Supplementary file

Methodology

Land use prediction using CA-ANN model and sensitivity analysis

CA-ANN model

The ANN is a machine learning method capable of capturing and representing complex interactions between inputs and outputs. The ANN is a network of nodes that is inspired by the simplification of neurons in the brain. It consists of a number of neurons or nodes that act in parallel to convert input data into output categories. An ANN is usually made up of three layers: input, hidden layers, and output. Each layer in a network has several neurons, depending on the application. Direct connections link each neuron to other neurons in the next layer. These links have a weight that represents the strength of the outgoing signal (Varoonchotikul, 2003). There are several types of networks that can be used in an ANN, and the best one depends on the problem and data available. Multi-layer perceptron (MLP) is perhaps the most commonly used neural network, according to Govindaraju (2000). Artificial neurons, or processing units, in MLP are organised in a layered configuration with an input layer, a processing ("hidden") layer (two hidden layers are used in complex topologies), and an output layer. By combining land-use change-conditioning variables, ANN-MLP was used to forecast the transitional probability model in this analysis. Elevation, slope, proximity to the urban area, agricultural land, sparse forest, scrubland, and water bodies were extracted from the 2000 and 2018 LULCs maps. The derived data was then processed using the Euclidean distance method in ArcGIS software to generate proximity parameters.

The CA model is a hybrid model that incorporates both cellular automata and Markov Chain concepts. The CA is made up of identical elements, such as cells, that are arranged in a regular and discrete space. The core theory behind CA is that any cell or pixel of LULC transition can be investigated using its existing situation and improvements in its surrounding cells or pixels. Using the Markov model, this incorporates transformation rules based on adjacent cells. The probabilities of original and later conditions are used to illustrate transition rules. The CA-Markov model creates a cell's or pixel's situation based on its initial state, the conditions of its neighbours, and a set of transformation rules. This depicts the complexities of transition using the proximity principle, which states that regions close to current areas of the same class are more likely to transition to different LULC classes. CA-Markov is known for its ability to forecast the complex dynamics of spatiotemporal patterns using a series of transformation rules. The model solves the problems associated with LULC transformations by acting on transition probabilities and using ANN-based suitability maps for each LULC category to generate accurate future predictions.

Sensitivity analysis

Random forest-based feature selection technique

Random Forest, developed by Breiman (Breiman, 2001), is one of the widely applied powerful ensemble supervised algorithms. This algorithm can be used to solve the regression problem, classification, and unsupervised learning. It has been extensively used in different aspects, such as natural hazard modeling, hydrology, LULC classification, and finance (Salam and Islam 2020; Chen et al., 2019; Talukdar and Pal, 2020).

The RF is the combination of the RSS and the bagging (Chen et al., 2019). The main advantages of the RF are the lower sensitivity to the multicollinearity test, which can handle the unbalanced and missing dataset. The RF model works in the following ways: (1) it produces subphases from the former data using the bootstrap re-sampling tool, which is equivalent to zero sizes in the former dataset, (2) it generates decision trees by applying the subphases, and (3) ultimately it produces the output by fusing the results of the prediction of all decision tree (*Ntree)* similarly. Chen et al., (2020) reported that the performance of the RF algorithm is mostly functioned by the number of *Ntree* and the features of the data which consisted of the subsets (*mtry*). A large value of *Ntree* led to higher times for modeling, whereas the low value produces more errors. In this work, RF was performed by the "Randomforest" package in R Studio 3.2 software. The RF was used for modeling fragmentation probability.

CART based sensitivity analysis

Classification and regression trees (CART) are considered as a non-parametric supervised machine learning method, which has been employed for classification and prediction (Nefeslioglu et al., 2010). The CART has become a popular data mining method because of its efficacy and easiness in solving a wide range of issues in the fields of agriculture, economics, engineering, and remote sensing research (Stenberg and Phillip, 1995; Waheed et al. 2006; Choubin et al. 2018). Generally, two types of CART have been used for modelings, such as classification trees and regression trees. For predicting a discrete variable, the classification trees have been employed while regression trees have been utilized for predicting a continuous parameter. It uses a stepwise tool to establish splitting rules (Stenberg and Phillip, 1995; Brieman et al., 1984). The classification tree is split into the space demarcated by the independent parameters dependent on the dependent parameters (Fakiola et al., 2010) and a continuous parameter e.g., regression tree, which forecasts the value of a dependent parameter based on numerous independent parameters. Contrasting the classification tree, the regression tree does not generate categories of dependent parameters. For this, the total sample has been partitioned into two or more homogeneous sets based on the most significant splitter in input variables. The key advantage of CART, a decision tree algorithm is its cross-validation nature (Brieman et al., 1984), which tries to identify overfitting problems that would lead to poor forthcoming forecasts. Another advantage of CART is that it has often triggered more precise predictions than other statistical tools (Kattan and Beck, 1995; Choubin et al. 2018). However, regression trees do not have pre-assigned categories, the output of this phase is a response value to each of the original values for the dependent parameter. In the present study, this method was implemented to derive the weight or power of influence of the different parameters for explaining the predicted fragmentation probability models. To the best of the authors' knowledge, this is the first study that employed CART based sensitivity analysis for finding the most sensitive parameters of the model.

