RESEARCH ARTICLE

Using the T-EFA method in a cellular automata-based urban growth simulation's calibration step

Ismail Ercument Ayazli[1] [iD] | Ahmet Emir Yakup[2] [iD] | Omer Bilen[3] [iD]

[1]Department of Geomatics Engineering, Sivas Cumhuriyet University, Sivas, Turkey

[2]Department of Architecture and Urban Planning, Hitit University, Corum, Turkey

[3]Department of Urban and Regional Planning, Bursa Technical University, Bursa, Turkey

Correspondence

Ismail Ercument Ayazli, Department of Geomatics Engineering, Sivas Cumhuriyet University, Sivas, Turkey.
Email: eayazli@cumhuriyet.edu.tr

Abstract

Changes in land cover driven by urban sprawl increase the threat of urbanization of forests and agricultural lands. Therefore, monitoring urban sprawl by creating simulation models is frequently carried out to understand sustainable city management. Cellular automata-based models are mostly preferred to reduce the damage led by urban sprawl, and the SLEUTH model is the most well known. Several methods have been developed for the SLEUTH model calibration step, such as optimum SLEUTH metrics and total exploratory factor analysis (T-EFA), to improve the model accuracy. This study aims to create a high-accuracy urban growth simulation model using low-resolution data, investigate the T-EFA method's success in the calibration step, and find the urban sprawl effects on land cover change. Istanbul was selected as our study area due to witnessing its tremendous urban sprawl since the 1950s. According to our results, the urban growth that occurred between 2000 and 2018 could be defined more closely to reality using the T-EFA method, and Istanbul will continue to grow until 2040, with approximately 428.7 km^2 of agricultural lands, 553.4 km^2 of forests, and 0.1 km^2 of wetlands being transformed to urban. In addition, the geologically risky areas under threat of urbanization will increase by 60% between 2018 and 2040.

1 | INTRODUCTION

The world population increases by an average of 1.2% every year (Alkan, Ozulu, Ilci, Tombus, & Sahin, 2017). Half of this population lives in cities with a higher job, education, health opportunity, and welfare level, and the urban population is expected to reach 68% in the world in 2050 (WUP, 2019). In addition to population growth, rent-oriented land-use policies also cause the urban space to expand physically and are defined as urban growth. Urban growth is a complex socioeconomic process that transforms the built environment and rural areas into urban settlements with increasing population, while shifting the spatial distribution of the population from rural areas to urban areas (UN, 2019). Urban growth has three components: natural population growth, migration, and reclassification (Stecklov, 2018). In addition to these, urban growth depends on the demographic changes in a country (Dyson, 2011), the size of the settlement unit (Batty, 2008), spatial planning policies (Angel, Parent, Civco, & Blei, 2011), and physical and local conditions specific to the urban area. When cities grow, they merge their populations and neighboring settlements that have been previously classified as rural areas (Siri, Brown, & Spielauer, 2010). The population growth in rural areas may cause settlements to be reclassified from rural to urban, thus accelerating urbanization (UN, 2019).

Generally, environmental, social, and economic problems are faced if urban growth, which causes natural areas and open spaces on urban fringes to be transformed into built residential areas, industrial areas, and trade centers, is not taken under control (EEA, 2006, 2016).

Land use/land cover (LULC) changes lead to the main environmental problems. Climate change as a result of deforestation and the destruction of agricultural lands and wetlands due to LULC changes causes problems such as an increase in land surface temperatures, global warming, drought and floods, epidemics, air and water pollution, lack of nutrition, and food insecurity (Bart, 2010; EEA, 2016; Emadodin, Reinsch, & Taube, 2019; Khorrami, Gunduz, Patel, Ghouzlane, & Najjar, 2019; Kusak & Kucukali, 2018; Loh et al., 2015; Selim & Demir, 2019; Tu, Xia, Clarke, & Frei, 2007). Moreover, the transformation of lands with natural disaster risk into settlements due to uncontrollable urban growth (Mestav Sarica, Zhu, & Pan, 2020), and the dimensions of social and economic losses caused by a possible disaster, are very high. In addition to these, increasing costs, especially of energy and transportation, cause the disruption of public services provided by local governments and an increase in poverty along with a change in demography and quality of life; these problems are addressed under the title of socioeconomic problems (Ayazli, Kilic, Lauf, Demir, & Kleinschmit, 2015; EEA, 2016).

Various models have been developed to minimize the damage of urban growth. The first-produced urban growth model theories are usually of radial form. After the 20th century, the approach in which cities consist of many dynamic subsystems has been accepted (Ayazli et al., 2015; Batty, 2007; Benenson & Torrens, 2004; Junfeng, 2003). In parallel with developments in computer technologies, it is possible to model cities' complex and dynamic structures nowadays. In line with this, simulation techniques are used to determine and model the behavior of a complex and dynamic system (Batty, 2007; Benenson & Torrens, 2004). In this context, cellular automata (CA)-based simulation methods are frequently used to model urban growth and LULC changes (Chaudhuri & Clarke, 2013; Clarke & Gaydos, 1998; Clewlow, 1989; Dennunzio, Formenti, & Kurka, 2012; Di Lena & Margara, 2008; Wang et al., 2021).

CA consists of five essential components: grid network, state, neighborhood, transition rules, and time. The grid network, which can be one or two-dimensional, is used to describe the space; each cell contains an automaton, and this cell contains the state of the automaton (Torrens, 2000). Transition rules determine the change in the states of automata over time according to the state of neighboring cells (Benenson & Torrens, 2004; Torrens, 2000). CA is the most straightforward representation of dynamic and spatial systems, and has high computational capability (White, Straatman, & Engelen, 2004). Therefore, it is frequently used to model LULC changes. In urban growth studies, its functionality should be increased by adding additional components to traditional CA (T-CA) (Torrens, 2000; White & Engelen, 1997; White et al., 2004). Cell states are divided into fixed or functional in this method, called modified CA (M-CA). Cells without urbanization potential are considered

fixed, while others are considered functional (Clarke, Hoppen, & Gaydos, 1997; Torrens, 2000; White et al., 2004). Neighborhood relationships are also expanded to include wider cells in M-CA (White & Engelen, 1997; White et al., 2004). Another difference between T-CA and M-CA is in the time component. Cell update in M-CA is asynchronous (Torrens, 2000).

Although CA is a proper modeling technique in urban sprawl studies, predefined transition rules might decrease modeling achievement because of the re-use of CA (Ayazli, 2019). Moreover, some projects may ignore quantitative exploration, even though they investigate qualitative analyses. Therefore, the correct specification of the growth features, which is carried out by calibration, has an important impact on the prediction accuracy of the simulation (Pontius & Malanson, 2005). The purpose of calibration is to minimize the differences between modeled and actual LULC, and the mathematical background of the calibration is formed by the objective function (Feng & Tong, 2018). For this reason, the calibration step that directly affects the model accuracy and makes the produced model reliable is the most crucial stage, and various calibration techniques have been developed, such as statistical or machine learning algorithms (Feng & Tong, 2018; Huang & Zhang, 2014; Liu, Li, Shi, Huang, & Liu, 2012). The other methods of determining the objective functions include simulated annealing, generalized pattern search, swarm intelligence, and evolutionary algorithms (Feng & Tong, 2018).

Many M-CA-based simulation models have been developed to monitor urban growth. The SLEUTH urban growth model, developed by Keith Clarke, is one of these models and is commonly used in academic studies since it is open-source and free software (Bihamta, Soffianian, Fakheran, & Gholamalifard, 2015; Dezhkam, Jabbarian Amiri, Darvishsefat, & Sakieh, 2013; Han, Hwang, Ha, & Byung Sik, 2015; Jantz, Goetz, & Shelley, 2004). After parameters are set in scenario files, simulation starts with the test in which the data used are first checked, continues with calibration (which is the most critical step affecting the model's accuracy), and ends with prediction (Clarke, Hoppen, & Gaydos, 1996; Silva & Clarke, 2002). Using the brute force calibration (BFC) method, the calibration stage aims to calculate the best-fit values of dispersion, breed, spread, road gravity, and slope coefficients. These calculations are made using the BFC method by applying the growth rules of "Spontaneous, New Spreading Center, Edge Growth, and Road-Influenced Growth." The primary purpose of the calibration stage is to calculate the best-fit value for each growth coefficient at the end of calibration by narrowing the five coefficient values between 0 and 100 in the coarse calibration step according to 13 spatial metrics. However, there is no consensus on which one of the 13 regression (R^2) scores to use during model calibration.

First, statistical and visual tests were used to calibrate the model tested in the San Francisco Bay region (Clarke et al., 1997). In the visual comparison, the R^2 values of "the urban area, the number of edge pixels, and the number of pixel clusters" criteria were calculated by control years, and the Lee–Sallee shape index, comparing the modeled and predicted growth, was used at the statistical stage (Clarke et al., 1996, 1997). Candau (2000) calibrated the model using the R^2 values of six criteria (compare, population, edge, cluster, mean cluster size, and Lee–Sallee) while examining urban growth in Santa Barbara (Candau, 2000). In the following years, the ongoing studies in San Francisco and Washington/Baltimore calculated 12 criteria for the urban growth model calibration, and four different statistical tests were applied, including the modified Lee–Sallee shape index (Clarke & Gaydos, 1998). The model was calibrated with a new criterion using the regression value calculated to compare the pixels representing LULC in Columbia with seven states on the east coast of the USA (Candau & Clarke, 2000), and the number of parameters increased to 13.

