging all TPMs (transition maps from non-urban classes to urban class). High map values indicate higher probability of change compared to neighboring pixels (Eastman 2015). In the non-zoning
mode (Figure 5(a1)), pixels with the highest probability of change were situated close to inner city. In the zoning mode (Figure 5(b1)) the highest density of pixels with the highest probability of change was found in district 5 and in sparse external city patches. The simulation process was based on the TPMs developed from the MLP neural network and the intensity of change derived from the MC analysis. Figure 5(a2,b2) show results of urban expansion simulation in 2013 in non-zoning and zoning modes, respectively.

3.5. Model validation

3.5.1. Validation of MC analysis

To validate MC estimated changes in the use of land, actual changes in land area (conversion to built-up area) were derived based on cross tabulation of classification data (1999 and 2013) in each district (Table 4). Differences between simulation modes’ output and reality were subsequently computed. This allows comparing the estimated quantity of changes using MC analysis in each district under zoning and non-zoning mode. The highest accuracy was reached with zero-differences; negative and positive values respectively indicate that simulation output was less or more than reality. Although the observed difference for non-zoning mode (−4.07) is less than the difference found using the zoning mode (−6.13), zoning mode performed better than non-zoning mode in all districts except 1, 2, 7, 8 and 11; differences observed in districts 1, 2 and 7 were quite moderate.

3.5.2. Kappa Index of Agreement (KIA)

The Kappa Index of Agreement (KIA) indicates the degree of agreement between simulated and actual changes (Pontius and Suedmeyer 2004). KIA was used to evaluate simulation accuracy and indicates the extent to which two maps agree in terms of

| District ID | Reality | Non-zoning simulation | Zoning simulation | Non-zoning simulation | Zoning simulation |
|-------------|---------|------------------------|-------------------|-----------------------|-------------------|
| 1           | 225     | 188                    | 178               | −16.44                | −20.89            |
| 3           | 398     | 576                    | 318               | 44.72                 | −20.10            |
| 5           | 1498    | 1205                   | 1368              | −19.56                | −8.68             |
| 7           | 239     | 222                    | 262               | −7.11                 | 9.62              |
| 10          | 450     | 368                    | 487               | −18.22                | 8.22              |
| 13          | 234     | 171                    | 173               | 26.92                 | −26.07            |
| 14          | 553     | 262                    | 523               | −52.62                | −5.42             |
| 16          | 925     | 405                    | 948               | −56.22                | 2.49              |
| 17          | 132     | 210                    | 129               | 59.09                 | −2.27             |
| 4           | 46      | 127                    | 59                | 176.09                | 28.26             |
| 9           | 28      | 90                     | 43                | 221.43                | 53.57             |
| 12          | 170     | 155                    | 171               | −8.82                 | 0.59              |
| 15          | 37      | 248                    | 53                | 570.27                | 43.24             |
| 2           | 805     | 759                    | 740               | −5.71                 | −8.07             |
| 8           | 527     | 578                    | 437               | 9.68                  | 17.08             |
| 11          | 171     | 189                    | 31                | 10.53                 | −81.87            |
| 18          | 142     | 100                    | 125               | −29.58                | −11.97            |
| Total area  | 7247    | 6952                   | 6803              | −4.07                 | −6.13             |
location (Pontius and Cheuk 2006). A 0.69 score was observed in the study area with the non-zoning mode increasing to 0.75 when the zoning mode was applied.

### 3.5.3. Validation through map overlay

To evaluate spatial accuracy, simulation results were compared with reality through map overlay. Figure 5(a3,b3) show validation maps for non-zoning and zoning modes, respectively, in 2013 where Hits (green) indicate correct simulated change, Misses (red) indicate non-simulated change and False alarms (yellow) indicate false simulated change. Amount of Hits was 3295 ha in zoning mode and 2750 ha in non-zoning mode (Table 5), improving accuracy by 19.8%. The highest improvements were observed in districts 16, 14, 10 and 5. Ratios of Hits to actual changes and Hits to simulated changes were respectively 0.406 and 0.352 in non-zoning mode, increasing to 0.484 and 0.423 under zoning mode. In central districts (1, 5, 7, 10, 13, 14, 16), accuracy (Hits to actual changes ratio) under zoning mode was higher than under non-zoning mode. In the inner city districts (4, 9 and 15) and in the external city districts (2, 8 and 11) both ratios were found higher under non-zoning mode compared with zoning mode. Taken together, simulation for districts in the external city (8, 18, 6, 11 and 2) was less accurate than in other districts. By contrast, districts 16, 17, 12, 5, 14, 9 and 1 had more accurate simulations.