Probability distribution function-based sensitivity analysis

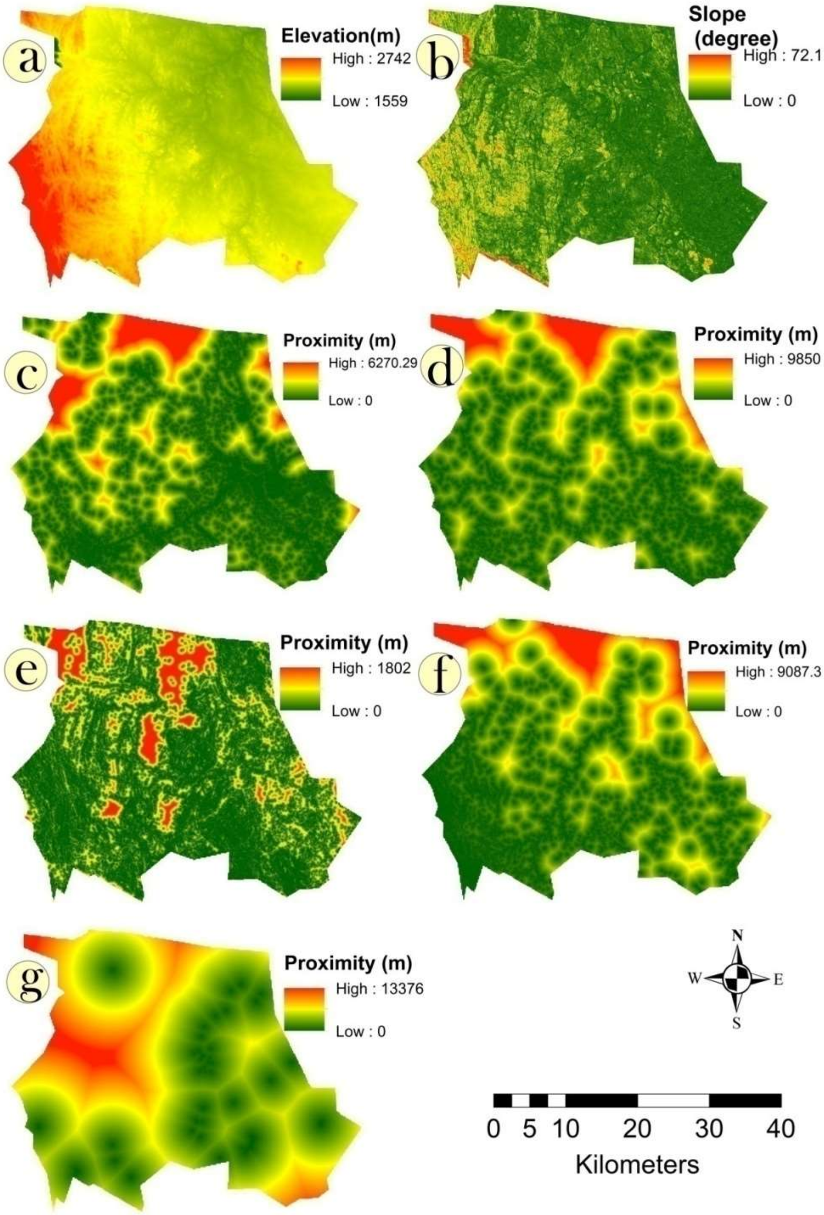
The probability distribution function (PDF) was calculated using the data obtained from the whole map of seven LULC changing parameters and land use suitability models for the data of 2000 and 2018. In our study, we used JASP (version 0.13.1.0) software to calculate PDF for seven parameters and land use suitability models

Pearson correlation-based sensitivity analysis

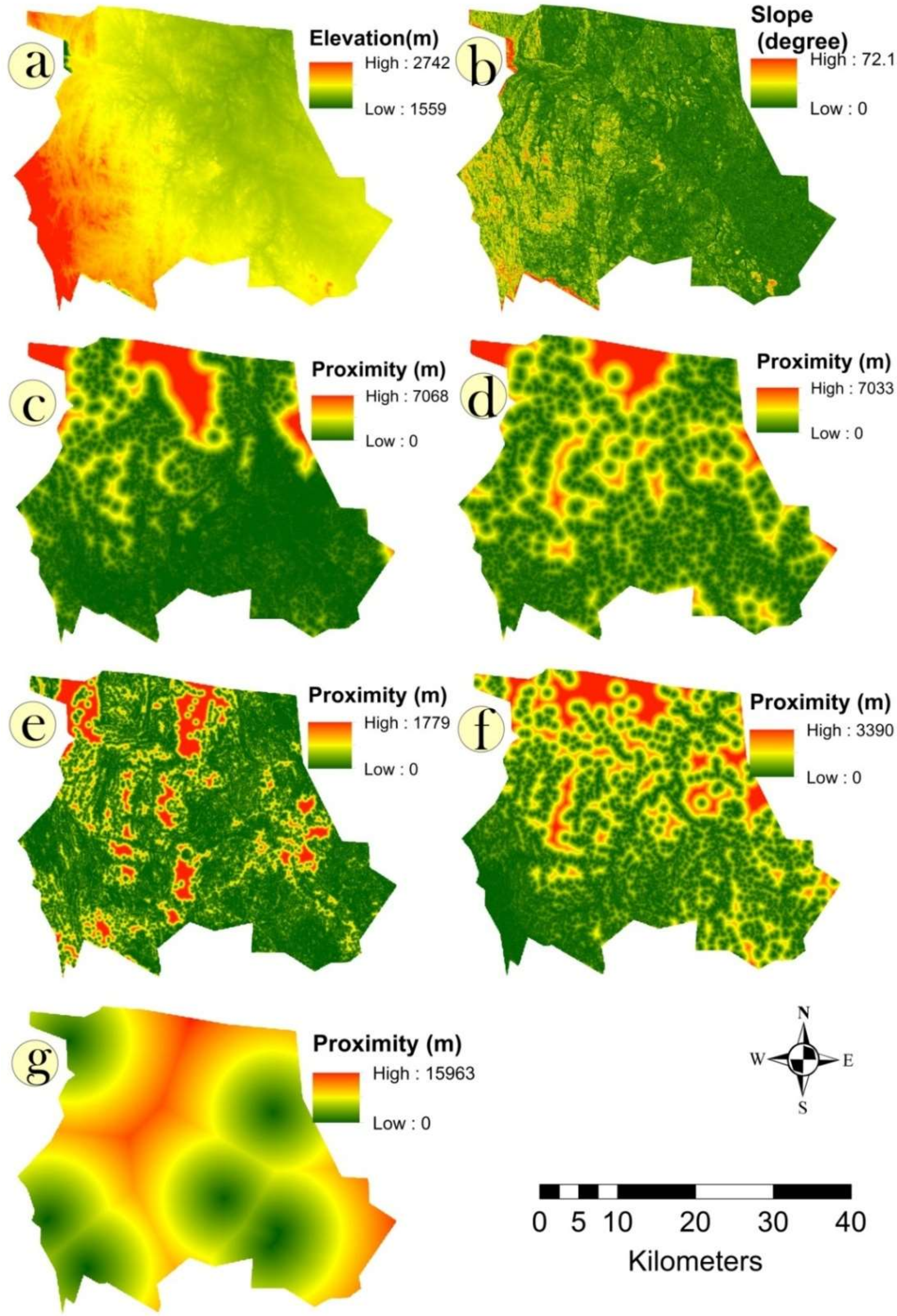
Pearson correlation coefficient was used to determine the association among seven LULC parameters and land use suitability methods for the data of 2000 and 2018. In this study, we used