The SLEUTH model was first tested in Porto and Lisbon in Europe, and the calibration stage was completed with 13 parameters (Silva & Clarke, 2002). Model calibrations used different combinations of these 13 parameters in studies conducted to date, and it is observed that the Lee–Sallee parameter is used as the primary criterion (Ayazli et al., 2015; Jantz et al., 2004; Sevik, 2006; Silva & Clarke, 2002; Yang & Lo, 2003). In 2007, Dietzel and Clarke developed the optimum SLEUTH metric (OSM) method, in which 7 of the 13 parameters of the SLEUTH model were used ("compare, population, edges, clusters, slope, x-mean, y-mean," f-match if land use was modeled) and claimed that the model would yield the most robust results with this method (Dietzel & Clarke, 2007).

New methods have been developed since the calibration process takes a long time in models created with BFC. Therefore, the self-adaptive genetic algorithm (GA) method was integrated with the model using machine learning algorithms. Although the SLEUTH GA method is five times faster than BFC, BFC is one step ahead in the model's goodness of fit (Saxena & Jat, 2020).

In 2019, Ayazli and Bilen aimed to reduce the number of parameters by grouping the intercorrelated ones among 13 calibration parameters by exploratory factor analysis (EFA). EFA is a dimension reduction technique in which the number of variables is reduced, thus calibrating the model using the regression scores representing the growth form in the study area in the best way (Ayazli & Bilen, 2019). The EFA method aims to calibrate the model using the grouping of intercorrelated parameters. Due to the low Lee–Sallee and OSM scores in the studies conducted in Istanbul, where the population growth is very high (Ayazli et al., 2015; Ayazli, Gul, Baslik, Yakup, & Kotay, 2019), attempts were made to increase the accuracy of the model by calibrating it, with high-resolution data, using the EFA method (Ayazli, 2019).

This study aims to create an urban growth simulation model, calibrating with the total exploratory factor analysis (T-EFA) method, because the OSM and Lee–Sallee scores were low in the previous studies using different resolution data (Ayazli et al., 2015, 2019). For this purpose, we used low-resolution (100 m) CORINE land cover data and found the urban growth impacts in the LULC change. In this context, all 12 spatial metrics representing urban growth were intended to be used with the T-EFA method of model calibration to improve the simulation model's accuracy. In line with this, we ask the following research questions:

- Can an optimum model calibration be performed using all spatial parameters symbolizing urban growth in the SLEUTH model, and what is the success of this calibration?
- Can the T-EFA method be an alternative calibration method for modeling with SLEUTH in rapidly growing cities?

Therefore, the land cover change impacts of urban sprawl were investigated by creating an urban growth simulation model (UGSM), whose calibration stages were completed using T-EFA from low-resolution data obtained from open data sources at short periods in settlement areas with rapid population growth. However, while previous studies have used the highest variance among the number of factors calculated in the EFA method, the factor scores are weighted and summed according to the variance values in the T-EFA method so that a single variable was calculated from 12 parameters (Compare, Pop, Edges, Clusters, Cluster Size, Leesalee, Slope, %Urban, Xmean, Ymean, Rad, and Fmatch). The calculations do not include the parameter Product due to multiplying the other 12 parameters. For our purposes, four-period CORINE land cover data and three-period transportation network data were used to determine the effects of urban growth on land cover change by minimizing the costs of urban growth monitoring studies. Two different urban growth models were produced using the T-EFA and OSM methods, and the results were compared with the actual land cover for 2018 by confusion matrices and calculated kappa statistics.

In line with the study purposes, Istanbul province, whose population increased by more than 50% (TSI, 2020) after the great Marmara earthquake on August 17, 1999, was selected as the project area. Although there are many studies in the literature conducted using the SLEUTH model specific to Istanbul (Ayazli et al., 2015, 2019; Mestav Sarica et al., 2020; Nigussie & Altunkaynak, 2017), a particular part of Istanbul or only one or a few of the 12 spatial metrics used in the SLEUTH model calibration were used in these studies. Our study aims to eliminate these deficiencies as follows:

- Istanbul and its surroundings constitute the study area in order to fully determine the urban sprawl impact on all forest areas, wetlands, and agricultural lands within the city boundaries.
- Furthermore, we aim to use the 12 spatial metrics' R^2 scores at the model calibration stage, based on the assumption that the SLEUTH model calibration parameters contain all the systematic effects affecting the city's growth in

the historical process that might or might not be observed. Therefore, T-EFA, a new method for model calibration, was used in the SLEUTH model calibration, and the results obtained are compared with the results obtained from the OSM method, which is the most accepted method in the literature, to test the model's success.

• Previous modeling studies using the EFA method have utilized the input data generated from high-resolution data. To determine the success of the T-EFA method in models produced using low-resolution and shorter-time-interval data, the transportation data and CORINE land cover data with a pixel size of 100 m obtained free of charge from open data sources were used.

In the calibration, a significant difference was obtained between the breed value calculated as "88" using the T-EFA method and the "1" value obtained from the OSM method. It was observed that the growth coefficients which best explained the growth form that occurred between 2000 and 2018 were obtained from the T-EFA method.

After the calibration, the prediction stage was initiated, and simulation models were created for 2040 according to the growth coefficients calculated using the T-EFA and OSM methods. As a result, it is expected that the urbanization pressure on forests, agricultural lands, and water basins in Istanbul will continue to increase in the following years.

The remainder of this article is organized in three sections. Section 2 explains the theoretical background of the simulation models created. Section 3 presents the study results, and these are compared with similar projects in the literature in Section 4.

2 | METHODOLOGY AND DATA PROCESSING

2.1 | Project area: Istanbul

The study area includes Istanbul and its surroundings. In addition to connecting the Asian and European continents, Istanbul is bordered by the Black Sea in the north and the Marmara Sea in the south, causing the city to grow on the east–west axis (Figure 1). Rapid urbanization is occurring in Istanbul, whose population has increased from 10,018,735 to 15,067,724 and increased by more than 50% between 2000 and 2018 (TSI, 2020). In the city, whose population has increased continuously since the 1950s, while growth continued on the east–west axis, the bridges built on the Bosphorus brought speed to the north, causing land cover change (Ayazli et al., 2015; Geymen, 2013) and increased urbanization pressure on forest areas, agricultural lands, and wetlands in Istanbul (Ayazli et al., 2015; Kucukmehmetoglu & Geymen, 2009). Istanbul is the city where the effects of global climate change experienced due to deforestation are the most intense in Turkey. Because of the drought in 2020, the occupancy rate decreased to 20% (ISKI, 2021).

Furthermore, the North Anatolian Fault line, which can generate Mw 7 and above earthquakes, passes through the south of Istanbul under the Marmara Sea. Depending on the population increase, areas defined as geologically risky in the Istanbul Environmental Plan (IEP) have either been built on or faced with the threat of housing. For these reasons, Istanbul was selected as the study area to serve as an example for many other projects worldwide. Thus, the modeling capability of urban growth model calibration under many conditions can be tested using the T-EFA method.

2.2 | Creating an urban growth simulation model with SLEUTH

For the Unix-based SLEUTH program to be used in the Windows operating system, a UNIX emulator Cygwin, completely free and open-source, is required. The structure of SLEUTH consists of mode process flows formed
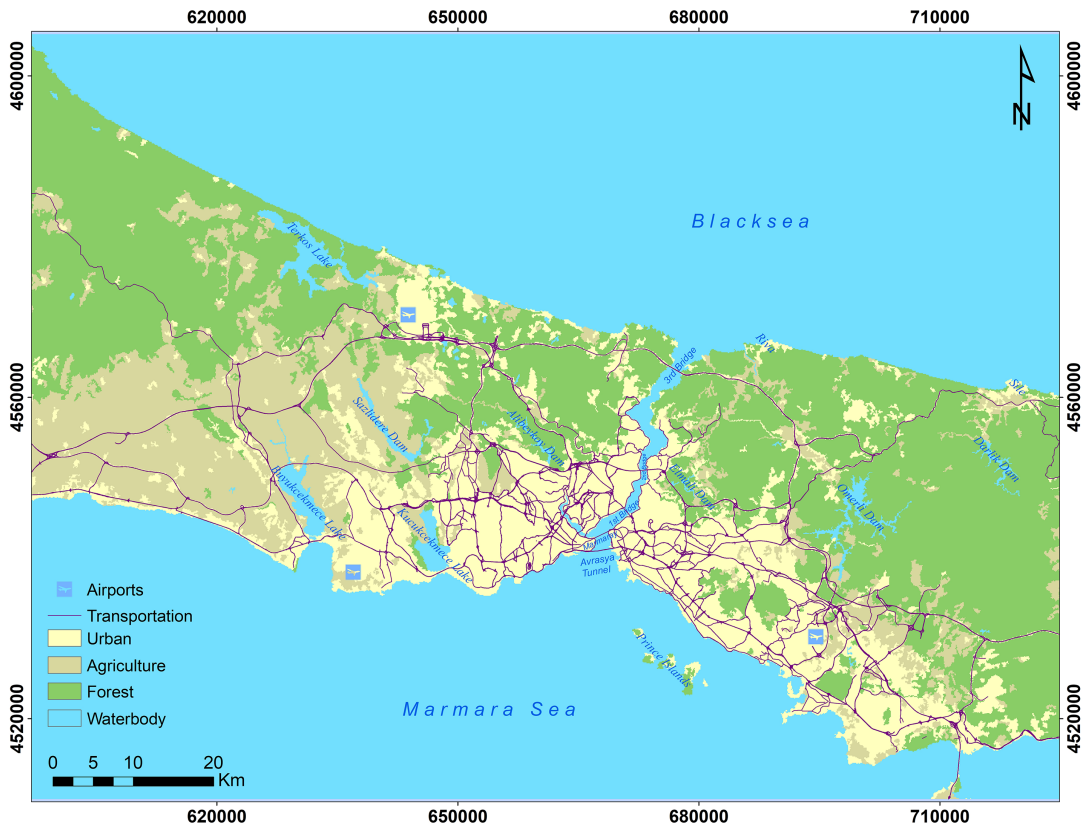
FIGURE 1 The project area

