Performance evaluation of land change simulation models using landscape metrics

Article in Geocarto International · March 2016
DOI: 10.1080/10106049.2016.1167967

4 authors, including:

Sadeq Dezhkam
University of Tehran
10 PUBLICATIONS  70 CITATIONS
SEE PROFILE

Bahman Jabbarian Amiri
University of Tehran
22 PUBLICATIONS  269 CITATIONS
SEE PROFILE

Ali A Darvishsefat
University of Tehran
75 PUBLICATIONS  463 CITATIONS
SEE PROFILE

Some of the authors of this publication are also working on these related projects:

Planning forest road network with low construction cost View project

All content following this page was uploaded by Sadeq Dezhkam on 17 May 2017.
The user has requested enhancement of the downloaded file. All in-text references underlined in blue are added to the original document and are linked to publications on ResearchGate, letting you access and read them immediately.
Performance evaluation of land change simulation models using landscape metrics

Sadeq Dezhkam, Bahman Jabbarian Amiri, Ali Asghar Darvishsefat & Yousef Sakieh

To cite this article: Sadeq Dezhkam, Bahman Jabbarian Amiri, Ali Asghar Darvishsefat & Yousef Sakieh (2016): Performance evaluation of land change simulation models using landscape metrics, Geocarto International, DOI: 10.1080/10106049.2016.1167967

To link to this article: http://dx.doi.org/10.1080/10106049.2016.1167967

Accepted author version posted online: 18 Mar 2016.
Published online: 08 Apr 2016.

Submit your article to this journal

Article views: 31

View related articles

View Crossmark data
Performance evaluation of land change simulation models using landscape metrics

Sadeq Dezhkama, Bahman Jabbarian Amiri, Ali Asghar Darvishsefat, and Yousef Sakieh

Faculty of Natural Resources, Department of Environmental Science, University of Tehran, Karaj, Iran; Faculty of Natural Resources, Department of Forestry, University of Tehran, Karaj, Iran; Department of Environment, Gorgan University of Agricultural Sciences and Natural Resources, Gorgan, Iran

ABSTRACT
This study proposes a landscape metrics-based method for model performance evaluation of land change simulation models. To quantify model performance at both landscape and class levels, a set of composition- and configuration-based metrics including number of patches, class area, landscape shape index, mean patch area and mean Euclidean nearest neighbour distance were employed. These landscape metrics provided detailed information on simulation success of a cellular automata-Markov chain (CA-Markov) model standpoint of spatial arrangement of the simulated map versus the corresponding reference layer. As a measure of model simulation success, mean relative error (MRE) of the metrics was calculated. At both landscape and class levels, the MRE values were accounted for 22.73 and 10.2%, respectively, which are further categorised into qualitative measurements of model simulation performance for simple and quick comparison of the results. Findings of the present study depict a hierarchical and multi spatial level assessment of model performance.

1. Introduction
Understanding, analysis and modelling land use/land cover (LULC) transformations have been an important area of research in professional community of spatial sciences (Herold et al. 2003, 2005; Braimoh & Onishi 2007; Jing & Zhiyuan 2011; Masnavi 2013; Asgarian et al. 2014; Ceccarelli et al. 2014; Hasani Sangani et al. 2014; Jaafari et al. 2015; Wakode 2014; Sakieh et al. 2015). Complexity, dynamic inheritance and uncertainty of natural systems necessitate an abstractive representation (representation of part of the reality for a specific purpose) of these environments for modelling-informed sustainable LULC practices. In addition, increasing computational ability and greater availability of spatial data as well as statistical and mathematical algorithms have accelerated development of functional models for simulating the dynamic process of LULC conversion. Spatially-explicit models are reliable tools for demonstrating and analysing of multi-dimensional phenomena. In recent decades, there has been a growing interest over development of dynamic micro-simulation methods such as agent-based (Hosseinali et al. 2013; Behzadi & Alesheikh 2013, Jokar et al. 2013b; Liu et al. 2014) and Cellular Automata (CA) models (Clarke et al. 1997; Al-Ahmadi et al. 2009; Feng et al. 2011;
Guan et al. 2011; Dezhkam et al. 2013; Sakieh et al. 2014a, 2014b), which facilitate integration of local factors affecting the LULC change process. The CA models apply a bottom-up approach to model and represent real-world processes. In CA class models, global properties of the system is controlled by local performance of the model parameters (Silva & Clarke 2002) and this feature is appropriate to be applied in a modelling practice, since dynamic nature of LULC change process comprises the same attributes (Guan et al. 2011). During recent years, for spatial modelling and analysis a variety of CA-based softwares have been developed, which included iCity (Stevens & Dragičević 2007), DINAMICA (Soares-Filho et al. 2002), CLUE-S (Verburg et al. 2002), SLEUTH (Clarke et al. 1997) and CA-Markov chain model (Eastman 2006).

The CA-Markov chain is a well-known LULC change model that can project mutual transformations of LULC categories into each other. A CA-Markov chain model integrated with GIS technology is an appropriate tool for modelling temporal change and spatial distribution of LULC types (Guan et al. 2011; Sang et al. 2011; Jokar et al. 2013a). The Markov process conducts temporal transformations among various LULC categories based on transition matrices; CA model controls spatial arrangement of the LULC changes in response to local rules and neighbourhood configuration (Santé et al. 2010; Mitsova et al. 2011); and GIS is capable to parameterize the model, compute the transition matrices and determine the transition rules based on neighbourhood arrangement (Batty et al. 1999). Regardless of the numerous applications of this approach, a robust and spatially-explicit evaluation of model performance is still a challenging concern to LULC modellers and practitioners. There are different methods for examining the performance of the LULC change model by which agreement between the simulated LULC and the observed LULC is measured. Methods included simple least squares regression (Silva & Clarke 2002; Rafiee, Mahiny, Khorasani, Darvishsefat et al. 2009), visual comparison, overall agreement and kappa-based indices derived from contingency table (Pontius 2000), relative operating characteristic (ROC) curve (Pearce & Simon 2000; Pontius & Schneider 2001; Pontius & Batchu 2003) and measures such as landscape metrics (Wu et al. 2009; Guan et al. 2011; Jafarnezhad et al. 2015). The kappa-based indices (Kquantity, Klocation, Kno etc.) are not assumed to be an appropriate indication of agreement for location or quantity-based accuracies and are potential to produce misleading information. Pontius and Millones (2011) exposed the indices properties mathematically and illustrate their limitations graphically focusing on kappa’s use of randomness as a baseline, and often-ignored transformation from an observed sample matrix to the estimated population matrix. ROC curve and simple least regression squares might not provide information on agreement between two categorical maps standpoint of spatial patterns of LULC. Therefore, a spatially-explicit performance evaluation method that considers spatial patterns of the simulated LULC maps is of great priority.

Spatial metrics are quantitative indicators of landscape composition- and configuration-based attributes. Applying the landscape metrics for describing environmental processes and spatial patterns are well-documented (e.g. Herold et al. 2003; Rafiee, Mahiny & Khorasani 2009; Paudel & Yuan 2012; Dezhkam 2013). However, few works have assessed applicability of the landscape metrics for model validation and performance evaluation. Herold et al. (2005) highlighted the applicability of the landscape metrics for measuring simulation success in terms of general agreement between spatial forms of urban development by comparing the simulated vs. observed LULCs. Wu et al. (2009) also evaluated SLEUTH model performance in Shenyang metropolitan in China applying multiple methods among which, the landscape metrics-based method had resulted in deep interpretation and evaluation of the model outputs regarding urban morphology. Mas et al. (2010b) confirmed once realistic simulated landscape (e.g. fragmentation) is an important issue, configuration-based metrics can show location and distribution of LULC classes. Guan et al. (2011) employed an integrated CA-Markov chain model incorporated with socio-economic factors in Saga, Japan and the landscape metrics-based evaluation of simulation success indicated close agreement between the simulated and the historical datasets.