SPSS (version 22) software to perform correlation coefficient analysis for seven LULC changing parameters and land suitability methods

List of figures

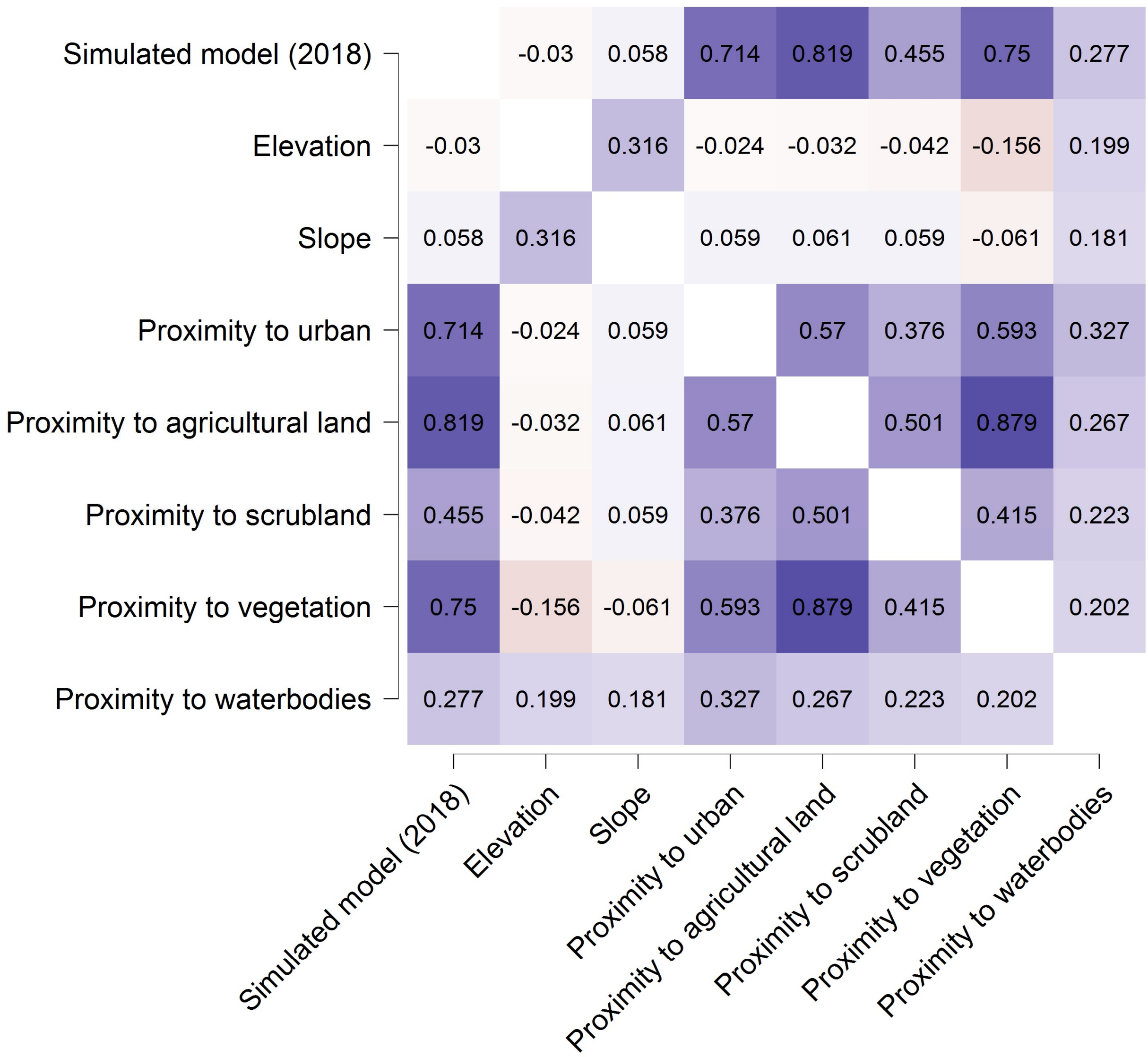


Supplementary figure 1 LULC change conditioning factors, such as (a) elevation, (b) slope, (c) proximity to urban area, (d) proximity to agricultural land, (e) proximity to scrubland, (f) proximity to sparse vegetation, (g) proximity to water bodies for predicting 2018 LULC