by the growth cycle, basic simulation, and test–calibration–prediction stages (Gigalopolis, 2020). In the growth cycle step, after assigning a unique value to each growth coefficient, the initial growth rules are applied, and the growth rate, which should be between the self-modifying parameters, is calculated (Clarke et al., 1997; Silva & Clarke, 2002). If the growth rate of the model, which is the sum of the four growth rules, does not take a value within "CRITICAL_HIGH" and "CRITICAL_LOW," the second-order behavioral rules, self-modifying rules, are applied (Clarke et al., 1997; Silva & Clarke, 2002). In an urban area with fast growth characteristics, the growth control parameters are multiplied by a number greater than one. Otherwise, if there is no or little growth, they are multiplied by a multiplier less than one (Clarke et al., 1997).

The primary simulation step is defined as the sequence of growth cycles that starts with the seed year of the first period settlement data used in the model and ends with the end time of the last period settlement data (Silva & Clarke, 2002). In other words, the growth cycle is generated by taking the difference between the value of the settlement data for the last year and the first year's value.

The first process flow mode is the test stage at which datasets are verified. In other words, it is the stage at which the compliance of the prepared input data and the arrangements made in the scenario file with the model standards are tested (Silva & Clarke, 2002). After the test is completed, the most crucial stage, calibration, is initiated.

Five growth coefficients are calculated according to four growth rules with Monte Carlo (MC) iteration at the calibration stage. According to the spontaneous growth rule, a function of the dispersion and slope coefficients, the suitability of a randomly selected pixel for urbanization is checked (Jantz et al., 2004). The breed coefficient controls whether a growth cell produced according to the spontaneous growth rule will be a new

spreading center and is a component of the new spreading center growth rule, together with the slope coefficient (Jantz et al., 2004; Silva & Clarke, 2002). In the edge growth rule controlled by the spread and slope coefficients, cells that have previously been a city or have newly become a city are multiplied (Jantz et al., 2004; Silva & Clarke, 2002). The road-influenced growth rule, consisting of the road gravity, dispersion, breed, and slope coefficients, replicates a newly formed urban cell along with the transportation network (Clarke & Gaydos, 1998; Jantz et al., 2004; Silva & Clarke, 2002).

Model calibration is applied to determine the growth parameters that best represent the future urban growth in the study area. The current situation and the model are compared by calculating the least-squares regression values of 13 metrics. Thus the model's accuracy is also measured. The OSM method is most commonly used to select R^2 scores. In the method developed by Dietzel and Clarke, coefficient ranges are selected by ranking the scores obtained from the multiplication of the Compare, Pop, Edges, Clusters, Slope, X-Mean, and Y-Mean criteria as in Equation (1) (Dietzel & Clarke, 2007). A value of the OSM score close to 1 indicates that the selected coefficient sequence represents urban growth in the study area with high accuracy, while a value close to 0 indicates that the representation of urban growth by the calculated coefficient sequence is weak (Dietzel & Clarke, 2007):

$$OSM = Compare * Pop * Edges * Clusters * Slope * X - mean * Y - mean \qquad (1)$$

2.3 | Coefficient selection using the T-EFA method

EFA emerges as a method frequently used in creating new theoretical structures in scientific research conducted using the inductive method. It is a parametric and multivariate statistical analysis method used to determine the underlying factor structure of a dataset. The EFA method, one of the factor analyses that have two different types, such as exploratory and confirmatory factor analysis, was preferred in this study. The variables used in EFA are expected to be measured on interval and ratio scales. In Equation (2), where X_1 is the score of the first subject for observed variable 1, b_q is the regression coefficient, F_q is the value of factor q of the subject, d_1 is the regression weight of the unique factor related to regression, and U_1 is the associated unique factor, EFA assumes that the observed variables are linear combinations of factors (Hatcher, 2007):

$$X_1 = b_1(F_1) + b_2(F_2) + \ldots + b_q(F_q) + d_1(U_1) \qquad (2)$$

There are different approaches regarding the sample size required for applying factor analysis. The general approach is that the sample size should be five times the most minor observed variables (Tabachnick & Fidell, 2013). Although there are different views, such as there should be at least 5, 10, 20, or 50 observations, or not less than 100 observations per variable (Alpar, 2011), if there are strong, reliable relationships and a small number of factors, the sample size can be decided as at least 50 provided that it is higher than the number of variables (Tabachnick & Fidell, 2013). Therefore, the correlation matrices of the variables to be included in EFA should also be used for factor analysis.

The Kaiser–Meyer–Olkin (KMO) value and Bartlett's test of sphericity (p) are used to determine the suitability of the correlation matrix for factor analysis. The KMO measure is the ratio of the total correlation variables' values to the sum of squares and the sum and partial correlation values to the sum of squares. It is stated that as this ratio approaches 1, the correlation pattern in the R-matrix is tight, and as it approaches 0, there is a spread in the pattern (Field, 2005). A tight correlation pattern between variables is desirable. Concerning this criterion, Kaiser (1974) indicated a ratio of 0.5 as an almost acceptable cut-off point while classifying KMO values 0.5–0.7 as medium, 0.7–0.8 as good, 0.8–0.9 as very good, and above 0.9 as perfect (Kaiser, 1974). The results of "Bartlett's test of sphericity" are another method used to determine the suitability of the correlation matrix for factor analysis. This test checks whether the correlation matrix obtained from the variables is the unit matrix. As a result of this

test, if the *p* value is less than 0.05, the usability of factor analysis is understood by saying that the correlation matrix is different from the unit matrix.

The *R*-matrix is a unit matrix that means the correlation coefficients between the variables are 0. Therefore, if there is no correlation between the variables, it will not be correct to talk about the existence of a typical cluster to explain the variables. However, "Bartlett's test of sphericity" is susceptible to sample size (Henson & Roberts, 2006; Zwick & Velicer, 1986). In this regard, it tends to yield significant results in large samples. Acceptable factor loadings by the number of observations were 0.30 for 350 and above units, 0.40 for 200, and 0.50 for 120 (Hair, 2006). Generally, in the literature, if the KMO value is higher than 0.45 (Balanza et al., 2007; Li, Yang, & Li, 2017) and the significance level of Bartlett's test of sphericity is less than 0.05 (Tabachnick & Fidell, 2013), it can be said that there is a fit between the sample and the EFA method.

To determine the test's reliability, common factor variance is calculated, and this value is expected to be high (Buyukozturk, 2020). While determining the number of factors, the cumulative percentage of explained variance (CTVE) should be at least 67%, and the eigenvalues should be higher than 1. Owing to this coefficient, how many factors explain the total variance cumulatively is also calculated. This is performed according to the principal component analysis (PCA) algorithm.

As a result of EFA, it is observed that variables are usually not concentrated around certain factors. Therefore, rotation is carried out to ensure the concentration of variables around these factors. Thus, we aim to ensure that the factors are not correlated. There are two types of rotation method, orthogonal and oblique (non-orthogonal). The most common orthogonal rotation methods are Varimax and Quartimax, while Oblimin and Promax are oblique rotation methods. The Varimax method is usually used in practice, and ensures that the number of variables with high factor loading on each factor is the lowest.

The EFA method calculates factor scores by grouping the intercorrelated data under the factor name. As for the T-EFA method, we aim to reduce the number of factors *n* to one by weighing these scores with the factor variance values (σ^2) calculated for each factor according to Equation (3):

$$\text{T - EFA} = \frac{\sum_{i=1}^{n} \sigma_i^2 * \text{Factor Score}_i}{\sum_{i=1}^{n} \sigma_i^2} \qquad (3)$$

Thus, 12 spatial parameters' regression scores are collected in a single group at the calibration stage, and the model contains all of the parameters' effects. These processes are calculated separately for each coarse, fine, and final calibration step.

2.4 | Preparation of input data

The first-level (urban, agricultural areas, wetlands, forests, and water bodies) CORINE land cover data with a pixel size of 100 m were used in the study. Four periods of LULC data for 2000, 2006, 2012, and 2018 were downloaded from the Copernicus Global Land Service (CGLS) website (CGLS, 2021). The slope data were produced using a digital elevation model (DEM) with a spatial resolution of 10 m, produced in 2006 by the General Directorate of Mapping (GDM) as compliance data. Geologically risky areas in the IEP, which entered into force in 2009 in order to identify possible settlement areas under earthquake risk in the future due to the expected great Istanbul earthquake, overlapped with the model produced for 2040. Finally, highways, link roads, the Bosphorus bridges, Marmaray, and Eurasia Tunnel routes were used as accessibility data. The current data were downloaded from the open data platform OpenStreetMap, and the transportation data for 2000 and 2015 were produced with the help of Google Earth™ and the Istanbul Metropolitan Municipality Application of City Map (ACM) (ACM, 2021).