Considering the limitations of the aforementioned methods, landscape metrics could be a valuable source of information on spatial pattern of LULC features to validate LULC simulation maps. Specifically, this study seeks to answer the following questions:
(1) Do landscape metrics have the potential for spatially-explicit performance evaluation of land change simulation models?

(2) What are the main benefits of adopting such an approach for measuring simulation success compared to other common indices such as Kappa index of agreement and overall accuracy?

2. Materials and methods

At the first step, LULC maps of the study area were generated. The resultant LULC maps were then divided into two groups: one group for model calibration and the other one for model simulation. The calibration layers were applied to train the CA-Markov chain model based on historical LULC changes in the study area. Applying the calibrated model, historical trend-based of LULC conversions of the subsequent period were then simulated. Employing a series of landscape metrics including Number of Patches (NP), Mean Patch Area (AREA_MN), Total Class Area (CA), Mean Euclidean Nearest Neighbour Distance (ENN_MN) and Landscape Shape Index (LSI), spatial patterns of the simulated LULC and the corresponding observed LULC maps were quantified and compared. Finally, a classification system has been proposed to systematically address the level of agreement between the simulated LULC and the corresponding observed LULC maps considering landscape composition and configuration attributes.

2.1. Study area

The study area encompasses administrative boundary of Rasht County in Guilan province, covers an area of 1215 km² (Figure 1), with 0.9 million inhabitants. Elevation from sea level varies between −27 and 760 m in the area. Average annual precipitation and average monthly temperature is 1500 mm and 15.8 °C, respectively. Annual relative moisture of the weather is 81%, as well. The northern part of the study area is bounded by the southern coast of the Caspian Sea and its southern part is covered by Hyrcanian forests. Due to its agricultural and industrial attractions, the study area has witnessed substantial increase in built-up areas. Hence, a considerable proportion of forest lands have in turn converted into agricultural fields to supply land for increasing population in this area (Iranian Statistics Center 2012).

2.2 Data processing

For conducting the current study, data sets included 1987, 1999 and 2011 Landsat Thematic Mapper (TM) images obtained from United States Geological Survey (USGS) database, topographic maps (1:25,000 scale), obtained from National Cartographic Centre (NCC) and ancillary data consisted of field survey information, aerial photography and GPS data.

The satellite images were geometrically corrected based on NCC topographic maps with an acceptable RMSE (less than one pixel) using nearest neighbourhood algorithm. Quality control indicated that no noticeable distortions such as striping, banding, sweep error and duplicate pixels were found. Following the co-registration of the imagery, a hybrid process including unsupervised (ISODATA clustering algorithm), supervised (maximum likelihood classifier) and on-screen visual digitization of LULC classes was conducted. The obtained layers were then merged and cross-tabulated across different node years for accuracy and consistency purposes. To ensure accuracy and consistency of classification process, the 1987 layer was first totally checked to remove salt-pepper effect and to modify LULC borders (after supervised and unsupervised classifications). Although very time-consuming, this process allowed us to produce a LULC map with highest possible accuracy. In the next step, the 1987 and 1999 maps were cross-tabulated and revised. In this case, as the pervious step, LULC boundaries were carefully delineated and the revised 1999 map was used for cross-tabulation with 2011 layer and the same process was also undertaken (forward revision). Similarly, a backward revision was then
carried out (LULC borders were carefully checked and modified to eliminate any possible inconsistency and misclassifications) (Mahiny & Clarke 2012). Minimum mapping unit (MMU) was decided by 0.4 hectare and polygons less than the threshold were merged into nearest neighbour that shares the longest border. The LULC maps were categorised according to first level of Anderson classification schema (1976), which includes built-up areas, agriculture, forest, barren lands, water bodies and wetland. In the next step, the accuracy of 2011 layer was then quantified because ground truth information was available only for this date (collected by GPS device). In this case, according to Congalton and Green (2009), in study locations with less than one million acre area and LULC categories fewer than 12, a minimum set of 50 samples is required to be collected for each category. Therefore, for our study area with 1215 km² (~300,233 acres), 300 samples (50 points for each category) were collected throughout the entire of study area. The kappa and overall accuracy values (as the most common approaches in accuracy evaluation of a classified image) were then calculated, which were within the acceptable limits (0.90 and 94%, respectively) (Congalton 1991). Regarding accuracy assessment of 1987 and 1999 maps, these layers not only modified based on 2011 map in backward revision, but also the accuracy evaluation was visually undertaken through pixel-by-pixel comparisons with various true and false colour composites of the corresponding dates (1, 2 and 3; 4, 3 and 2; 3, 2 and 1; 2, 3 and 4; 1, 4 and 5; 1, 4 and 7; 5, 4 and 3; 7, 5 and 2), so that no apparent inconsistency was present in the maps in the end. The IDRISI Taiga, ArcGIS 9.3, Microsoft Excel and Fragstats 3.3 were applied for data preparation, model execution and landscape metrics calculation.
2.3 Model calibration

This study employed a coupled CA-Markov chain model in a GIS environment to evaluate the calibration success in terms of measuring the agreement between spatial patterns of the simulated and observed (ground truth) maps. Figure 2 illustrates overall research steps, which have been taken to conduct the present work.

The Markov chain analysis is a theory rooted in the process of the formation of Markov random process systems for the anticipation and optimised control (Jiang et al. 2009). The Markov model not only provides quantification of transformation states between multiple LULC categories, but can also explain the transfer rate among various LULC features. It is commonly implemented in the anticipation of geographical settings with no aftereffect event (Benito et al. 2010; Guan et al. 2011), which has now become an important anticipating algorithm in geographic researches (Veldkamp & Lambin 2001; Sang et al. 2011; Xin et al. 2012).

The performance of CA models is influenced by uncertainties resulting from the interactivity between model parameters, structures and the quality of input data. CA models are capable detailed and dynamic spatial simulation and possess natural affinity to raster GIS and remotely sensed data. In CA models, space is separated into regular cells and the status of each specific cell is determined according to the state of the cell itself in addition to the status of its neighbouring cells at a previous time step. The transition rule is another principal component of CA models that conducts the transition process among different possible states for each cell. The overall behaviour of the system will be decided by the integrated actions of all locally defined transition functions, and therefore, the state of the system become updated in discrete time intervals.

CA-Markov is an integrated Cellular Automat-Markov Chain prediction algorithm that combines a parameter of spatial contiguity in addition to knowledge of the probable spatial arrangement of transitions to Markov chain process. The Markov model concentrates on the quantity in anticipations for LULC changes. Integration of spatial elements is weak in Markov model and the model does not consider various types of LULC transformations in spatial extents (Eastman 2006). The CA model has

![Research flowchart.](image-url)
a powerful conceptualisation of space, which is strong capability for modelling dynamic and complex systems. The CA-Markov chain model, which combines the principles of Markov and CA models, is an innovative tool about time series and space for the advantage of anticipating.