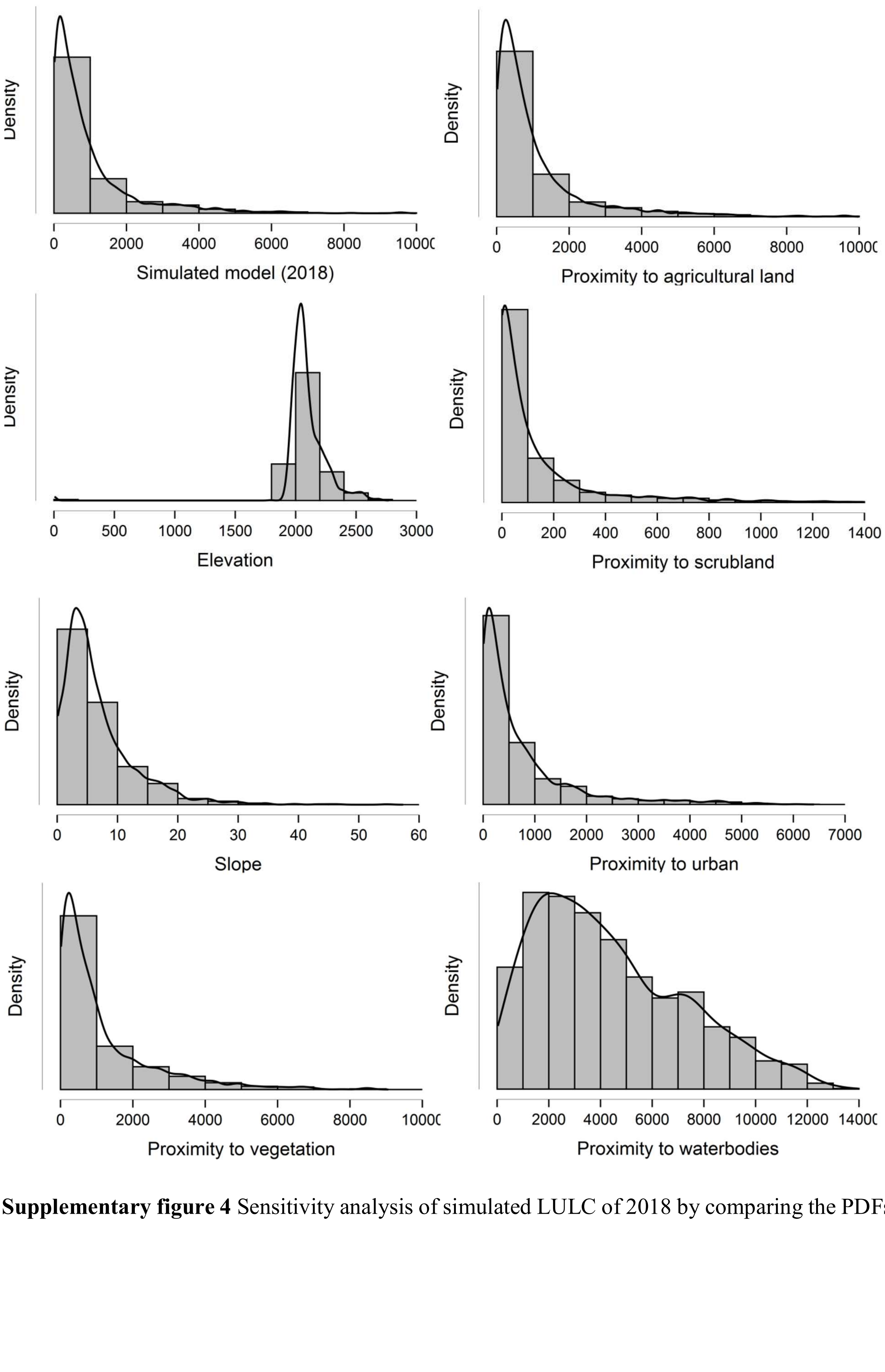


Supplementary figure 2 LULC change conditioning factors, such as (a) elevation, (b) slope, (c) proximity to urban area, (d) proximity to agricultural land, (e) proximity to scrubland, (f) proximity to sparse vegetation, (g) proximity to water bodies for predicting 2028 LULC.

Explanation: Lower elevation and slope provide favorable conditions for urban expansion. While regions, which located near the urban area have higher chances to be converted to the urban area. On contrary to this, high distance from core forest, water bodies, agricultural land have been the deciding factors for the transformation of the urban area. Supplementary figures 1 and 2 showed the LULC change conditioning factors for 2018 and 2028. The conditioning factors for predicting 2018 were prepared from the data of 2000, while the conditioning factors for predicting 2028 were generated from the data of 2018. However, supplementary figure 1 showed that the eastern and northeastern parts of the study area have lower elevation and slope, which is a suitable condition for constructing new habitat. While the northern part of the study area was very far from the existing urban area (6270m in 2000 and 7068 m in 2018). Therefore, the conversion to the urban area in this region is very difficult, even in the future. On the other hand, agricultural land, sparse vegetation, and scrubland in the eastern and northeastern parts of the study area were situated very close to the urban area in 2000. Therefore, these areas had a higher chance to be converted to the urban area in 2018. Similarly, the closeness of these areas toward the urban area increased significantly in 2018, which indicates the conversion of these areas to urban areas in 2028 (supplementary figure 2).