The overall classification accuracy rate (OCAR) and kappa values were calculated to determine the accuracy of CORINE LULC data. The accuracy values of four periods were calculated by randomly generating 50 control points for each class in ArcGIS. These calculated values were appropriate for use in the study since they were above 80%.

The input data required by the SLEUTH model should have specific standards. Therefore, the input data were generated in the same resolution (numbers of rows and columns are equal), in the same datum and projection, in 8-bit radiometric resolution, and in accordance with the appropriate naming format conditions with the help of ArcGIS in the ".gif" format. While 0 represents a non-existent or empty value in all images, those taking a value between 1 and 256 are values that we can define as living or existent (Gigalopolis, 2021). Three different datasets were prepared as input data: (1) a pixel size of 400 m for coarse calibration; (2) 200 m for fine calibration; and (3) 100 m for final and predict.

After the input data were prepared, the UGSM was produced, as shown in Figure 2. At the first test stage, the compliance of the input data with the model standards was checked. If a problem is encountered, the test process should be repeated by checking the compliance of the input data with the standards.

Calibration, which is the second stage, is the most crucial section in line with the purposes of this study. As a result of the calibration, completed in four steps, the growth coefficients used in the prediction are calculated. The calibration stage performed with T-EFA also starts with the coarse calibration step, and after its completion, the "control_stats.log" file produced is opened in the SPSS software, and a single weighted factor score is calculated from 12 spatial metrics using the T-EFA method, a dimension reduction process. Ranges are determined for the five growth coefficients corresponding to the three highest T-EFA scores, and the next calibration step is initiated.
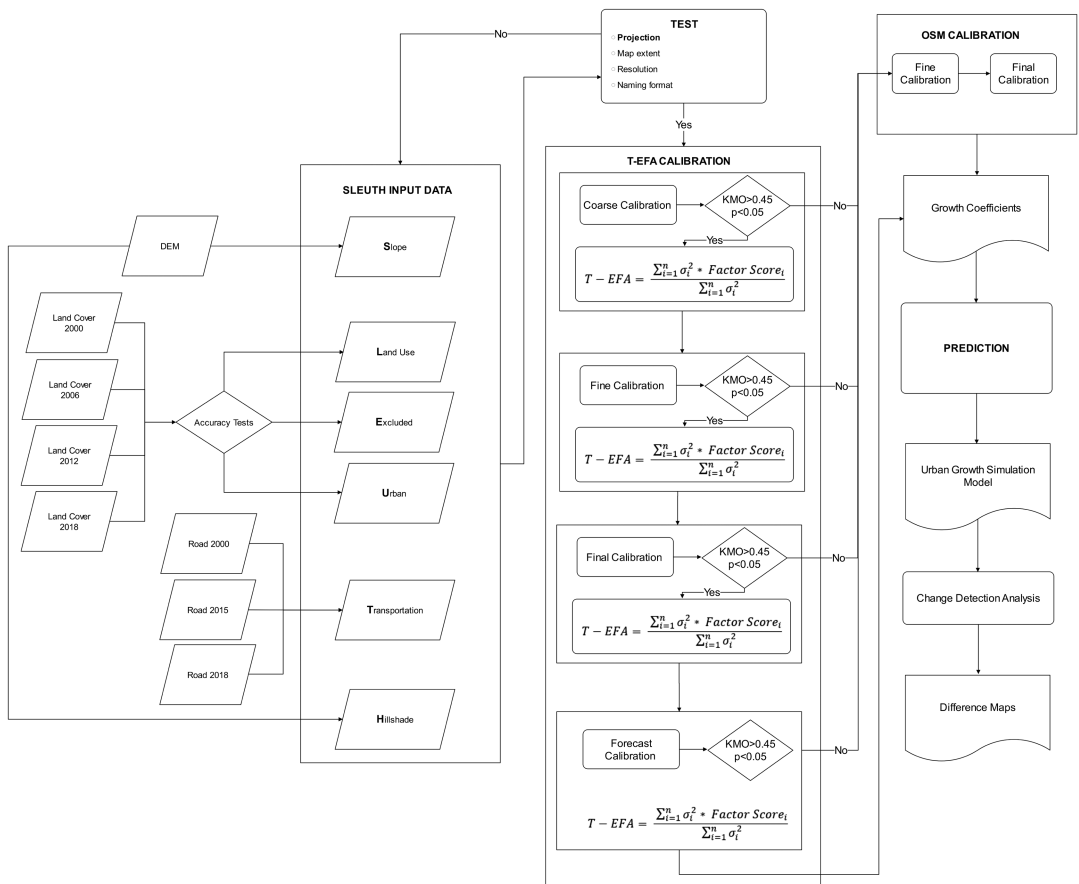

FIGURE 2 Model calibration with the T-EFA method and UGSM production

These processes are repeated at each step. To use the T-EFA method, the KMO and p values, in which the sample's fit is tested, are calculated in all three calibration steps. If KMO is higher than 0.45 and p is less than 0.05, the T-EFA method can be used at the calibration stage. Otherwise, it will be more appropriate to use the OSM method for calibration.

After completing the calibration stage, the prediction is started, at which the UGSM will be produced. At the final stage, the effects of urban growth are determined by conducting a change detection analysis between the prediction model produced to determine the effects of urban growth and the land cover input data of the last period.

To measure the success of the T-EFA method, a second model using the OSM method at the calibration stage was produced, and the results obtained were compared.

3 | RESULTS

3.1 | Urban growth in Istanbul

This study aimed to determine the urban sprawl effects on land cover change by producing urban growth simulation models with short-term low-resolution data obtained from open data sources. A CA-based simulation model for 2040 was created for Istanbul. As the LULC data required for the model, land cover data of settlement areas, agricultural lands, forests, wetlands, and water bodies with a pixel size of 100 m and a minimum mapping unit (MMU) of 25 ha were used. In this context, the CORINE land cover data provided free-of-charge for 2000, 2006, 2012, and 2018 were downloaded online. An accuracy analysis was performed for each land cover class to decide whether to use the CORINE data in the study. The OCAR and kappa values were calculated by generating 50 random points for each class (Table 1). Since more than 80% accuracy was calculated in all periods, it was appropriate to use CORINE data in the study.

Between 2000 and 2018, 14% of agricultural areas and 7% of forests were converted into settlements; in other words, urban growth will continue in Istanbul. Moreover, it was identified that 15% of settlement areas in 2018 were in geologically risky areas. According to the difference map generated to show urban growth between 2000 and 2018 (Figure 3), it is observed to be concentrated between the Istanbul Airport and the Third Bridge in the northwest of the European side, between Sazlıdere Dam Lake and Alibeykoy Dam Lake, around the link road of the Third Bridge in Basaksehir district and to the west of Buyukcekmece Lake in the south, on the coast of the Marmara Sea. On the Anatolian side, urban growth occurs in the north, in Riva and Sile under the effect of the Third Bridge, around the bridge link road passing through Sultanbeyli and the forest in the west of Omerli Dam, and the east of Sabiha Gokcen Airport in the south. This determined growth complies with the four growth rules of SLEUTH software.

The urban growth that has occurred in 18 years is in the form of "Spontaneous and New Spreading Center" in Istanbul Airport and especially forest lands, in the form of "Edge Growth" between Sazlıdere and Alibeykoy Dam Lakes and in Riva and Sile, and the form of "Road-Influenced Growth" along the Third Bridge route and in the south, along the coast.

TABLE 1 Accuracy analysis results for CORINE LULC data

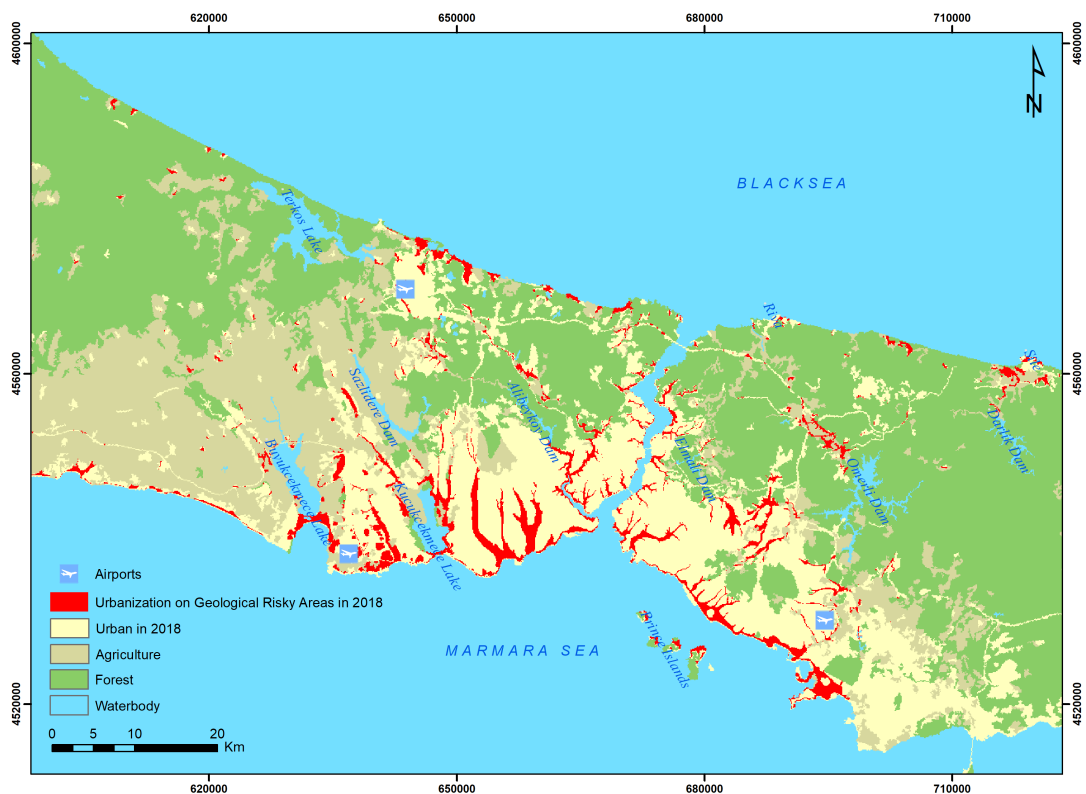| Year	2000	2006	2012	2018
OCAR (%)	93	93	92	87
Kappa (%)	91	91	89	82