### 3.6. Simulating urban expansion in 2025

Spatial zoning was used for simulating expansion of Mashhad in 2025. The study area was partitioned into 88 zones based on density of changed areas over 1999–2013 (Figure 6(a)). Adjacent zones with similar distribution of changed pixels were subsequently merged, obtaining a total of 25 districts. The shift from 18 to 25 districts is

| District ID | Zoning simulation (Ha) | Non-zoning simulation (Ha) | Hits | False alarms | Misses | Zoning simulation (Ha) | Non-zoning simulation (Ha) | Hits | False alarms | Misses | Zoning | Non-zoning |
|-------------|------------------------|-----------------------------|------|--------------|--------|------------------------|-----------------------------|------|--------------|--------|---------|------------|
| 1           | 104.7                  | 135                         | 103.7| 84.6         | 137    | 0.587                  | 0.551                        | 0.436| 0.431        |
| 3           | 117.3                  | 339                         | 167.8| 409.14       | 290    | 0.368                  | 0.291                        | 0.257| 0.366        |
| 5           | 833.8                  | 709                         | 708.1| 497          | 834.75 | 0.61                   | 0.587                        | 0.54 | 0.549        |
| 7           | 83.52                  | 153.27                      | 81.72| 140.58       | 167.13 | 0.335                  | 0.368                        | 0.352| 0.328        |
| 10          | 247.3                  | 248.85                      | 210.87| 339.21       | 0.508  | 0.426                  | 0.498                        | 0.498| 0.316        |
| 13          | 83                     | 159                         | 81.72| 89.46        | 160.65 | 0.483                  | 0.477                        | 0.343| 0.337        |
| 14          | 309                    | 257                         | 181.3| 80.91        | 384.93 | 0.591                  | 0.691                        | 0.546| 0.32         |
| 16          | 640                    | 327                         | 113  | 674.91       | 0.675  | 0.721                  | 0.662                        | 0.302| 0.292        |
| 17          | 79.56                  | 56.88                       | 94.41| 116          | 42.84  | 0.616                  | 0.449                        | 0.583| 0.688        |
| 4           | 32.85                  | 42.12                       | 85.14| 39.69        | 0.55   | 0.331                  | 0.402                        | 0.515| 0.515        |
| 9           | 31.59                  | 42.33                       | 45.45| 44.46        | 34.47  | 0.736                  | 0.76                          | 0.395| 0.568        |
| 12          | 108.5                  | 97.3                        | 86.85| 68.76        | 119    | 0.632                  | 0.558                        | 0.527| 0.422        |
| 15          | 17                     | 19.7                        | 27.54| 220.59       | 0.321  | 0.111                  | 0.463                        | 0.734| 0.734        |
| 2           | 231.7                  | 635.7                       | 534.33| 615          | 0.313  | 0.296                  | 0.267                        | 0.267| 0.267        |
| 8           | 84.6                   | 477                         | 118.3| 459.18       | 443.34 | 0.194                  | 0.205                        | 0.151| 0.211        |
| 11          | 21.5                   | 151                         | 34.92| 153.54       | 138.51 | 0.682                  | 0.185                        | 0.124| 0.201        |
| 18          | 34.47                  | 145.35                      | 33.93| 65.7         | 145.89 | 0.275                  | 0.341                        | 0.191| 0.188        |
| Total area  | 3295                   | 4485                        | 2750 | 4021         | 5052   | 0.484                  | 0.406                        | 0.423| 0.352        |
Figure 6. Urban expansion simulation in 2025. District boundaries (a), transition potential map (b), simulated urban expansion (c).
motivated with changes in urban boundaries and socioeconomic characteristics during 1986–1999 and 1999–2013, reflecting distinct growth rates and land-use change patterns. Figure 6(b) shows TPM derived from MLP neural network at the district scale. According to this map, adjacent land to mountainous areas southwest of the city and free land spatially separated from consolidated urban areas have more potential for urbanization. Figure 6(c) illustrates simulation of the city expansion in 2025. According to simulation, 5818 ha will add to urban land-use in the 2013–2025 time intervals. The major growth has occurred in patches physically separated from consolidated city. Urban land-use will expand mostly on open land (vegetation and agriculture land) along north, northeast, east and southeast directions. Mountainous land in southern and southwestern parts of the metropolitan region acts as a natural barrier for future urban expansion.