In this study, model calibration and simulation were accomplished based on a historical growth scenario. In this case, this study assumes that LULC changes occurred between 1999 and 2011 are influenced based on their historical profile (changes between 1987 and 1999) and performance of the model was quantified based on historical scenario. Accordingly, the following steps were undertaken to execute the model (according to Mitsova et al. (2011)):

Step 1: Markov transition probabilities matrix, transition areas matrix, and a set of conditional probability images were generated using Markov process. Frequency of cells within a LULC category, which would be converted into other classes under a given probability threshold, is calculated by transition matrices. The 1987 and 1999 LULC maps were applied in this phase (Figure 3). Figure 4 and Tables 1 and 2 depict the outputs of Markov process, which have been derived from this step.

Step 2: CA-Markov procedure was conducted to integrate the influence of neighbourhood configuration and transition probabilities matrix since Markov process does not incorporate spatial dimensions of LULC transitions that are likely to occur. In this case, each cell in the lattice is interactive with the states of immediate neighbours and different states of cells are based on transition rules. The 5 × 5 moving window was chosen and the 1999 LULC map was introduced as a basis image. It should be noted that Wang et al. (2012) evaluated the performance of the CA-Markov chain model in response to several parameters including spatial resolution and filter size. In case of cell size, they concluded that performance of the model was very good at the resolution ranged between 30 and 120 m. They suggested that the best resolution to model land-use change is 30 m (as our study), which fits to the remotely sensed data. Regarding filter size, they report prediction accuracy of the model is quite stable with less than 1% change in its accuracy between the filter of 3 × 3 to 13 × 13. In another study Pan et al. (2010) have been pointed that small cell size and neighbourhood size generate improper expression of the land-use transition. Therefore, referring to relevant studies that applied CA-Markov chain model to simulate LULC changes (Mitsova et al. 2011; Sang et al. 2011), and to prevent data loss

![Image](Figure 3. Input data layers for model calibration.)
Figure 4. Conditional images of LULC categories derived from Markov chain algorithm between the years 1987 and 1999.

Table 1. Transition probabilities matrix derived from Markov algorithm between LULC maps of the years 1987 and 1999 of the Rasht County.

<table>
<thead>
<tr>
<th>Given LULC type</th>
<th>Built-up area</th>
<th>Agriculture</th>
<th>Forest</th>
<th>Water body</th>
<th>Barren land</th>
<th>Wetland</th>
</tr>
</thead>
<tbody>
<tr>
<td>Built-up area</td>
<td>1.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
</tr>
<tr>
<td>Agriculture</td>
<td>0.0372</td>
<td>0.9491</td>
<td>0.0007</td>
<td>0.0087</td>
<td>0.0039</td>
<td>0.0003</td>
</tr>
<tr>
<td>Forest</td>
<td>0.0132</td>
<td>0.0480</td>
<td>0.9265</td>
<td>0.0018</td>
<td>0.0106</td>
<td>0.0000</td>
</tr>
<tr>
<td>Water body</td>
<td>0.0061</td>
<td>0.1062</td>
<td>0.0051</td>
<td>0.8527</td>
<td>0.0299</td>
<td>0.0000</td>
</tr>
<tr>
<td>Barren land</td>
<td>0.0634</td>
<td>0.6150</td>
<td>0.0111</td>
<td>0.0445</td>
<td>0.2659</td>
<td>0.0000</td>
</tr>
<tr>
<td>Wetland</td>
<td>0.0030</td>
<td>0.3095</td>
<td>0.0011</td>
<td>0.0014</td>
<td>0.0000</td>
<td>0.6851</td>
</tr>
</tbody>
</table>

Table 2. Transition areas matrix derived from Markov algorithm between LULC maps of the years 1987 and 1999 of the Rasht County.

<table>
<thead>
<tr>
<th>Given LULC type</th>
<th>Built-up area</th>
<th>Agriculture</th>
<th>Forest</th>
<th>Water body</th>
<th>Barren land</th>
<th>Wetland</th>
</tr>
</thead>
<tbody>
<tr>
<td>Built-up area</td>
<td>12714.84</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Agriculture</td>
<td>3224.34</td>
<td>82240.11</td>
<td>63.27</td>
<td>753.3</td>
<td>341.55</td>
<td>24.75</td>
</tr>
<tr>
<td>Forest</td>
<td>188.82</td>
<td>686.61</td>
<td>13265.55</td>
<td>25.47</td>
<td>151.29</td>
<td>0.36</td>
</tr>
<tr>
<td>Water body</td>
<td>21.51</td>
<td>376.56</td>
<td>18.18</td>
<td>3024</td>
<td>106.02</td>
<td>0</td>
</tr>
<tr>
<td>Barren land</td>
<td>171.09</td>
<td>1659.15</td>
<td>30.06</td>
<td>120.15</td>
<td>717.39</td>
<td>0.09</td>
</tr>
<tr>
<td>Wetland</td>
<td>5.13</td>
<td>537.3</td>
<td>1.98</td>
<td>2.34</td>
<td>0</td>
<td>1189.35</td>
</tr>
</tbody>
</table>
resulting from oversimplification ($7 \times 7$ filter size and bigger), we decided to use $5 \times 5$ filter size. The transition probability matrix and the transition areas matrix derived from the previous stage were imported into the analysis as the model requirement. Each category was then set with a probability threshold for converting into other LULC types based on the calculated transition probabilities matrix in the previous step.

Step 3: Assigning the number of iterations to estimate transition probabilities. It would be noteworthy that the iteration in the CA process indicates an annually-based temporal resolution between simulation node years. In other word, taking spatiotemporal scale of the study into account, the model is able to tract annual changes in LULC categories and finer scales might prevent proper calibration of the transition rules (Eastman 2006). Therefore, in this phase, 12 iterations were considered (which equals to time period between 1987 and 1999) and the 2011 LULC map was consequently simulated. The moving window is applied to conduct land conversions in response to transition probabilities and the state of adjacent cells as a CA function.

2.4 Model performance evaluation

2.4.1 Calculation of kappa and overall accuracy indices

The accuracy of the simulated LULC maps was examined applying confusion matrix-based indices. Kappa and overall accuracy were calculated using Equations (1) and (2) (Congalton & Green 2009). Figure 5 illustrates the simulated LULC map against the reference image.

\[
\text{Overall accuracy} = \frac{\sum_{i=1}^{k} n_{ii}}{n}
\]

where, \( k \) is number of LULC classes, \( n_{ij} \) denotes correctly classified pixels of LULC class \( i \) and \( n \) stands for the total number of the pixels. The percent of correctly classified pixels relative to the whole set of the pixels is calculated by this measure.

Figure 5. The simulated LULC map using CA-Markov model versus reference LULC map produced using hybrid classification.
where, $p_o$ is the real agreement between predicted and reference LULC map and $p_c$ is the chance of agreement between predicted and reference LULC layer.

### 2.4.2 Landscape metrics based performance evaluation

For measuring the agreement between the simulated and the reference maps standpoint of spatial patterns of the LULC categories, a series of landscape metrics were calculated. To compare the simulated- and reference layer-derived metrics, and to evaluate the model performance, Relative Error (RE) and Mean Relative Error (MRE) for the model simulation were measured using Equations (3) and (4) below:

$$
RE = \left[ \frac{M_p - M_r}{M_r} \right] \times 100
$$

where, $M_p$ stands for the value of the landscape metric derived from the simulated map and $M_r$ refers to the value of the landscape metric extracted from the reference LULC map.

$$
MRE = \frac{1}{n} \sum_{i=1}^{n} RE_i
$$

where, $RE_i$ is relative error of the model for each LULC class, which is calculated for each metric and $n$ stands for number of all estimated relative errors.