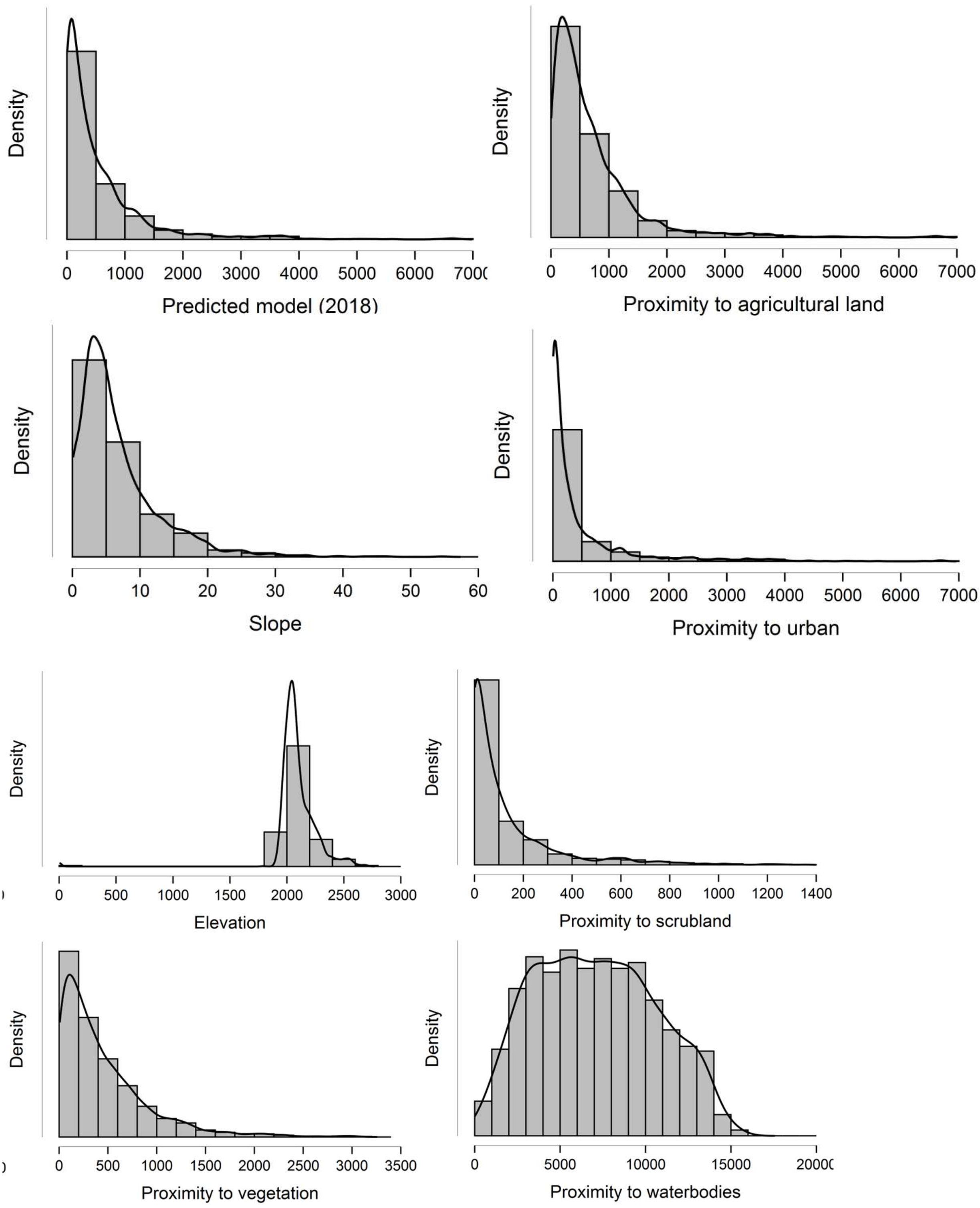
Supplementary figure 3 Correlation between simulated LULC for 2018 and LULC change conditioning parameters



Supplementary figure 4 Sensitivity analysis of simulated LULC of 2018 by comparing the PDFs

Explanation: In the present study, we prepared PDFs and correlation coefficient for seven independent parameters and LULC probability or suitable models of 2000 for simulating the LULC of 2018 to exploring the data pattern of the parameters and comparing the shape of the PDFs to analyze the sensitivity (Supplementary figures 3-4). Except for elevation, seven LULC change conditioning parameters have almost identical data distribution patterns. The shape of the PDFs showed that the data distribution of seven parameters is not normally distributed and skewed rightward (Supplementary figure 4). On the other hand, the shape of PDFs or data distribution pattern was quite different and normally distributed. However, we also prepared a PDF for the land-use suitability model of 2000 for simulating the LULC map of 2018. It also showed a similar shape of PDF or data distribution pattern. The shape of PDFs of proximate to urban, vegetation, and agricultural land were almost identical with the shape of PDF of a land suitable model. While the shape of PDFs of proximate to water bodies, and elevation were quite different from the shape of a land suitable model. Therefore, based on the analysis of the data distribution pattern, it can be stated that proximity to urban, vegetation, and agricultural land can influence or control maximally the land suitable model. We also computed the relation between the land suitability model and independent variables using Pearson's correlation coefficient. Supplementary figure 3 showed that proximity to the urban area had a higher correlation (r:0.714) at the significance level of <0.01. While the elevation had a very low correlation with the land suitable model.