4. Discussion

While appropriate identification of urban districts allows improving model accuracy, effective methodologies classifying districts into homogeneous groups are still required. Many of the models and algorithms used to partitioning space into homogeneous zones (e.g. cluster analysis, graph partitioning, mathematical programming) are efficient only when applied to problems in which the number of zones (M) and/or basic units (N) is low, or with a set of less restrictive spatial conditions, as in cases where centers of a given region are predetermined (Moreno-Regidor et al. 2012). In this study, we used a Thiessen polygons method based on density and distribution of changed pixels in the calibration period. Thanks to the use of spatial information and computational geometry in the zoning process, determining number and boundaries of urban districts according to settlement morphology proved to be more flexible than using administrative divisions. However, regularity in polygon boundaries may prevent a precise matching of polygons with urban districts’ boundaries. Using improved versions of Voronoi diagrams, e.g. power Voronoi diagrams, additively weighted Voronoi diagrams (AWVDs) or multiplicatively weighted Voronoi diagrams (MWVDs) by smoothing boundary of zones, may contribute to an effective identification of homogeneous districts.

Another limitation in the use of Thiessen polygons in urban growth simulations is that changing polygons could impact final model’s outcome; adding or removing a point in a specific location would influence the state of surrounding polygons, impacting simulation results in a given district. In this sense, points have to be chosen based on the spatial structure of the changed area. Unsuitable merging of neighboring districts can negatively influence model accuracy. To overcome this problem, chosen districts should have similar growth patterns that can be verified using a Moran’s spatial autocorrelation analysis.

Because cities are expanding and their boundaries are changing over time, the position of district’s centers may change. Differential point selection and polygon design according to urban dynamics would help predicting realistic urban spatial structures. Accuracy is another important issue in land change simulation when evaluating overall model performance. To evaluate performance of Thiessen polygons, overall accuracy should be compared with other approaches after modeling.
land change at districts generated using different zoning systems. Kappa indexes comparing observed accuracy relative to a baseline accuracy expected due to randomness, were used in this study. However, in practical applications, randomness is an uninteresting, irrelevant and/or misleading baseline. Also the standard Kappa and its computational variants result rather complicated, difficult to understand and unhelpful to interpret (Pontius et al. 2011). In this sense, allocation agreement and disagreement was regarded as useful criterion to validate simulation results. Allocation indexes are computed from map overlay (or cross-tabulation matrix) of reference maps and simulated maps. Overlay produces maps that show three types of information, namely Hits (agreement), False alarms (disagreement) and Misses (Olmedo et al. 2015). Simple indexes can be derived by linear composition of these variables with the aim at validating different approaches in land change simulation.

Based on empirical evidences illustrated in our study, Mashhad’s districts were classified into three categories based on land-use patterns and amount of change: (i) fringe districts adjacent to open land and vegetation, displaying rapid urbanization rates; (ii) external city districts with moderate urbanization rate and a lower simulation accuracy compared with fringe districts; (iii) inner city districts experiencing moderate urban changes due to restricted availability of buildable land.

5. Conclusions

Land-use simulation is important for a variety of planning and management issues as well as for academic research (Deng et al. 2008). In this study, a spatial zoning approach was used to improve calibration and to increase accuracy of a model simulating urban growth in Mashhad. Under the hypothesis that quantity of change is proportional to urbanization rate in earlier time intervals, use of spatial zoning increased the accuracy of MC analysis, containing the impact of heterogeneity in land-use and urban growth patterns on simulation outcomes and contributing in the identification of place-specific, relevant drivers of urban growth.

Disclosure statement

No potential conflict of interest was reported by the authors.