FIGURE 3 Land cover in 2018

3.2 | Model calibration with T-EFA and OSM

In line with the standards of the SLEUTH model, input data were created in the form of 8-bit.gif images, and CA-based modeling studies were initiated. After the test stage was completed successfully, two different calibration procedures were performed under the same conditions using the T-EFA and OSM methods. The first step of calibration, which is completed in three steps, starts with coarse calibration. In the coarse calibration performed using input data with a pixel size of 400 m, the initial value was set to 0, the end value to 100, the step value to 25 in both methods, and 3125 growth cycles were generated. After the coarse calibration step, which was completed in 49 min 23 s, in the first method using the T-EFA method, KMO was calculated as 0.800, p as 0.000, and CTVE score as 83.300%, and 12 parameters were divided into three intercorrelated classes. Each class was weighted with its variance value, the T-EFA value was calculated, and thus a single parameter was obtained to determine the coefficient range. The coefficient ranges used in the fine calibration step were selected according to the three highest T-EFA scores, and the fine calibration step was initiated.

The fine calibration step, in which input data of a 200 m pixel size according to the selected coefficient range were used, was completed in 9 h 25 s by making 7776 calculations in eight iterations. As a result of the calculated KMO 0.780, p 0.000, and CTVE 82.000% values, it was decided that the sample was also suitable for the fine calibration step of the T-EFA method. The T-EFA value was determined by weighing each of the four factor scores obtained with its calculated variance values, and the coefficient ranges for final calibration were selected according to the three highest T-EFA scores.

In the final calibration step, 4500 growth cycles were generated, and the number of iterations was set to 10. After the procedures were completed in 1 day 6 h 4 min 49 s, the KMO was calculated as 0.868, p 0.000, and

TABLE 2 Model's calibration stages and the calculated coefficients

| | Coarse step (5 iterations) | | Fine step (8 iterations) | | Final step (10 iterations) | | | |
| | Start–step–stop | | Start–step–stop | | Start–step–stop | | Growth coefficient values | |
	T-EFA	OSM	T-EFA	OSM	T-EFA	OSM	T-EFA	OSM
Dispersion	0–25–100	0–25–100	0–15–75	25–1–25	28–1–32	25–1–25	38	25
Breed	0–25–100	0–25–100	0–15–75	0–25–100	73–1–77	1–1–1	88	1
Spread	0–25–100	0–25–100	0–5–25	25–1–25	23–1–27	25–1–25	32	25
Slope	0–25–100	0–25–100	50–10–100	75–1–75	50–2–60	75–1–75	13	75
Road	0–25–100	0–25–100	0–15–75	0–5–25	30–3–45	0–5–25	50	15
Growth cycle	3125	3125	7776	30	4500	6		
Time	49′ 23″	49′ 23″	9^h 00′ 25″	2′ 34″	1^{day} 6^h 04′ 49″	2′ 17″		
KMO/p/CTVE (%)	80.0/0.0/83.3		78.0/0.0/82.0		86.8/0.0/96.1			

TABLE 3 Comparison of the calibration accuracy by confusion matrices

Calibration technique	Land cover 2018	Model (ha)	Model (ha)	Total (ha)	Recognition (%)
T-EFA	Actual	111,601	19,333	130,934	85
	Actual	38,696	1,062,131	1,100,827	96
	Total	150,297	1,081,465	1,231,761	95
OSM	Actual	106,150	23,164	129,314	82
	Actual	23,200	1,077,611	1,100,811	98
	Total	129,350	1,100,775	1,230,125	96

CTVE 96.132%. Thus, it was decided to use the T-EFA method of the sample, and each of the two factor scores obtained was weighted with its variance values calculated, and the growth coefficients to be used at the prediction stage were calculated according to the highest T-EFA score (Table 2). The breed, slope, and road gravity coefficients were 88, 13, and 50 in the T-EFA; 1, 75, and 15 in the OSM method. The dispersion and spread coefficients were calculated as 38 and 32 in the T-EFA method, and 25 in the OSM method.

The calibration stage was repeated with the OSM method to compare the T-EFA results, and the OSM was completed in approximately 1 h. The highest OSM values calculated for the coarse, fine, and final calibration steps were 0.11952540, 0.00019234, and 0.00594196, respectively.

3.3 | Model validation

Two confusion matrices were produced from actual and model data between 2000 and 2018 in order to interpret the calibration accuracy. The calculated total recognition values are 95% T-EFA and 96% OSM. In other words, it can be expressed that the success rate of both methods is at par. However, the T-EFA method is 3% more successful in predicting urban growth, albeit slightly. On the other hand, the OSM method performed 2% better in identifying areas that will not be urbanized (Table 3).

As very close values were obtained from the confusion matrices, another cell-based comparison method—kappa statistics—was used between actual land cover data 2018 and two sets of modeled data. Fifty random control points were generated for each urban and non-urban dataset. The kappa and OCAR scores were 78 and 89% for the T-EFA, whereas they were 74 and 87% for the OSM method. Model validation results obtained from the kappa statistics are illustrated in Figure 4. T-EFA modeled the urbanized pixel finer than OSM according to the "New Spreading Center" and "Road-Influenced Growth" rules, due to higher breed and road growth coefficients (Figure 4a,b). However, neither method could detect the urban growth at Istanbul International Airport (Figure 4c) because the airport was constructed by a cabinet decision made in 2012 and was not located in the Istanbul Master Plan prepared in 2010.

3.4 | Creating urban growth models for 2040

According to the simulation model produced for 2040 after the prediction stage, it was determined that urban growth would continue between 2018 and 2040. According to the model created with the growth coefficients by the T-EFA method, 36.5% of agricultural lands (42,872 ha), 8.5% of wetlands (13 ha), and 20.6% of forests (55,339 ha) were converted into settlement areas. In contrast, according to the OSM method, these rates were 8% (~9604 ha) in agricultural lands, 1% (~1 ha) in wetlands, and 3% (~7704 ha) in forests (Table 4 and Figure 5). Both simulation models observed a decrease in agricultural lands, wetlands, and forest areas, while urban areas increased.
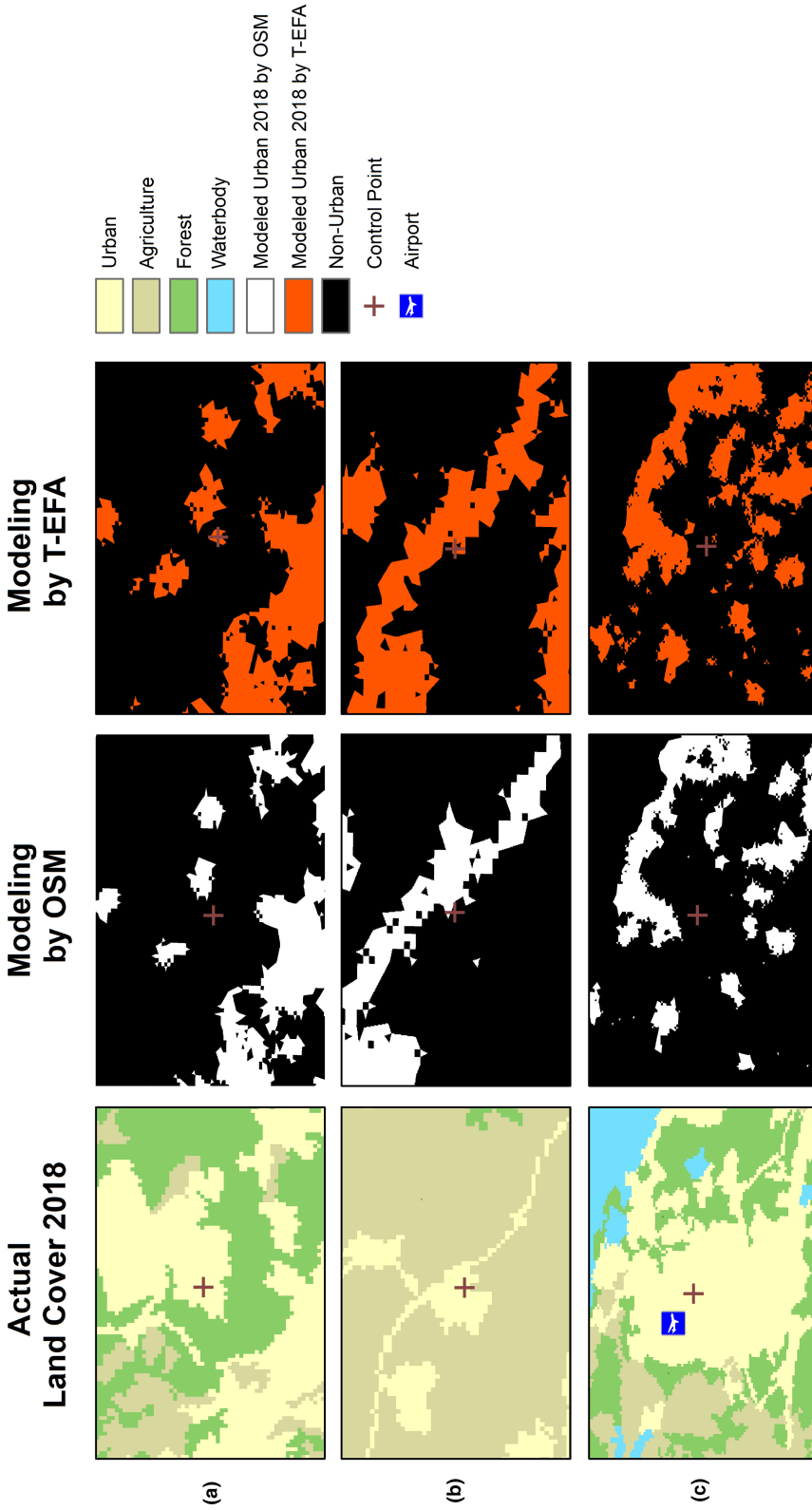
FIGURE 4 Model validations of the T-EFA and OSM according to kappa statistics: (a) for the breed coefficient; (b) for the road growth coefficient; and (c) could not be detected by T-EFA and OSM