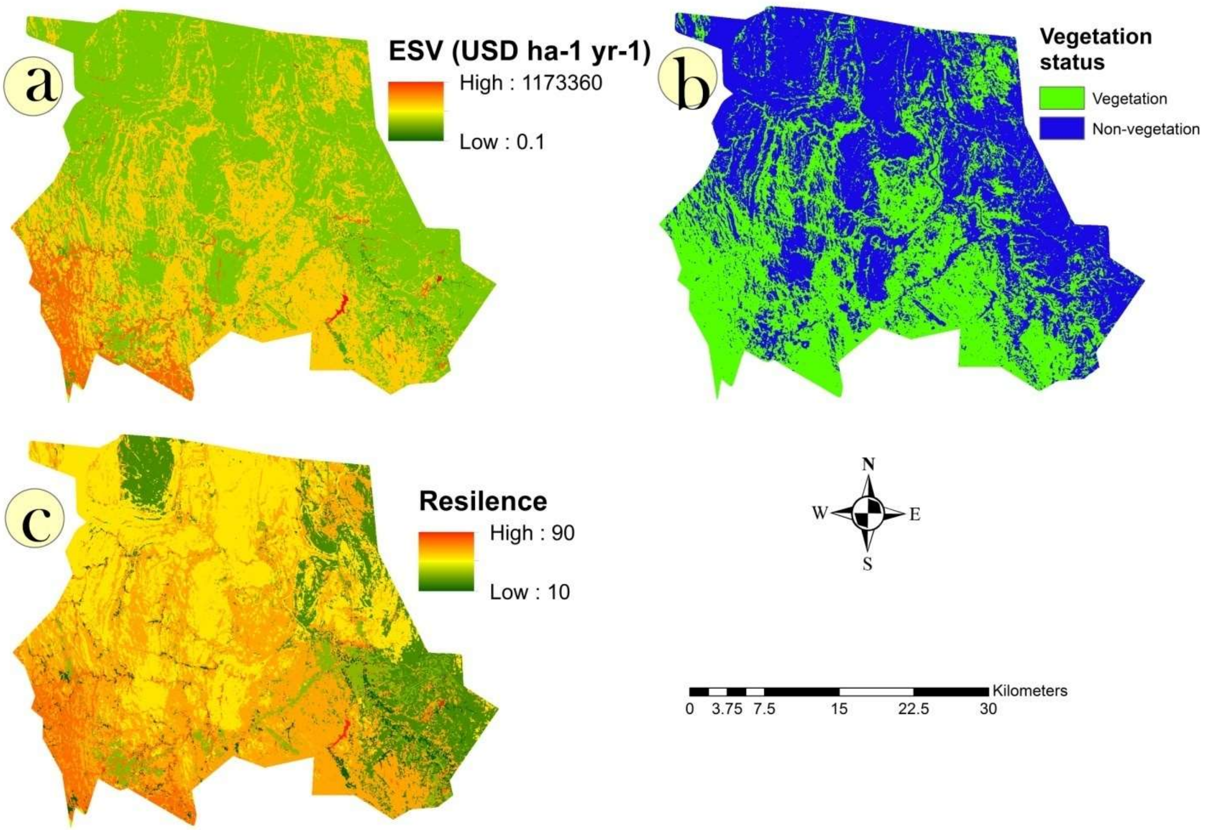
In the present study, we prepared PDFs for seven independent parameters and LULC probability or suitable models of 2018 for projecting the LULC of 2028 to exploring the data pattern of the parameters and comparing the shape of the PDFs to analyze the sensitivity (Supplementary figure 5). Except for elevation and proximity to water bodies, six LULC change conditioning parameters have almost identical data distribution patterns. The shape of the PDFs showed that the data distribution of seven parameters is not normally distributed and skewed rightward. On the other hand, the shape of PDFs or data distribution pattern of elevation and proximity to water bodies were quite different and normally distributed. However, we also prepared a PDF for the land-use suitability model of 2018 for projecting the LULC map of 2028. It also showed a similar shape of PDF or data distribution pattern. The shape of PDFs of proximate to agricultural land, vegetation, and urban were almost identical with the shape of PDF of a land suitable model. While the shape of PDFs of proximate to water bodies and elevation was quite different from the shape of a land suitable model.


Density
0
1000
2000
3000
4000
5000
6000
7000
Predicted model (2018)
Density
0
1000
2000
3000
4000
5000
6000
7000
Proximity to agricultural land
Density
0
10
20
30
40
50
60
Slope
Density
0
1000
2000
3000
4000
5000
6000
7000
Proximity to urban
Density
0
500
1000
1500
2000
2500
3000
Elevation
Density
0
200
400
600
800
1000
1200
1400
Proximity to scrubland
Density
0
500
1000
1500
2000
2500
3000
3500
Proximity to vegetation
Density
0
5000
10000
15000
Proximity to waterbodies