References

Assunção, R.M., et al., 2006. Efficient regionalization techniques for socio-economic geographical units using minimum spanning trees. *International Journal of Geographical Information Science, 20* (7), 797–811. doi:10.1080/13658810600665111
Bell, E.J., 1974. Markov analysis of land use change – an application of stochastic processes to remotely sensed data. *Socio-Economic Planning Sciences, 8* (6), 311–316. doi:10.1016/0038-0121(74)90034-2
Benito, P.R., et al., 2010. Land use change in a Mediterranean metropolitan region and its periphery: assessment of conservation policies through CORINE Land Cover data and Markov models. *Forest Systems, 19*, 315–328. doi:10.5424/fs/2010193-8604
Boots, B. and South, R., 1997. Modeling retail trade areas using higher-order, multiplicatively weighted Voronoi diagrams. *Journal of Retailing*, 73 (4), 519–536. doi:10.1016/S0022-4359(97)90033-6

Cao, K., et al., 2014. Calibrating a cellular automata model for understanding rural-urban land conversion: A Pareto front-based multi-objective optimization approach. *International Journal of Geographical Information Science*, 28 (5), 1028–1046. doi:10.1080/13658816.2013.851793

Casaer, J., et al., 1999. Analysing space use patterns by Thiessen polygon and triangulated irregular network interpolation: a non-parametric method for processing telemetric animal fixes. *International Journal of Geographical Information Science*, 13 (5), 499–511. doi:10.1080/136588199241175

Eastman, J.R., 2015. IDRISI TerrSet, guide to GIS and image processing, manual version 18.0. Worcester, MA: Clark University.

Geneletti, D., 2012. Assessing the impact of alternative land-use zoning policies on future ecosystem services. *Environmental Impact Assessment Review*, 40, 25–35. doi:10.1016/j.eiar.2012.12.003

Jantz, C.A., et al., 2010. Designing and implementing a regional urban modeling system using the SLEUTH cellular urban model. *Computers, Environment and Urban Systems*, 34, 1–16. doi:10.1016/j.compenvurbsys.2009.08.003
Lambin, E.F. and Geist, H.J., 2006. Land-use and land-cover change: local processes and global impacts. Berlin: Springer.

Li, X., et al., 2011. Coupling urban cellular automata with ant colony optimization for zoning protected natural areas under a changing landscape. International Journal of Geographical Information Science, 25 (4), 575–593. doi:10.1080/13658816.2010.481262

Li, X. and Yeh, A.G.O., 2002. Neural-network-based cellular automata for simulating multiple land use changes using GIS. International Journal of Geographical Information Science, 16, 323–343. doi:10.1080/13658810210137004

Liu, X., et al., 2011. Zoning farmland protection under spatial constraints by integrating remote sensing, GIS and artificial immune systems. International Journal of Geographical Information Science, 25 (11), 1829–1849. doi:10.1080/13658816.2011.557380

Luo, G., et al., 2010. Combining system dynamic model and CLUE-S model to improve land use scenario analyses at regional scale: a case study of Sangong watershed in Xinjiang, China. Ecological Complexity, 7 (2), 198–207. doi:10.1016/j.ecocom.2010.02.001

Matthews, R.B., et al., 2007. Agent-based land-use models: a review of applications. Landscape Ecology, 22 (10), 1447–1459. doi:10.1007/s10980-007-9135-1

Moreno-Regidor, P., Lacalle, J.G.L.D., and Manso-Callejo, M.-Á., 2012. Zone design of specific sizes using adaptive additively weighted Voronoi diagrams. International Journal of Geographical Information Science, 26 (10), 1811–1829. doi:10.1080/13658816.2012.655742

Morillo, M.C., García Cepeda, F., and Martínez-Cuevas, S., 2016. The application of spatial analysis to cadastral zoning of urban areas: an example in the city of Madrid. Survey Review, 1–10. doi:10.1080/00396265.2015.1113029

Mozumder, C. and Tripathi, N.K., 2014. Geospatial scenario based modelling of urban and agricultural intrusions in Ramsar wetland Deepor Beel in Northeast India using a multi-layer perceptron neural network. International Journal of Applied Earth Observation and Geoinformation, 32, 92–104. doi:10.1016/j.jag.2014.03.002

Ochoa, A., Bernabé, B., and Ochoa, O., 2009. Towards a parallel system for demographic zonification based on complex networks. Journal of Applied Research and Technology, 7 (2), 218–232.

Okabe, A., et al., 2000. Spatial Tessellations. Concepts and Applications of Voronoi Diagrams. Chichester: Wiley.