TABLE 4 Conversion to urban settlements 2000–2018 and 2018–2040

To urban settlements	2000–2018		2018–2040 (T-EFA)		2018–2040 (OSM)	
From agricultural areas	14%	17,496 ha	37%	42,872 ha	8%	9604 ha
From forests	7%	19,066 ha	21%	55,339 ha	3%	7704 ha
From wetlands	0%	0 ha	9%	13 ha	1%	1 ha
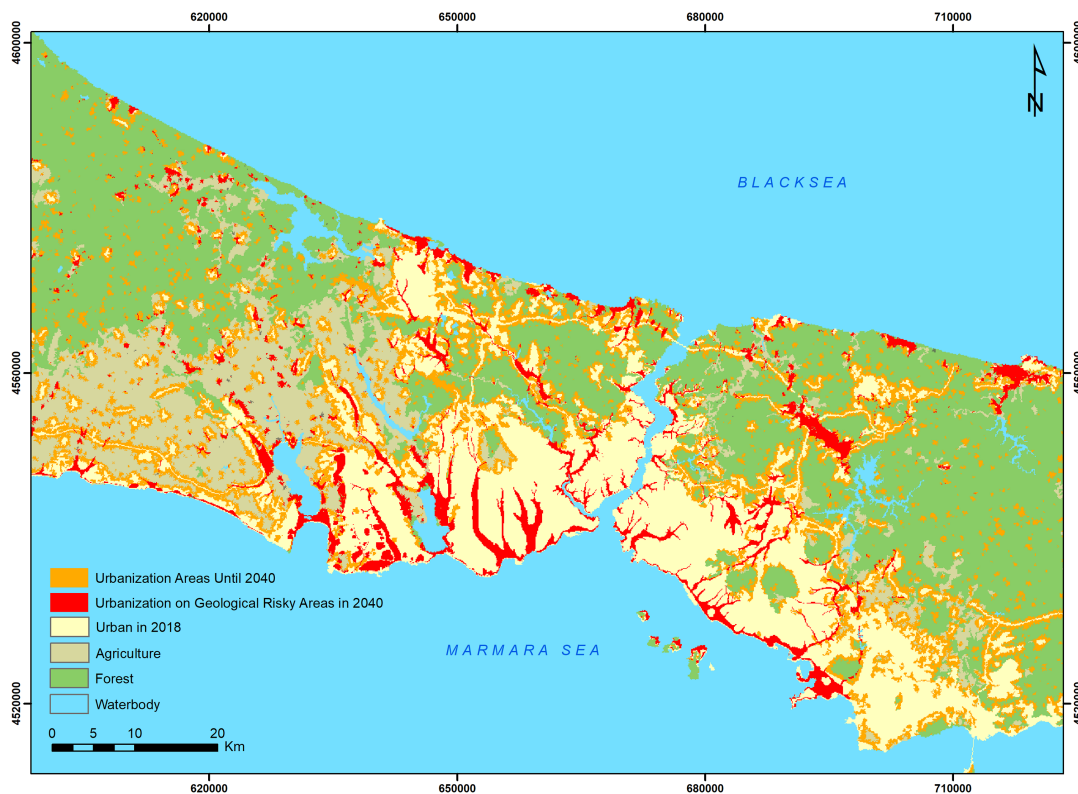
FIGURE 5 Land cover in 2040

As shown in Table 3, it was determined that the model results created using the T-EFA method were close to the change rates between 2000 and 2018. According to the model created using T-EFA, urban growth between 2018 and 2040 in Istanbul shows similarity with the growth trend between 2000 and 2018. While urban growth is concentrated around Istanbul Airport and the Third Bridge in the north, it continues along the coast in the south-west and occurs in forest lands through which ring road connections pass in the middle section.

4 | DISCUSSION AND CONCLUSIONS

Two simulation models were created using open data sources to determine the urban sprawl effects on land cover change. In this context, first-level CORINE land cover data with a pixel size of 100 m required in the model were obtained free-of-charge from CGLS, and accessibility data were obtained free-of-charge from the OpenStreetMap, Google Earth™, and Istanbul City Guide websites. Slope data were generated from the DEM data produced by GDM in 2006 and obtained through an inter-institutional agreement within the scope of another

project in previous years. The UGSMs were generated using CA-based, open-source, and free SLEUTH software to minimize project costs. Due to the low resolution of the data obtained from open data sources, the accuracy of the simulation model created may be low. Therefore, the T-EFA method was used at the calibration stage of the UGSM in order to improve model accuracy, and the calibration results were compared with the model results calibrated using the OSM method.

At the beginning of the study, the OCAR and kappa values belonging to four periods of land cover were calculated to determine the accuracy of CORINE land cover data. Since the values calculated by randomly generating 50 points for each class in each period were above 80%, it was deemed appropriate to use the CORINE land cover data as input data in the model.

Upon examining the land cover data for 2000 and 2018, it was observed that Istanbul continues to grow, depending on the population increase. This growth occurred mainly around the existing settlement areas of Sile and Riva on the Anatolian side under the effect of the Third Bridge passing through the north of Istanbul, in Sultanbeyli district and the forests in the west of Omerli Dam, where the bridge link road passes. At the same time, it is concentrated around Istanbul Airport, within the boundaries of the Basaksehir district, around the bridge link road passing through the forest area between Sazlıdere and Alibeykoy Dams on the European side. This growth overlaps with studies claiming that the bridges built in Istanbul caused the city to grow northward (Ayazli et al., 2015; Geymen, 2013). Furthermore, there is growth on the east–west axis along the coast in the south, though only a small amount.

The urban growth coefficients support these aforementioned urban sprawl characteristics. A considerable difference emerged in the breed and road gravity coefficients, even though the dispersion and spread values were calculated close to each other. While the coefficients obtained with T-EFA are 88 for the breed and 50 for the road gravity, the values were 1 and 15 in the OSM. The lower value of the slope coefficient is compatible with the previous studies implemented in Istanbul (Ayazli et al., 2015, 2019).

The calibration results showed that the OSM method remained insufficient in explaining the growth according to the "New Spreading Center" and "Road-Influenced Growth" rules. According to the confusion matrices and kappa analysis, the growth form in Istanbul between 2000 and 2018 was better explained with the T-EFA method. However, to use the T-EFA method, the condition of sample suitability must be fulfilled in each calibration step. Unfortunately, this situation emerges as a disadvantage of the T-EFA method. Thus, the KMO, p, and CTVE values, in which the fit of the model was checked, were calculated separately for each of the coarse, fine, and final steps, and since the KMO value for each calibration step was more significant than 0.45, p was less than 0.05, and CTVE was higher than 67%, the T-EFA method could be used in calibration. Another disadvantage of the T-EFA method is the processing time. While the calibration procedure with T-EFA took approximately 2 days, the OSM method was completed faster, in 54 min 14 s.

Two different calibration procedures were conducted using the T-EFA and OSM methods. In the calibration step, the success rates of both methods are very close to each other, and the T-EFA method can be considered a robust calibration method like the OSM method. Especially as an alternative method, T-EFA may be preferred in regions where fragile ecosystems are located, because its rate of detecting urban growth is higher than that of OSM.

After the prediction stage, two different urban growth simulation models were generated for 2040 according to the T-EFA and OSM methods. As a result of the temporal change analysis, the conversion rates were calculated from forests, agricultural lands, and wetlands to urban areas between 2018 and 2040 (Table 4). In contrast, the rates determined according to the model produced with T-EFA are approximately 4.5 times higher in agricultural lands and seven times in forest areas than OSM; this rate increases to 8.5 times in wetlands. While urban growth simulation models with higher accuracy can be produced from low-resolution data using the T-EFA method, models with lower accuracy can be obtained in a shorter time using OSM. The important thing here is to determine which calibration method to use, in which the growth coefficients that best represent urban growth are calculated. Then, urban growth should be monitored using the most appropriate calibration method for the study area,

aiming to minimize possible damage. The T-EFA method can also be considered a good alternative for calibrating CA-based SLEUTH software in future studies.