According to the calculated RE and MRE values, a classification system has been proposed to evaluate simulation success in terms of deviation of the simulated metrics from actual data sets (Table 3). Aiming to simple and quick interpretation, RE values were categorised into qualitative descriptions of model performance, which are represented in Tables 7–9.

Accuracy of landscape metrics is heavily dependent on that of the LULC maps from which, they are computed. Accordingly, the proposed classification system is based on acceptable accuracy of the classified maps. Congalton and Green (2009) noted that an overall accuracy level of 85% can be adopted as a cutoff between acceptable and unacceptable results. This cutoff was first described by Anderson (1976) and later was applied in a numerous works (e.g. Tang et al. 2008; Paudel & Yuan 2012; Rienow & Goetzke 2014; Fan & Myint 2014). Thus, the classification system is proposed on the basis of 15% of classification error.

Process and function in a landscape are influenced by its structure (Forman & Godron 1986; Turner et al. 2001; Botequila et al. 2006; Farina 2006). Hence, geospatial models that are able to simulate landscape structural changes are of great interest for decision makers and policy makers. In turn, landscape structural attributes (composition and configuration) merit to be considered for performance evaluation of land change simulation models.

---

**Table 3.** Proposed relative error classification table for evaluating model performance in terms of variations between landscape metrics’ values in the simulated and actual data.

<table>
<thead>
<tr>
<th>Absolute relative error (%)</th>
<th>Level of agreement (between the simulated and reference maps)</th>
<th>Model performance</th>
</tr>
</thead>
<tbody>
<tr>
<td>0–15</td>
<td>High</td>
<td>Excellent</td>
</tr>
<tr>
<td>15–30</td>
<td>Good</td>
<td>Good</td>
</tr>
<tr>
<td>30–45</td>
<td>Average</td>
<td>Moderate</td>
</tr>
<tr>
<td>More than 45</td>
<td>Low</td>
<td>Weak</td>
</tr>
</tbody>
</table>
2.4.3 Landscape metrics selection procedure

Over the recent decades there has been a proliferation of statistical indices of landscape pattern (O’Neill et al. 1999; Turner et al. 2001; Herold et al. 2002, 2003; Dietzel et al. 2005; DiBari 2007; Weng 2007; Tang et al. 2008; Su et al. 2011; Tian et al. 2011). While these attempts have been providing researchers and practitioners with detailed information of landscape structure, it has also generated a considerable source of confusion. The large number of available metrics presents a major challenge for those who address them to decide on optimum set of the metrics that best describe landscape structure. However, it is challenging to know a priori what combination of the metrics will adequately answer a specific question.

Selecting a set of landscape metrics can be a difficult task (Cushman et al. 2008). This is due to many indices simultaneously quantify multiple characteristics of the landscape structure, which confounds the composition and configuration of the landscape (McGarigal & Marks 1995; Gustafson 1998). Besides, some metrics are logically redundant since they are different ways of quantifying the same basic information. In addition, metrics might provide overlapped information not because that basically measure a similar property of a structure, but because for the targeted landscape, where different characteristics of the structure are correlated.

During the last two decades there has been an effort to determine if the principal elements of landscape structure can be analysed by a small set of independent metrics (Riitters et al. 1995; Griffith et al. 2000; Cushman et al. 2008; Schindler et al. 2008). These studies have reported that patterns can be detected and quantified by a small set of the metrics. Cushman et al. (2008) suggest three criteria of universality, strength and consistency for parsimony in selection of the metrics (for definition of each criterion readers are referred to the mentioned article). Accordingly, we reviewed a series of recent articles in this field that applied landscape metrics for different purposes including spatiotemporal quantification of a landscape pattern, model performance evaluation and landscape designing (Table 4). Finally, by considering the following reasons (Botequila et al. 2006; Cushman et al. 2008); the optimum set of landscape metrics was selected to be specifically employed for model performance evaluation in this study:

1. These metrics should quantify fundamental aspects of a landscape structure including number, area, size, distance and shape of the patches. These concepts accordingly reflect shape complexity, nearest neighbour distance, patch dominance and aggregation of the landscape, which are among universal, consistent and powerful components of a landscape (Cushman et al. 2008);
2. They should be simple and easy to understand and interpret;
3. They should be reliable when applied together for a specific purpose (e.g. model performance evaluation).

Interestingly, in accordance to our literature review, metrics including Number of Patches (NP), Class Area (CA), Landscape Shape Index (LSI), Mean Euclidean Nearest Neighbour Distance (ENN_MN) and Mean Patch Area (AREA_MN) were frequently used over very different regions (Table 5). This reflects the fact that this set of the metrics have not only implemented and interpreted by many investigators, they effectively quantify and describe class-level and landscape-level properties of a region. In addition, these metrics are easy to interpret and understand and they could produce reliable results when applied together in different study areas.

NP metric indicates number of patches in the landscape of patch type (class) and equals the number of patches of the corresponding patch type (class). NP quantifies fundamental components of a landscape structure and fragmentation. This metric is best used in comparative analyses between landscapes with similar extents and LULC classification. (Botequila et al. 2006). This metric was applied to clearly address model’s ability in correctly simulation of patch numbers at both landscape and class levels. The formula is as follows (Mc Garigal & Marks 1995):

\[ NP = n_i \]
Table 4. The employed landscape metrics in different study areas and for different purposes including spatiotemporal quantification of landscape pattern, model performance evaluation and landscape designing.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of patches (nP)</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td></td>
<td>*</td>
<td>*</td>
<td></td>
<td></td>
<td>*</td>
<td>*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Class area (CA)</td>
<td></td>
<td>*</td>
<td>*</td>
<td></td>
<td>*</td>
<td>*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Edge density (ED)</td>
<td></td>
<td>*</td>
<td>*</td>
<td></td>
<td>*</td>
<td>*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Largest patch index (LPI)</td>
<td></td>
<td>*</td>
<td>*</td>
<td></td>
<td>*</td>
<td>*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Percentage of landscape (PLAND)</td>
<td></td>
<td>*</td>
<td>*</td>
<td></td>
<td>*</td>
<td>*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Landscape shape index (LSI)</td>
<td></td>
<td>*</td>
<td>*</td>
<td></td>
<td>*</td>
<td>*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Euclidean mean nearest neighbour distance (ENN_MN)</td>
<td></td>
<td>*</td>
<td>*</td>
<td></td>
<td>*</td>
<td>*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean patch area (AREA_MN)</td>
<td></td>
<td>*</td>
<td>*</td>
<td></td>
<td>*</td>
<td>*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Area-weighted mean patch area (AREA_AM)</td>
<td></td>
<td>*</td>
<td>*</td>
<td></td>
<td>*</td>
<td>*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Area weighted mean patch fractal dimension (AWMPFD)</td>
<td></td>
<td>*</td>
<td>*</td>
<td></td>
<td>*</td>
<td>*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean patch fractal dimension (FRAC_MN)</td>
<td></td>
<td>*</td>
<td>*</td>
<td></td>
<td>*</td>
<td>*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean patch fractal dimension (MPFD)</td>
<td></td>
<td>*</td>
<td>*</td>
<td></td>
<td>*</td>
<td>*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean perimeter to area ratio (PARA_MN)</td>
<td></td>
<td>*</td>
<td>*</td>
<td></td>
<td>*</td>
<td>*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Core area index (CAI)</td>
<td></td>
<td>*</td>
<td>*</td>
<td></td>
<td>*</td>
<td>*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Shannon's diversity index (SHDI)</td>
<td></td>
<td>*</td>
<td>*</td>
<td></td>
<td>*</td>
<td>*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Shannon evenness index (SHI)</td>
<td></td>
<td>*</td>
<td>*</td>
<td></td>
<td>*</td>
<td>*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean patch shape index (MSI)</td>
<td></td>
<td>*</td>
<td>*</td>
<td></td>
<td>*</td>
<td>*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Area-weighted mean shape index (AWMSI)</td>
<td></td>
<td>*</td>
<td>*</td>
<td></td>
<td>*</td>
<td>*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Clumpiness index (CLUMP)</td>
<td></td>
<td>*</td>
<td>*</td>
<td></td>
<td>*</td>
<td>*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Contagion index (CONTAG)</td>
<td></td>
<td>*</td>
<td>*</td>
<td></td>
<td>*</td>
<td>*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Connectance (CONNECT)</td>
<td></td>
<td>*</td>
<td>*</td>
<td></td>
<td>*</td>
<td>*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Interspersion and juxtaposition index (IJ)</td>
<td></td>
<td>*</td>
<td>*</td>
<td></td>
<td>*</td>
<td>*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Aggregation index (AI)</td>
<td></td>
<td>*</td>
<td>*</td>
<td></td>
<td>*</td>
<td>*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note: * indicates implemented metrics in the study.
where, \( n_i \) indicates number of patches in the landscape of patch class \( i \).