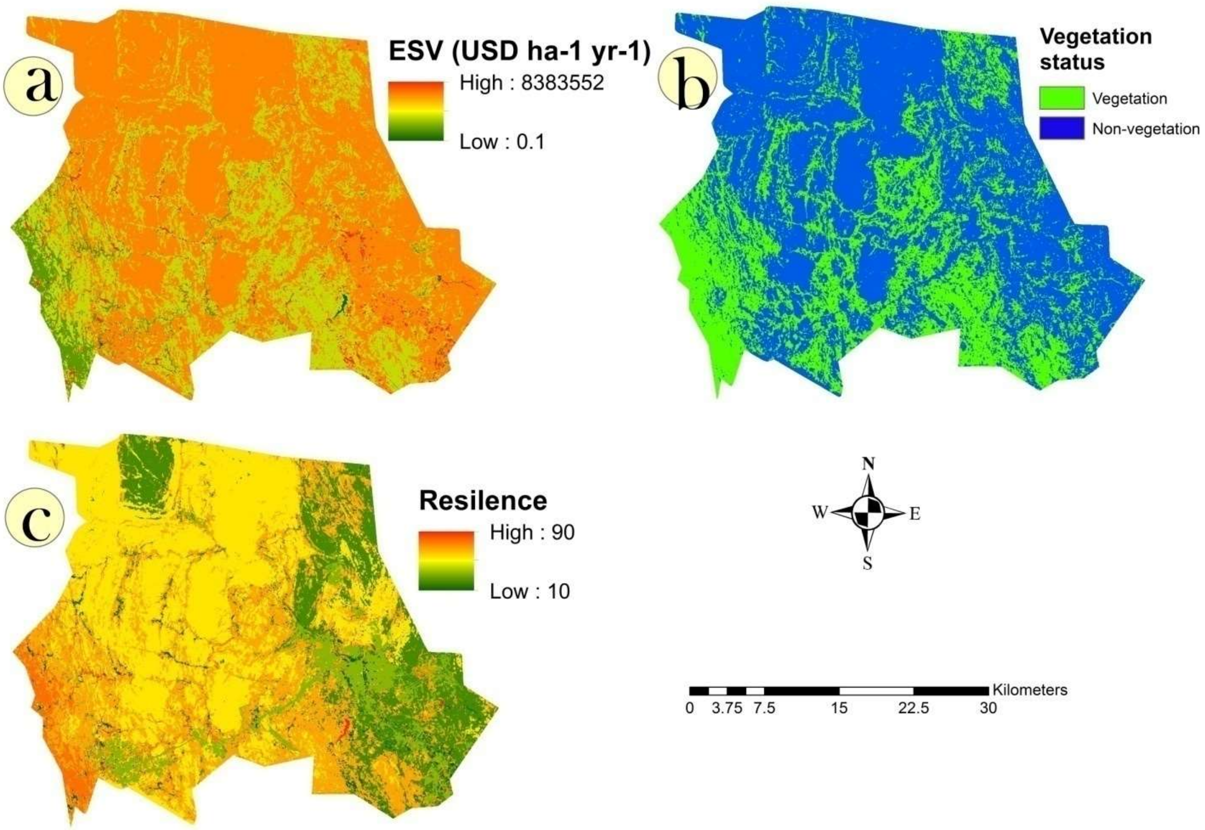
Supplementary figure 5 Sensitivity analysis of simulated LULC of 2028 by comparing the PDFs

Explanation:

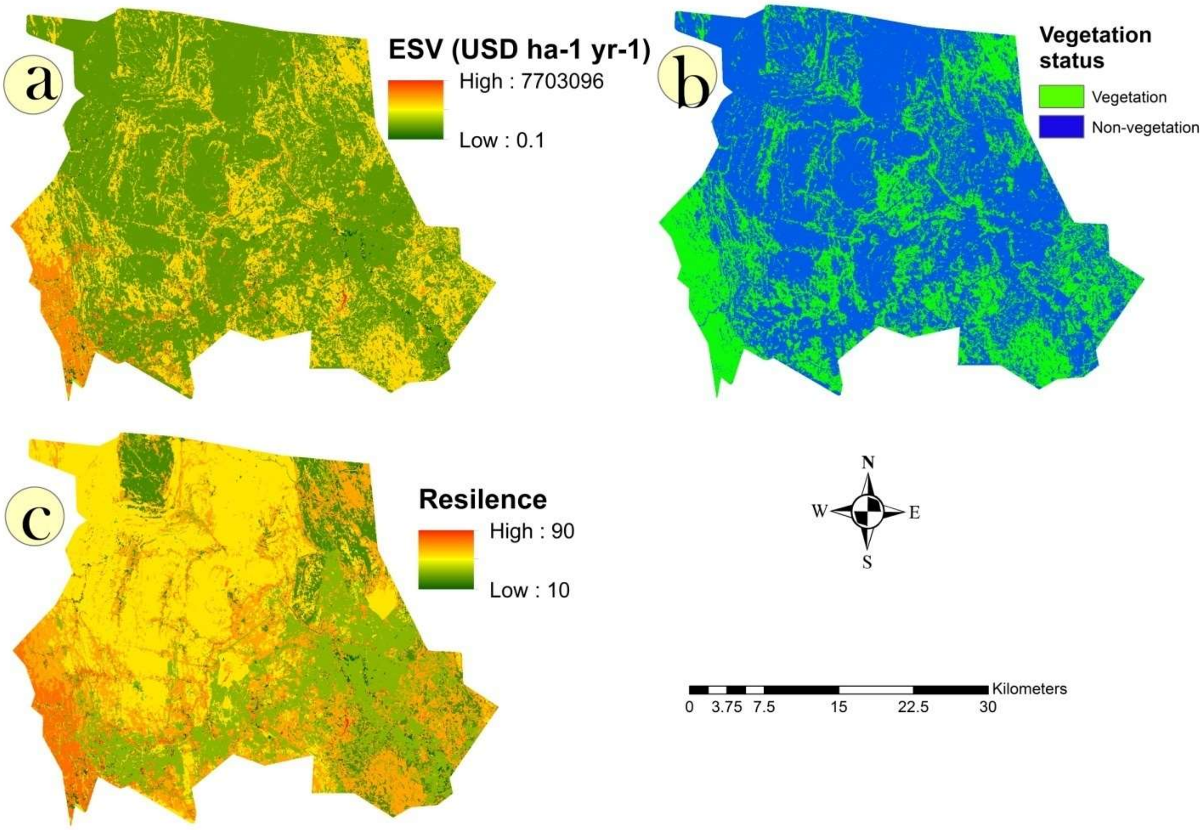
In the present study, we prepared PDFs for seven independent parameters and LULC probability or suitable models of 2018 for projecting the LULC of 2028 to exploring the data pattern of the parameters and comparing the shape of the PDFs to analyze the sensitivity (Supplementary figure 5). Except for elevation and proximity to water bodies, six LULC change conditioning parameters have almost identical data distribution patterns. The shape of the PDFs showed that the data distribution of seven parameters is not normally distributed and skewed rightward. On the other hand, the shape of PDFs or data distribution pattern of elevation and proximity to water bodies were quite different and normally distributed. However, we also prepared a PDF for the land-use suitability model of 2018 for projecting the LULC map of 2028. It also showed a similar shape of PDF or data distribution pattern. The shape of PDFs of proximate to agricultural land, vegetation, and urban were almost identical with the shape of PDF of a land suitable model. While the shape of PDFs of proximate to water bodies and elevation was quite different from the shape of a land suitable model. Therefore, based on the analysis of the data distribution pattern, it can be stated that proximity to agricultural, scrubland, vegetation, and urban can influence or control maximally the land suitable model. We also computed the relation between the land suitability model and independent variables using Pearson's correlation coefficient. Figure 5 showed that proximity to the urban area had a higher correlation (r: 0.569) at the significance level of <0.01. While the elevation had a very low correlation with the land suitable model.