There is also the North Anatolian Fault line, which passes about 20 km south of Istanbul and where very severe earthquakes have occurred throughout history. According to the report prepared jointly by Istanbul Metropolitan Municipality and Bosphorus University Kandilli Observatory and Earthquake Research Institute Department of Earthquake Engineering in 2020, an earthquake scenario of Mw 7.5 would damage 17% of the buildings in Istanbul at medium level and above. It has been calculated that approximately 12,400–14,150 people will lose their lives, and a financial loss of around $15 billion will occur (Cakti et al., 2020). While urbanization in geologically risky areas covered 20,494 ha in 2018, it will increase to 32,740 ha in 2040 by the T-EFA calibration technique. Unfortunately, this increment will increase the costs and loss of life and property.

Data production costs are an essential budget item in studies monitoring urban growth. Nowadays, the number of open data sources is increasing every day, and the accuracy of the data obtained from these sources has reached satisfactory levels. This study has proven that high-accuracy urban growth simulation models can be generated by minimizing data production costs. However, scientific studies alone are not sufficient in such projects. For the produced scientific results to show their effect, the support of the city administrators for these projects is vital in terms of the correct management of resources. In this context, to minimize the possible damage that may be experienced, it is necessary to determine the boundaries of urban growth, include them in plans, and prepare the necessary legal and administrative infrastructure to protect these boundaries.

Consequently, the rapid population increase in Istanbul and the urban growth brought about by it have caused land cover change. In line with this, especially considering the urban growth structure of Istanbul, it is thought that the unbuilt areas and their surroundings in Istanbul are under intense urbanization pressure. Decision-makers can use urban growth simulation models as a prediction tool while developing policies to control urban growth. Simulation models should be produced under alternative scenarios to evaluate the possible effects of infrastructure projects triggering dynamic and complex land cover changes in the metropolis of Istanbul. Moreover, considering the earthquake risk, necessary policies should be developed urgently to slow down or prevent potential urbanization in risky regions.

CONFLICT OF INTEREST

All authors declare that they have no conflicts of interest.

DATA AVAILABILITY STATEMENT

Research data are not shared.

ORCID

Ismail Ercument Ayazli 🔟 https://orcid.org/0000-0003-0782-5366
Ahmet Emir Yakup 🔟 https://orcid.org/0000-0002-1789-4448
Omer Bilen 🔟 https://orcid.org/0000-0001-7198-8421

REFERENCES

ACM. (2021). *Application of city map*. Retrieved from https://sehirharitasi.ibb.gov.tr/

Alkan, R. M., Ozulu, M., Ilci, V., Tombus, F., & Sahin, M. (2017). Usability of GNSS technique for cadastral surveying. In T. Yomralioglu & J. McLaughlin (Eds.), *Cadastre: Geo-information innovations in land administration* (pp. 266–286). Cham, Switzerland: Springer. https://doi.org/10.1007/978-3-319-51216-7_8

Alpar, R. (2011). *Applied multi-variable statistical methods*. Ankara, Turkey: Detay Yayincilik.

Angel, S., Parent, J., Civco, D., & Blei, A. (2011). *Making room for a planet of cities*. Cambridge, MA: Lincoln Institute of Land Policy.

Ayazli, I. E. (2019). Monitoring of urban growth with improved model accuracy by statistical methods. *Sustainability*, *11*(20), 5579. https://doi.org/10.3390/su11205579

Ayazli, I. E., & Bilen, O. (2019). Using exploratory factor analysis to improve the calibration of SLEUTH urban growth models. *Fresenius Environmental Bulletin*, *28*(2), 975–979.

Ayazli, I. E., Gul, F. K., Baslik, S., Yakup, A. E., & Kotay, D. (2019). Extracting an urban growth model's land cover layer from spatio-temporal cadastral database and simulation application. *Polish Journal of Environmental Studies*, *28*(3), 1063–1069. https://doi.org/10.15244/pjoes/89506

Ayazli, I. E., Kilic, F., Lauf, S., Demir, H., & Kleinschmit, B. (2015). Simulating urban growth driven by transportation networks: A case study of the Istanbul third bridge. *Land Use Policy*, *49*, 332–340. https://doi.org/10.1016/j.landusepol.2015.08.016

Balanza, R., Garcia Lorda, P., Perez-Rodrigo, C., Aranceta, J., Bulló, M., & Salas-Salvadó, J. (2007). Trends in food availability determined by the Food and Agriculture Organization's food balance sheets in Mediterranean Europe in comparison with other European areas. *Public Health Nutrition*, *10*, 168–176. https://doi.org/10.1017/S1368980007246592

Bart, I. L. (2010). Urban sprawl and climate change: A statistical exploration of cause and effect, with policy options for the EU. *Land Use Policy*, *27*(2), 283–292. https://doi.org/10.1016/j.landusepol.2009.03.003

Batty, M. (2007). *Cities and complexity: Understanding cities with cellular automata, agent-based models, and fractals*. Cambridge, MA: MIT Press.

Batty, M. (2008). The size, scale, and shape of cities. *Science*, *319*, 769–771. https://doi.org/10.1126/science.1151419

Benenson, I., & Torrens, P. (2004). *Geosimulation: Automata-based modeling of urban phenomena*. New York, NY: John Wiley & Sons.

Bihamta, N., Soffianian, A., Fakheran, S., & Gholamalifard, M. (2015). Using the SLEUTH urban growth model to simulate future urban expansion of the Isfahan Metropolitan Area. *Iran. Journal of the Indian Society of Remote Sensing*, *43*(2), 407–414. https://doi.org/10.1007/s12524-014-0402-8

Buyukozturk, S. (2020). Factor analysis: Basic concepts and its use in scale development. *Educational Administration: Theory and Practice*, *32*, 470–483.

Cakti, E., Safak, E., Hancilar, U., Sesetyan, K., Bas, M., Kilic, O., … Kara, S. (2020). *Project of updating possible earthquake loss estimates in Istanbul Province*. Ankara, Turkey: BOUN-IMM.

Candau, J. (2000). Calibrating a cellular automaton model of urban growth in a timely manner. In *Proceedings of the Fourth International Conference on Integrating GIS and Environmental Modeling (GIS/EM4): Problems, Prospects and Research Needs*, Banff, AB, Canada (pp. 2–8).

Candau, J., & Clarke, K. C. (2000). Probabilistic land cover transition modeling using deltatrons. In *Proceedings of the 2000 URISA Annual Conference*, Orlando, FL (pp. 1–11).

CGLS. (2021). *Copernicus Global Land Service*. Retrieved from https://land.copernicus.eu/global/

Chaudhuri, G., & Clarke, K. C. (2013). The SLEUTH land use change model: A review. *International Journal of Environmental Resources Research*, *1*(1), 88–104.

Clarke, K., & Gaydos, L. (1998). Loose-coupling a cellular automaton model and GIS: Long-term urban growth prediction for San Francisco and Washington/Baltimore. *International Journal of Geographical Information Science*, *12*(7), 699–714. https://doi.org/10.1080/136588198241617

Clarke, K., Hoppen, S., & Gaydos, L. J. (1996). Methods and techniques for rigorous calibration of a cellular automaton model of urban growth. In *Proceedings of the Third International Conference on Integrating GIS and Environmental Modeling*, Santa Fe, NM (pp. 1319–1328).

Clarke, K., Hoppen, S., & Gaydos, L. (1997). A self-modifying cellular automaton model of historical urbanization in the San Francisco Bay area. *Environment and Planning B: Planning and Design*, *24*(2), 247–261. https://doi.org/10.1068/b240247

Clewlow, L. (1989). *Cellular automata and dynamical systems* (Unpublished PhD dissertation). Coventry, UK: University of Warwick.

Dennunzio, A., Formenti, E., & Kurka, P. (2012). Cellular automata dynamical systems. In G. Rozenberg, T. Bäck, & J. N. Kok (Eds.), *Handbook of natural computing* (pp. 25–75). Berlin, Germany: Springer. https://doi.org/10.1007/978-3-540-92910-9_2

Dezhkam, S., Jabbarian Amiri, B., Darvishsefat, A., & Sakieh, Y. (2013). Simulating the urban growth dimensions and scenario prediction through sleuth model: A case study of Rasht County, Guilan, Iran. *GeoJournal*, *79*, 591–604. https://doi.org/10.1007/s10708-013-9515-9

Di Lena, P., & Margara, L. (2008). Computational complexity of dynamical systems: The case of cellular automata. *Information and Computation*, *206*, 1104–1116. https://doi.org/10.1016/j.ic.2008.03.012

Dietzel, C., & Clarke, K. (2007). Toward optimal calibration of the SLEUTH land use change model. *Transactions in GIS*, *11*, 29–45. https://doi.org/10.1111/j.1467-9671.2007.01031.x

Dyson, T. (2011). The role of the demographic transition in the process of urbanization. *Population and Development Review*, *37*(s1), 34–54. https://doi.org/10.1111/j.1728-4457.2011.00377.x

EEA. (2006). *Urban sprawl in Europe: The ignored challenge*. Retrieved from https://www.eea.europa.eu/tr/publications/briefing_2006_4

EEA. (2016). *Urban sprawl in Europe: Joint EEA-FOEN*. Retrieved from https://www.eea.europa.eu/publications/urban-sprawl-in-europe

Emadodin, I., Reinsch, T., & Taube, F. (2019). Drought and Desertification in Iran. *Hydrology*, 6(3), 66. https://doi.org/10.3390/hydrology6030066

Feng, Y., & Tong, X. (2018). Calibration of cellular automata models using differential evolution to simulate present and future land use. *Transactions in GIS*, 22(2), 582–601. https://doi.org/10.1111/tgis.12331

Field, A. (2005). *Discovering statistics using SPSS*. Thousand Oaks, CA: SAGE Publications.