**CA metric** is a measure of landscape composition. It provides a basis for calculating many of the class- and landscape-based indices. CA is most effective when it is applied in conjunction with one or more configuration metrics in a complementary manner. For example, when CA is applied with mean patch size (AREA_MN) and/or patch density (PD) the overall extent and subdivision of each LULC category can be better quantified (Botequila et al. 2006). This metric indicates the sum of the areas (m\(^2\)) of all patches of the corresponding patch type, divided by 10,000 to convert into hectares. CA approaches 0 as the patch type becomes increasing rare in the landscape. This measure is able to describe to what extent the model can simulate real area of specific category (e.g. built-up areas). The equation is (Mc Garigal & Marks 1995):

\[
CA = \sum_{j=1}^{n} a_{ij} \left( \frac{1}{10000} \right)
\]

where, \( a_{ij} \) is area (m\(^2\)) of patch \( ij \).

**AREA_MN metric** addresses mean size of the patches (hectares) at both landscape and class levels (\( \geq 0 \), without limit). This is one of the most important pieces of information that an analyst can acquire about a landscape. AREA_MN is a fairly intuitive measure of landscape structure. It has many potential applications so that may serve as rough indicators of landscape function in combination with other landscape metrics (Botequila et al. 2006). This metric is of significant potential for assessing the model performance in terms of real size of patches belonging to each category, especially highly dynamic ones. The formula is as follows (Mc Garigal & Marks 1995):

\[
AREA_MN = \frac{\sum_{j=1}^{m} a_j}{m}
\]

where, \( a_j \) is the size of patch \( j \) and \( m \) is the total number of patch types (classes).

**ENN_MN metric** is mean shortest distance among patches that share the same category. This is a measure of patch context and patch isolation and approaches 0 as distance to the nearest neighbour decreases. ENN_MN provides a means for an analyst to mathematically describe distribution of LULC categories across a landscape (Botequila et al. 2006). ENN_MN is an appropriate indication of the model’s ability in explaining isolation of patches belonging to a category form their nearest neighbours, and thereby, mirrors model performance in terms of simulating the patch locations and their distance from each other. This measure can clearly reflect spatial variation of a dynamic class such as built-up areas, which is very probable to fill gaps in locations that had already been discontinuous. The formula is as follows:

---

**Table 5.** Descriptions of landscape metrics used in the study (Mc Garigal & Marks 1995).

<table>
<thead>
<tr>
<th>Concept</th>
<th>Type of metric</th>
<th>Range</th>
<th>Units</th>
<th>Abbreviation</th>
<th>Landscape metrics</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fragmentation</td>
<td>Landscape composition</td>
<td>( NP &gt; 0 ), without limit</td>
<td>None</td>
<td>NP</td>
<td>Number of patches</td>
</tr>
<tr>
<td>Fragmentation, Area</td>
<td>Landscape configuration</td>
<td>AREA_MN &gt; 0, without limit</td>
<td>Hectare</td>
<td>AREA_MN</td>
<td>Mean patch area</td>
</tr>
<tr>
<td>Area</td>
<td>Landscape composition</td>
<td>( CA &gt; 0 ), without limit</td>
<td>Hectare</td>
<td>CA</td>
<td>Total class area</td>
</tr>
<tr>
<td>Isolation</td>
<td>Landscape configuration</td>
<td>ENN &gt; 0, without limit</td>
<td>Metres</td>
<td>ENN_MN</td>
<td>Mean Euclidian nearest neighbour Distance</td>
</tr>
<tr>
<td>Shape, aggregation index</td>
<td>Landscape configuration</td>
<td>( LSI \geq 1 ), without limit</td>
<td>None</td>
<td>LSI</td>
<td>Landscape shape index</td>
</tr>
</tbody>
</table>
where, $h_{ij}$ is distance (m) from patch $ij$ to the nearest neighbouring patch of the same type (class), based on patch edge-to-edge distance, computed from cell centre to cell centre (McGarigal & Marks 1995).

LSI metric provides a standardised measure of total edge or edge density that adjusts for the size of the landscape (≥1, without limit). LSI metric is equal to 1 when the landscape consists of a single square (or almost square) patch; LSI increases without limit as landscape shape becomes more irregular and/or as the length of the edge within the landscape increases. This metric is indicative of agreement between the simulated shape of the patches within a category and their correspondence category in the reference map. The shape of LULC patches is an important feature in determining how they affect ecological characteristics of the landscape. Patch shape strongly influences magnitude and nature of interaction of a patch with its surrounding neighbourhood, principally via edge effects and cross-boundary processes (Botequila et al. 2006). The formula is (McGarigal & Marks 1995):

$$LSI = \frac{e_i}{\min e_i}$$

where $e_i$ denotes the perimeter of category $i$ and $\min e_i$ refers minimum perimeter of class $i$.

3. Results and discussion

3.1 Kappa and overall accuracy indices-based performance evaluation

Kappa and overall accuracy indices for the entire landscape are presented in Table 6. The overall accuracy and kappa index were accounted for 94.34% and 0.90, respectively, which denote that the simulated LULC map has high level of consistency with the reference data. Forest and wetland classes received the highest kappa value followed by agricultural fields, built-up areas, barren lands and water bodies, accounting for 0.99, 0.99, 0.90, 0.87, 0.86 and 0.74, respectively. Lower kappa value of model performance for water category is attributable to regular geometric shape of water reservoirs and linear distribution of river patches in the landscape that make this category more sensitive to quantity-based measures of model performance and tend to decrease more dramatically.

As given in Table 6, there is high degree of consistency between the simulated LULC vs. the reference LULC map regarding percent of the correctly simulated pixels.

3.2 Landscape metrics-based performance evaluation

For providing information on the landscape pattern and shape of the LULC maps, landscape metrics were calculated for both the simulated and the reference LULC layers.