Okabe, A. and Suzuki, A., 1997. Locational optimization problems solved through Voronoi diagrams. European Journal of Operational Research, 98, 445–456. doi:10.1016/S0377-2217(97)80001-X

Olmedo Jr, M.T.C., et al., 2015. Comparison of simulation models in terms of quantity and allocation of land change. Environmental Modelling & Software, 69, 214–221. doi:10.1016/j.envsoft.2015.03.003

Pijanowski, B.C., et al., 2014. A big data urban growth simulation at a national scale: configuring the GIS and neural network based land transformation model to run in a high performance computing (HPC) environment. Environmental Modelling & Software, 51, 250–268. doi:10.1016/j.envsoft.2013.09.015

Pinto, N.N. and Antunes, A.P., 2010. A cellular automata model based on irregular cells: application to small urban areas. Environment and Planning B: Planning and Design, 37, 1095–1114. doi:10.1068/b36033

Poelmans, L. and Van-Rompaey, A., 2009. Complexity and performance of urban expansion models. Computers, Environment and Urban Systems, 34 (1), 17–27. doi:10.1016/j.compenvurbsys.2009.06.001

Pontius Jr, R.G. and Cheuk, M.L., 2006. A generalized cross-tabulation matrix to compare soft-classified maps at multiple resolutions. International Journal of Geographical Information Science, 20 (1), 1–30. doi:10.1080/13658810500391024

Pontius Jr, R.G. and Suedmeyer, B., 2004. Components of agreement in categorical maps at multiple resolutions. In: R.S. Lunetta and J.G. Lyon, eds. Remote sensing and GIS accuracy assessment. Boca Raton: CRC Press, 233–251.
Pontius Jr, RG., et al., 2011. Death to Kappa: birth of quantity disagreement and allocation disagreement for accuracy assessment. *International Journal of Remote Sensing*, 32(15), 4407–4429.

Qiu, B.W. and Chen, C.C., 2008. Land use change simulation model based on MCDM and CA and its application. *Acta Geographica Sinica*, 63, 165–174.

Rafiee, R., et al., 2009. Simulating urban growth in Mashhad City, Iran through the SLEUTH model (UGM). *Cities*, 26, 19–26. doi:10.1016/j.cities.2008.11.005

Santé, I., et al., 2016. A simulated annealing algorithm for zoning in planning using parallel computing. *Computers, Environment and Urban Systems*, 59, 95–106. doi:10.1016/j.compenvurbsys.2016.05.005

Shen, T.-Y., et al., 2016. A simulated annealing algorithm for zoning in planning using parallel computing. *Computers, Environment and Urban Systems*, 35, 116–125. doi:10.1016/j.compenvurbsys.2010.10.003

Wang, F., et al., 2011. Identifying dominant factors for the calibration of a land-use cellular automata model using Rough Set Theory. *Computers, Environment and Urban Systems*, 35, 116–125. doi:10.1016/j.compenvurbsys.2010.10.003

Thiessen, A.H., 1911. Precipitation averages for large areas. *Monthly Weather Review*, 39, 1082–1089.

Vega, A.P., Mas, J.-F., and Zielinska, A.L., 2012. Comparing two approaches to land use/cover change modeling and their implications for the assessment of biodiversity loss in a deciduous tropical forest. *Environmental Modelling & Software*, 29 (1), 11–23. doi:10.1016/j.envsoft.2011.09.011

Thapa, R.B. and Murayama, Y., 2011. Scenario based urban growth allocation in Kathmandu Valley, Nepal. *Landscape and Urban Planning*, 105, 140–148. doi:10.1016/j.landurbplan.2011.12.007

Wei, B.C. and Chai, W.Y., 2004. A multi-objective hybrid metaheuristic approach for GIS-based spatial zoning model. *Journal of Mathematical Modelling and Algorithms*, 3, 245–261. doi:10.1023/B:JMMA.0000038615.32559.af

Williams, J.C., 2002. A zero-one programming model for contiguous land acquisition. *Geographical Analysis*, 34 (4), 330–349. doi:10.1111/gean.2002.34.issue-4

Wu, F., 1998. Simulating urban encroachment on rural land with fuzzy-logic-controlled cellular automata in a geographical information system. *Environmental Management*, 53, 293–308.

Yang, Q., Li, X., and Shi, X., 2008. Cellular automata for simulating land use changes based on support vector machines. *Computers & Geosciences*, 34 (6), 592–602. doi:10.1016/j.cageo.2007.08.003

Yang, Q.S. and Li, X., 2007. Integration of multi-agent systems with cellular automata for simulating urban land expansion. *Scientia Geographica Sinica*, 27, 542–548.