Geymen, A. (2013). Impacts of Bosphorus bridges on the Istanbul metropolitan settlement areas. *Land Degradation and Development*, 24(2), 156–169. https://doi.org/10.1002/ldr.1114

Gigalopolis. (2020). *Project Gigalopolis: Structure*. Retrieved from http://www.ncgia.ucsb.edu/projects/gig/About/bkStructure.html

Gigalopolis. (2021). *Project Gigalopolis: Inputs*. Retrieved from http://www.ncgia.ucsb.edu/projects/gig/About/dtInput.htm

Hair, J. F. (2006). *Multivariate data analysis* (6th ed.). Hoboken, NJ: Pearson Prentice Hall.

Han, H., Hwang, Y., Ha, S., & Byung Sik, K. (2015). Modeling future land use scenarios in South Korea: Applying the IPCC special report on emissions scenarios and the SLEUTH model on a local scale. *Environmental Management*, 55(5), 1064–1079. https://doi.org/10.1007/s00267-015-0446-8

Hatcher, L. (2007). *A step-by-step approach to using SAS for factor analysis and structural equation modeling*. Cary, NC: SAS Press.

Henson, R., & Roberts, J. (2006). Use of exploratory factor analysis in published research: Common errors and some comment on improved practice. *Educational and Psychological Measurement*, 66, 393–416. https://doi.org/10.1177/0013164405282485

Huang, B., & Zhang, W. (2014). Sustainable land-use planning for a downtown lake area in central China: Multiobjective optimization approach aided by urban growth modeling. *Journal of Urban Planning and Development*, 140(2), 04014002. https://doi.org/10.1061/(asce)up.1943-5444.0000186

ISKI. (2021). *ISKI dam occupancy rate*. Retrieved from https://www.iski.istanbul/web/tr-TR/baraj-doluluk

Jantz, C., Goetz, S., & Shelley, M. (2004). Using the SLEUTH urban growth model to simulate the impacts of future policy scenarios on urban land use in the Baltimore-Washington metropolitan area. *Environment and Planning B: Planning and Design*, 31(2), 251–271. https://doi.org/10.1068/b2983

Junfeng, J. (2003). *Transition rule elicitation for urban cellular automata models, case study: Wuhan, China* (Unpublished M.S. thesis). Enschede, The Netherlands: International Institute for Geo-Information Science and Earth Observation.

Kaiser, H. (1974). An index of factorial simplicity. *Psychometrika*, 39, 31–36. https://doi.org/10.1007/BF02291575

Khorrami, B., Gunduz, O., Patel, N., Ghouzlane, S., & Najjar, M. (2019). Land surface temperature anomalies in response to changes in forest cover. *International Journal of Engineering and Geosciences*, 4(3), 149–156. https://doi.org/10.26833/ijeg.549944

Kucukmehmetoglu, M., & Geymen, A. (2009). Urban sprawl factors in the surface water resource basins of Istanbul. *Land Use Policy*, 26(3), 569–579. https://doi.org/10.1016/j.landusepol.2008.08.007

Kusak, L., & Kucukali, U. (2018). Outlier detection of land surface temperature: Kucukcekmece region. *International Journal of Engineering and Geosciences*, 4, 126–132. https://doi.org/10.26833/ijeg.404426

Li, S., Yang, Z., & Li, H. (2017). Statistical evaluation of no-reference image quality assessment metrics for remote sensing images. *ISPRS International Journal of Geo-Information*, 6, 133. https://doi.org/10.3390/ijgi6050133

Liu, X., Li, X., Shi, X., Huang, K., & Liu, Y. (2012). A multi-type ant colony optimization (MACO) method for optimal land use allocation in large areas. *International Journal of Geographical Information Science*, 26(7), 1325–1343. https://doi.org/10.1080/13658816.2011.635594

Loh, E. H., Zambrana-Torrelio, C., Olival, K. J., Bogich, T. L., Johnson, C. K., Mazet, J. A. K., … Daszak, P. (2015). Targeting transmission pathways for emerging zoonotic disease surveillance and control. *Vector-Borne and Zoonotic Diseases*, 15(7), 432–437. https://doi.org/10.1089/vbz.2013.1563

Mestav Sarica, G., Zhu, T., & Pan, T. C. (2020). Spatio-temporal dynamics in seismic exposure of Asian megacities: Past, present and future. *Environmental Research Letters*, 15(9), 094092. https://doi.org/10.1088/1748-9326/ababc7

Nigussie, T. A., & Altunkaynak, A. (2017). Modeling the effects of Project Canal Istanbul on the urban extent and hydrological response of Ayamama watershed, Istanbul. In *Proceedings of the 2017 World Environmental and Water Resources Congress*, Sacramento, CA (pp. 1–11). Reston, VA: ASCE. https://doi.org/10.1061/9780784480601.001

Pontius, G. R., & Malanson, J. (2005). Comparison of the structure and accuracy of two land change models. *International Journal of Geographical Information Science*, 19(2), 243–265. https://doi.org/10.1080/13658810410001713434

Saxena, A., & Jat, M. K. (2020). Analysing performance of SLEUTH model calibration using brute force and genetic algorithm-based methods. *Geocarto International*, *35*(3), 256–279. https://doi.org/10.1080/10106049.2018.1516242

Selim, S., & Demir, N. (2019). Detection of ecological networks and connectivity with analyzing their effects on sustainable urban development. *International Journal of Engineering and Geosciences*, *4*(2), 63–70. https://doi.org/10.26833/ijeg.443114

Sevik, O. (2006). *Application of Sleuth model in Antalya* (Unpublished MS thesis). Ankara, Turkey: Middle Eastern Technical University.

Silva, E., & Clarke, K. (2002). Calibration of the SLEUTH urban growth model for Lisbon and Porto, Portugal. *Computers, Environment and Urban Systems*, *26*, 525–552. https://doi.org/10.1016/S0198-9715(01)00014-X

Siri, J. G., Brown, Z., & Spielauer, M. (2010). Simulation modeling of the long-term evolution of local malaria transmission and acquired immunity in the context of urban growth and urban–rural travel. *Malaria Journal*, *9*(2), P47. https://doi.org/10.1186/1475-2875-9-S2-P47

Stecklov, G. (2018). *The components of urban growth in developing countries*. Preprint. Retrieved from https://osf.io/preprints/socarxiv/4zk5b/

Tabachnick, B., & Fidell, L. (2013). *Using multivariate statistics* (6th ed.). London, UK: Pearson Education.

Torrens, P. M. (2000). *How cellular models of urban systems work: 1. Theory* (CASA Working Paper No. 28). London, UK: Centre for Advanced Spatial Analysis, University College London.

TSI. (2020). *Indicators*. Retrieved from https://biruni.tuik.gov.tr/medas/?kn=95&locale=tr

Tu, J., Xia, Z. G., Clarke, K. C., & Frei, A. (2007). Impact of urban sprawl on water quality in eastern Massachusetts, USA. *Environmental Management*, *40*(2), 183–200. https://doi.org/10.1007/s00267-006-0097-x

UN. (2019). *World urbanization prospects: The 2018 revision*. New York, NY: United Nations. https://doi.org/10.18356/b9e995fe-en

Wang, R., Feng, Y., Wei, Y., Tong, X., Zhai, S., Zhou, Y., & Wu, P. (2021). A comparison of proximity and accessibility drivers in simulating dynamic urban growth. *Transactions in GIS*, *25*(2), 923–947. https://doi.org/10.1111/tgis.12707

White, R., & Engelen, G. (1997). Cellular automata as the basis of integrated dynamic regional modelling. *Environment and Planning B: Planning and Design*, *24*, 235–246. https://doi.org/10.1068/b240235

White, R., Straatman, B., & Engelen, G. (2004). Planning scenario visualization and assessment: A cellular automata based integrated spatial decision support system. In M. F. Goodchild & D. G. Janelle (Eds.), *Spatially integrated social science* (pp. 420–442). Oxford, UK: Oxford University Press.

WUP. (2019). *World urbanization prospects: The 2018 revision*. New York, NY: United Nations. Retrieved from https://population.un.org/wup/Download/Files/WUP2018-F02-Proportion_Urban.xls

Yang, X., & Lo, C. P. (2003). Modelling urban growth and landscape changes in the Atlanta metropolitan area. *International Journal of Geographical Information Science*, *17*(5), 463–488. https://doi.org/10.1080/1365881031000086965

Zwick, W., & Velicer, W. (1986). Comparison of five rules of determining the number of components to retain. *Psychological Bulletin*, *99*(3), 432–442. https://doi.org/10.1037/0033-2909.99.3.432