3.2.1 Class metrics-based performance evaluation

Figure 6 (a, b, c, d and e) demonstrates differences between the simulated and the reference maps in terms of spatial arrangement of LULC categories. In addition, Tables 7, 8 and 9 presents detailed information on the RE of the model calibration.

According to Table 7, there are negative (underestimated) and positive (overestimated) RE values for each metric of LULC categories. In this case, the model tends to overestimate NP values for barren, agriculture and wetland categories (RE = 53.65, 31.90 and 25.00%, respectively). In contrast, the
The model tends to produce negative NP values for built-up, water and forest land features (RE = −31.00, −13.60 and −3.13%, respectively). Level of agreement for the simulated and the observed data are

**Figure 6.** Differences between the simulated and reference LULC classes (1: Built-up, 2: Agriculture, 3: Forest, 4: Water, 5: Barren land and 6: Wetland) regarding (a) NP (Number of Patches), (b) AREA_MN (Mean Patch Area/hectare), (c) CA (Class Area/hectare), (d) LSI (Landscape Shape Index) and (e) ENN_MN (Mean Euclidean Nearest Neighbour Distance/metre) metrics values.

**Table 7.** Relative error and mean relative error values for each LULC category and landscape metric at class level.

<table>
<thead>
<tr>
<th>Metric/LULC category</th>
<th>NP</th>
<th>AREA_MN</th>
<th>CA</th>
<th>ENN_MN</th>
<th>LSI</th>
<th>MRE_{LULC} (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Built-up</td>
<td>31.00</td>
<td>33.00</td>
<td>−8.26</td>
<td>14.55</td>
<td>−15.00</td>
<td>20.36</td>
</tr>
<tr>
<td>Agriculture</td>
<td>−31.90</td>
<td>−25.81</td>
<td>−2.15</td>
<td>78.30</td>
<td>−13.19</td>
<td>30.27</td>
</tr>
<tr>
<td>Forest</td>
<td>−3.13</td>
<td>3.89</td>
<td>0.65</td>
<td>1.81</td>
<td>−6.42</td>
<td>3.18</td>
</tr>
<tr>
<td>Water</td>
<td>−13.60</td>
<td>−11.57</td>
<td>−23.00</td>
<td>4.16</td>
<td>−15.00</td>
<td>13.46</td>
</tr>
<tr>
<td>Barren land</td>
<td>53.65</td>
<td>51.03</td>
<td>132.07</td>
<td>−20.30</td>
<td>−4.00</td>
<td>52.21</td>
</tr>
<tr>
<td>Wetland</td>
<td>25.00</td>
<td>25.00</td>
<td>2.10</td>
<td>−30.21</td>
<td>9.54</td>
<td>18.37</td>
</tr>
<tr>
<td>MRE_{metric} (%)</td>
<td>26.38</td>
<td>25.05</td>
<td>28.03</td>
<td>23.22</td>
<td>10.52</td>
<td>MRE_{class level} = 22.73%</td>
</tr>
</tbody>
</table>

Downloaded by [Yousef Sakieh] at 08:25 08 April 2016
judged to be high (forest and water), good (wetland), average (built-up and agriculture), and Low (barren), respectively.

In terms of AREA_MN metric, the model generates overestimated results for barren, built-up, wetland and forest categories (RE = 51.03, 33.00, 25.00 and 3.89%, respectively) and by way of contrast there are negative values for agriculture and water land parameters (RE = −25.81 and −11.57%). Comparing the simulated and reference maps, forest and water categories represented high, wetland and agriculture categories indicated good and built-up areas demonstrated average level of agreement.

In case of the CA metric, there was a considerable overestimated value for barren category, which demonstrated 132.07% of RE value, followed by wetland and forest categories (RE = 2.10 and 0.65%) Conversely, there were underestimated results for water, built-up and agriculture land elements with RE values of −23.00, −8.26 and −2.15%, respectively. Degree of similarity regarding CA metric for forest, built-up areas, agriculture, water bodies, wetland and barren lands were high, high and low, respectively.

For ENN_MN metric, there are underestimated results for wetland and barren categories (RE = −30.21 and −20.30%). Agriculture, built-up, water and forest classes depicted exaggerated RE values (78.30, 14.55, 4.16 and 1.81%, respectively). Model goodness-of-fit for water, forest and built-up areas indicated excellent performance, and in terms of barren and wetland categories, good and moderate performances were recorded followed by weak model behaviour for the agricultural fields.

Regarding LSI metric, the model simulation outputs depicted an over-predicted result for wetland category with RE value of 9.54%. In contrary, there are underestimated results for built-up, water, agriculture, forest and barren land features (RE = −15.00, −15.00, −13.19, −6.42 and −4.00%, respectively.

### Table 8. Level of agreement between the simulated and reference LULC maps for all categories at class level.

<table>
<thead>
<tr>
<th>Metric/category</th>
<th>NP</th>
<th>AREA_MN</th>
<th>CA</th>
<th>ENN_MN</th>
<th>LSI</th>
<th>Agreement</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Built-up</td>
<td>Average</td>
<td>Average</td>
<td>High</td>
<td>High</td>
<td>High</td>
<td>Good</td>
<td>Average</td>
</tr>
<tr>
<td>Agriculture</td>
<td>Average</td>
<td>Good</td>
<td>High</td>
<td>High</td>
<td>Low</td>
<td>Good</td>
<td>Average</td>
</tr>
<tr>
<td>Forest</td>
<td>High</td>
<td>High</td>
<td>High</td>
<td>High</td>
<td>High</td>
<td>Average</td>
<td>The highest agreement</td>
</tr>
<tr>
<td>Water</td>
<td>High</td>
<td>Low</td>
<td>Good</td>
<td>High</td>
<td>High</td>
<td>High</td>
<td>High</td>
</tr>
<tr>
<td>Barren land</td>
<td>Low</td>
<td>Low</td>
<td>Low</td>
<td>Good</td>
<td>Good</td>
<td>Good</td>
<td>The lowest agreement</td>
</tr>
<tr>
<td>Wetland</td>
<td>Good</td>
<td>Good</td>
<td>High</td>
<td>Good</td>
<td>Average</td>
<td>Good</td>
<td>The highest agreement</td>
</tr>
</tbody>
</table>

### Table 9. Level of model performance for all LULC categories and landscape metrics estimations at class level.

<table>
<thead>
<tr>
<th>Metric/category</th>
<th>NP</th>
<th>AREA_MN</th>
<th>CA</th>
<th>ENN_MN</th>
<th>LSI</th>
<th>Model Performance</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Built-up</td>
<td>Moderate</td>
<td>Moderate</td>
<td>Excellent</td>
<td>Excellent</td>
<td>Excellent</td>
<td>Good</td>
<td>The highest performance</td>
</tr>
<tr>
<td>Agriculture</td>
<td>Moderate</td>
<td>Good</td>
<td>Excellent</td>
<td>Weak</td>
<td>Excellent</td>
<td>Moderate</td>
<td>The lowest performance</td>
</tr>
<tr>
<td>Forest</td>
<td>Excellent</td>
<td>Excellent</td>
<td>Good</td>
<td>Good</td>
<td>Excellent</td>
<td>Excellent</td>
<td>The highest performance</td>
</tr>
<tr>
<td>Water</td>
<td>Excellent</td>
<td>Weak</td>
<td>Excellent</td>
<td>Good</td>
<td>Excellent</td>
<td>Good</td>
<td>The highest performance</td>
</tr>
<tr>
<td>Barren land</td>
<td>Good</td>
<td>Good</td>
<td>Excellent</td>
<td>Moderate</td>
<td>Excellent</td>
<td>Good</td>
<td>The lowest performance</td>
</tr>
<tr>
<td>Wetland</td>
<td>Good</td>
<td>Good</td>
<td>Excellent</td>
<td>Good</td>
<td>Excellent</td>
<td>Good</td>
<td>The highest performance</td>
</tr>
</tbody>
</table>

Downloaded by [Yousef Sakieh] at 08:25 08 April 2016
respectively). Considering LSI metric, the model performance was evaluated as *excellent* in terms of all LULC categories.

On the whole, taking the set of five metrics into account, the model performance was better than others in simulating forest lands. MRE value for this category was scored at 3.18% for the total set of the landscape metrics. In this point, LSI metric with −6.42% and CA metric with 0.65% variation depicted the lowest and the highest agreements vs. the reference LULC layer.

Except for LSI and ENN_MN metrics, the model performance was least successful in terms of barren category, which depicted low agreement referring to the rest of the calculated metrics. MRE recorded at 52.21% for barren category which indicates the simulation process was least successful in terms of the spatial pattern of the LULC categories.

Generally, MRE value derived from class level metrics measured at 22.73% that reflects the good performance of the model.

### 3.2.2 Landscape level metrics

Figure 7 and Table 10 represent variations between the calculated metrics and RE values for model results, referring to the simulated and reference layers at landscape level. The LSI index indicated the greatest deviation between model outputs and corresponding categories in the reference map (under-estimated result with RE value of −13.05%). On the other hand, ENN_MN metric depicted the lowest variation among others with 6.4% difference from its actual value. According to the results, there is a high agreement between the simulated and the reference maps for the total number of the metrics at landscape level. The MRE parameter estimated at 10.12% at landscape level, which mirrors the fact that model has an excellent performance through the simulation process.

Applying quantity-based indicators of simulation success alone cannot inform the LULC modellers and practitioners on model performance in terms of spatial metrics and morphological attributes. In fact, the spatially-explicit predictive models require the spatially-explicit performance evaluation methods which provide additional level of knowledge on model behaviour. Accordingly this study proposed a schema for model performance evaluation based on landscape pattern analysis. The unique characteristic of the proposed methodology is that the landscape metrics established a basis for hierarchical and multi spatial level analysis of the model behaviour, which allows quantifying simulation success in terms of landscape structure and spatial patterns.
By comparing the resultant MRE values at class and landscape levels (22.73 and 10.12%, respectively), it can be drawn that there are considerable differences between model performances at these two levels. According to Tables 7 and 10, the final MRE value was within the acceptable limits (less than 15%) at landscape level (Congalton & Green 2009). However, MRE values were approximately increased two- or three-fold at class level (especially in case of NP, AREA_MN and ENN metrics). This is an important finding which substantiates a basis for simultaneous analysis of model performance at multiple levels. This may be explained by calculation method and more incorporated information at the class level analysis of landscape pattern.

In this study, an integrated CA-Markov chain model was calibrated applying historical data sets of the years 1987 and 1999. Model calibration takes advantage from both Markov chain process for calculating transition matrix and CA components. The simulation success of the calibrated model was evaluated by a series of landscape metrics. The 1999 LULC map was considered as initial seed year and the 2011 LULC map was simulated using 12 CA iterations. According to the results, the model performance was the best one for forest category and the LULC change model simulates the spatial pattern of this category quite similar to the reality. One contributing factor in model success for simulating forest patches is that this class has experienced insignificant amount of area change (6.16%) compared to others. In the case of forest category, the results were consistent with findings of Guan et al. (2011).

On the contrary, barren lands showed the sharpest differences between the simulated and the reference maps regarding the metrics’ values at both class and landscape levels. Therefore, the model tended to simulate more fragmented landscape. Maybe it could be attributable to the fact that this category had the most portion of mutual conversion exchange with other classes. Namely, 51.89% of the barren lands were transformed into other classes (between 1987 and 2011), which means that barren areas are highly dynamic and thereby difficult to be modelled in this study site (Figure 5). Referring to the metrics’ values for built-up areas, the simulated results indicated that CA-Markov chain model tended to simulate urban areas as being too compact (Figure 5). Herold et al. (2005) and Wu et al. (2009) reported that the SLEUTH model has a similar behaviour in replicating spatial pattern of built-up areas. In contrast to the built-up areas, the model tended to simulate more fragmented landscape for agricultural fields referring to NP and ENN_MN metrics, which is in accordance with the findings of Mas et al. (2010b). Regardless of high accuracy obtained from kappa and overall accuracy indices, the spatial pattern of water bodies is not properly captured by the model.

Generally, simulation success of the model can be summarised into following items:

1. there is a tendency in which the model simulates higher number of patches (NP) for agriculture, barren and wetland and a compact landscape for built-up category;
2. patch shape (LSI) in the simulated results became simpler and bigger in size and patches became farther in distance from each other;
3. integrating conclusions 1 and 2, model outputs are simplistic against the observed spatial pattern of reference LULC map;
4. according to the results (Tables 7 and 10), performance of CA-Markov chain model was more successful at landscape level (excellent performance) than class level (good performance);

<table>
<thead>
<tr>
<th>Metric/value</th>
<th>Reference map</th>
<th>Simulated map</th>
<th>Relative error</th>
<th>Level of agreement</th>
<th>Model performance</th>
</tr>
</thead>
<tbody>
<tr>
<td>NP</td>
<td>1626.00</td>
<td>1469</td>
<td>−9.96</td>
<td>High</td>
<td>Excellent</td>
</tr>
<tr>
<td>AREA_MN</td>
<td>74.82</td>
<td>83.10</td>
<td>11.06</td>
<td>High</td>
<td>Excellent</td>
</tr>
<tr>
<td>ENN_MN</td>
<td>318.52</td>
<td>338.90</td>
<td>6.40</td>
<td>High</td>
<td>Excellent</td>
</tr>
<tr>
<td>LSI</td>
<td>26.08</td>
<td>22.68</td>
<td>−13.05</td>
<td>High</td>
<td>Excellent</td>
</tr>
<tr>
<td>MRE_landscape level</td>
<td>10.12</td>
<td>10.12</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 10. Metrics value, relative error, mean relative error of model simulation results and level of agreement for reference and simulated LULC maps at landscape level.
(5) synoptic consideration of the results reveals that the model performed well in simulating spatial pattern of LULC classes.

There are over hundred statistical indices of landscape structure at both levels of land-use category and landscape (Mc Garigal et al. 2002). It is important to comprehend the empirical linkages among these measures so that a wise combination of the metrics for model performance evaluation can be decided. It is therefore helpful to decide on the smallest and more consistent set of the metrics (i.e. parsimony in landscape metrics) to identify a suite of structure elements that together effectively describe major independent aspects of landscape structure (Cushman et al. 2008).

As illustrated in this study, landscape metrics are useful tools to distinguish model performance in simulating a landscape structure at landscape and class levels. Class-level metrics characterise model performance in terms of patch size, shape and neighbourhood of a land-use category. Landscape-level analyses illustrate model’s simulation success for spatial structure of patch mosaics with multiple categories, which finally provide information on overall landscape heterogeneity, texture or graininess (Cushman et al. 2008). In addition, at class-level analysis it is of importance to distinguish between composition and configuration because they are conceptually different dimensions of a landscape structure (Fahrig 2002).

It is of value to note that inclusion of several correlated metrics might increase redundancy and produce exaggerated results. Likewise, failure to select metrics that address distinct aspect of the landscape structure will reduce dimensionality of landscape structure (Cushman et al. 2008). Choice of specific sets of the metrics would be based on the research questions being analysed and on known metric behaviour (Neel et al. 2004). It is important to note that our list of the metrics would not be sufficient to capture the full characteristics of any particular study location. Rather, it is possible that in other study areas these metrics due to their universality feature might consistently be present, but unique feature of a landscape structure may also be present that necessitate their specific metrics to be selected.

Categorical LULC maps inherently suffer from uncertainty. Thus, the derived landscape metrics are sensitive to LULC misclassification and inherits uncertainties (Hess 1994; Wickham et al. 1997; Mas et al. 2010a). According to Lechner et al. (2012) there are three major sources that introduce uncertainty in spatial analyses including (1) classification scheme uncertainty, (2) spatial scale and (3) classification error.

Applying discrete classification procedures in ecological studies can yield in spatial uncertainty as there are no certain borders between LULC categories. Therefore, categorization of the landscape can be based on immeasurable ways (Burrough 1996; Arnot et al. 2004; Schmit et al. 2006) and borders between land features inherit subjectivism (Chapman et al. 2005). In addition, imprecise definition of LULC categories is another source of uncertainty in classification scheme. Differences in classification schemes will influence the result of satellite images classification.

Spatial scale is another source of uncertainty in landscape analyses. There are five scale dependent factors that play a significant role in characterisation of landscape pattern and pose a considerable influence on landscape metrics calculation (Lechner et al. 2012): (1) pixel size, (2) MMU, (3) smoothing, (4) thematic resolution and (5) spatial extent.

There are several options to analyse spatial uncertainty such as identifying and outlining scale dependent factors and sensitivity analysis (Lechner et al. 2012). Studies that undertake spatial analysis should include an explanation regarding scale dependent factors related to the data (e.g. MMU, extent, etc.) and the description of the scale at which the environmental phenomenon operates through the landscape. In addition, the efficiency of classification scheme and explicit explanation of all scale dependent factors can also significantly contribute in addressing uncertainty.

There are also some considerations to address the effect of uncertainty including employment of finer spatial resolution than the phenomena of interest, using a larger extent than the targeted ecological process and a MMU that removes patches too small to be related to the studied question. These considerations, along with more accurate classification procedures (e.g. hybrid method) and a
classification scheme with specificity to the studied ecological phenomenon can decrease some of the probable sources of spatial uncertainty.

Accordingly, spatial resolution of the study was decided to be 30 m (Landsat image cell size). This choice is on the basis that data at finer resolutions might unnecessarily increase data size and data at coarser levels can cause information loss. Thematic resolution is based on the purpose and scale of the study, which follows first level of Anderson (1976) classification scheme. In addition, considering spatial resolution of the data used for LULC classification prevented us to select a finer thematic resolution. Regarding scale of the study (1:25,000), MMU was decided to be 0.4 hectare (133 pixels), since mapping units smaller than 2.5 × 2.5 mm were too small to be visually detected and digitised. Using hybrid classification method, LULC borders were carefully digitised and revised for various node years mainly to produce highly accurate LULC maps and to support landscape metrics calculations. Finally, a statement of accuracy for LULC map of 2011 is provided to quantitatively address the spatial accuracy of the classification process.

4. Conclusion

The processes and functions of a landscape are influenced by changing its structure, composition and configuration (Forman & Godron 1986; Turner et al. 2001; Botequila et al. 2006; Farina 2006). Geospatial LULC change models as innovative planning tools are of great interest for analysing, modelling and understanding of landscape structure and functions. Accordingly, performance evaluation of such models needs to incorporate fundamental attributes of landscape structure, which provides further information on simulation behaviour of LULC change model. Many landscape metrics have been developed and implemented in different research areas during the last decades; however, a limited number of studies have applied landscape metrics (without technical details) as spatial indicators of model performance. The current paper adopted a different approach to evaluate the performance of a CA-Markov chain model, which is reflected in resultant RE values of metrics calculations. According to the results, the landscape metrics employed in this study indicated considerable potential to measure simulation success in terms of the following items:

(1) Considering spatial details at multiple levels and (2) Considering morphological and structural features of LULC patches such as size, number, distance, shape and complexity. The first refers to unique ability of the metrics in hierarchical and multilevel performance evaluation compared to simple conventional methods such as kappa coefficient and overall accuracy. This method has potential to simultaneously quantify simulation success at three levels including a specific patch, class and landscape. The later addresses the method’s ability in providing information on simulation accuracy of each LULC category in terms of neighbourhood characteristics and spatial patterns. Thus, the LULC modellers and practitioners can compare different models in a spatially-explicit manner and decide on best-performing simulation method in terms of replicating a landscape structure. This property can be interesting for ecologists that tend to measure the effect of structural changes on a specific landscape function (e.g. carbon sequestration, nutrient cycling, etc.).

According to Cushman et al. (2008) landscape metrics possess attributes such as universality, strength and consistency. This implies on a fact that landscape structures in different regions share similar characteristics including size, neighbourhood similarity, fragmentation and dominance, which can be studied by the same set of the metrics. Synoptic consideration of the results obtained from our literature review and those derived from Cushman et al. (2008) demonstrates that implemented metrics in this study have frequently been used in different study locations with significant contribution for analysing the structure of the landscape. It is also reported that those landscape metrics with low universality are unique dimension of landscape structure that do not exist in majority of land-use categories and regions. Accordingly, it is possible to declare that there seem to be both universal and idiosyncratic features of landscape structure. Therefore, other spatial simulation models in different study areas would require different set of the metrics. For instance, land allocation models need
different sets of the metrics compared to more specific studies such as designing habitat networks in protected areas.

It should be noted that the performance of the CA-Markov chain model was evaluated under a historical scenario. In other word, this study assumed that urban growth profile between 1987 and 1999 is expected to continue in the subsequent time period (1999–2011). As a topic of further research, the effect of scenario definition on model performance can also be evaluated by using landscape metrics. Some model parameters including conditional images, CA transition rules and transition areas matrix can be modified in prediction phase of the model run (1999–2011). This enables the user to introduce additional alternatives such as accelerated urbanisation or compact urban growth in a specific study. The effect of each alternative on model performance can be evaluated based on the recommended procedure in this paper. Therefore, those alternatives with better model performance in terms of landscape metrics can be implemented to more realistically project urban growth trajectory into the uncertain future.

Finally, the following subjects merit further research for testing, developing and improving the proposed approach in this paper:

1. For selecting an optimum and best-descriptive set of the metrics, it is necessary to measure and compare simulation success of diverse spatial algorithms through various metrics in different study areas.

2. Finally, the following questions are of interest to get explored in future studies; what are the effects of changing spatial scale (grain size and extent) (Turner et al. 1989; Wu et al. 2002; Wu 2004) on land change simulation models performance and how they can be quantified using landscape metrics?

Disclosure statement

No potential conflict of interest was reported by the authors.

ORCID

Bahman Jabbarian Amiri http://orcid.org/0000-0003-1382-786X
Yousef Sakieh http://orcid.org/0000-0003-1814-7783

References


Pontius RG Jr, Batchu K. 2003. Using the relative operating characteristic to quantify certainty in prediction of location of land cover change in India. Trans GIS. 7:467–484.