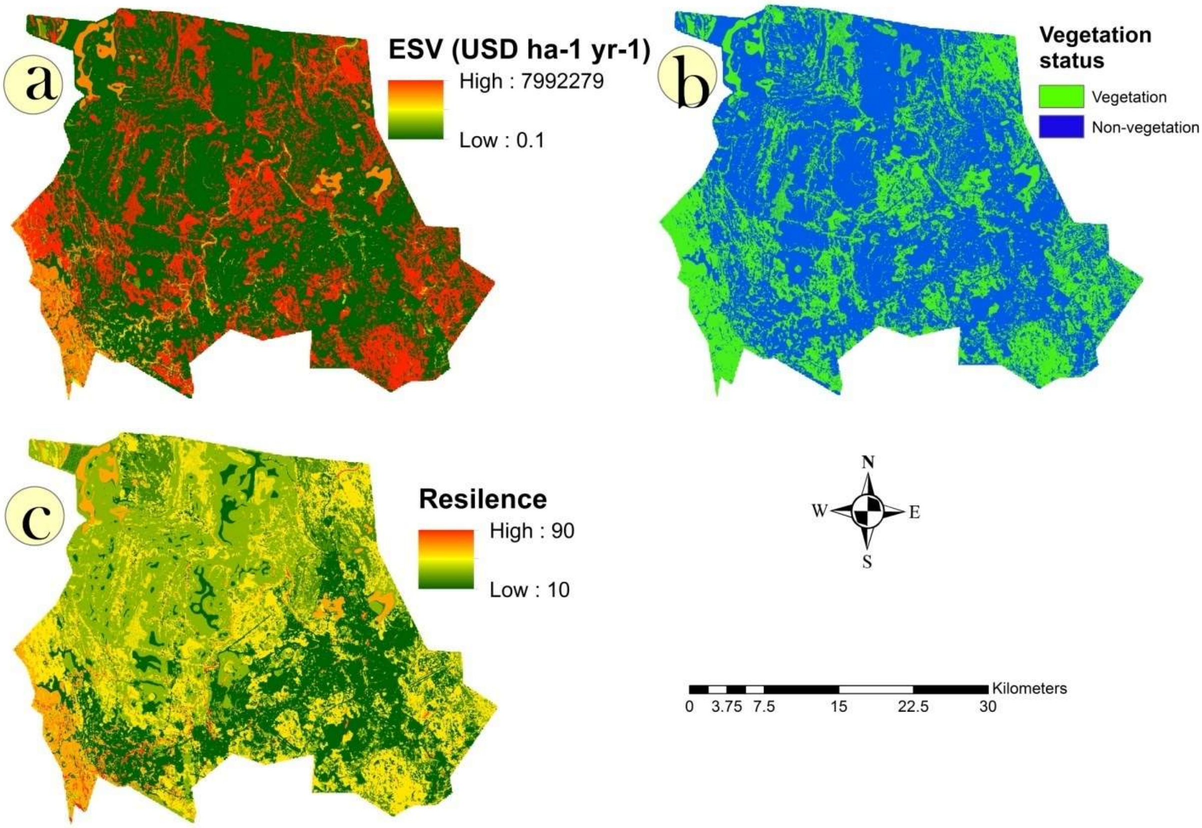
Supplementary figure 6 parameters for modeling ecosystem health conditions in 1990, such as (a) ESV, (b) vegetation status (NDVI), (c) resilience



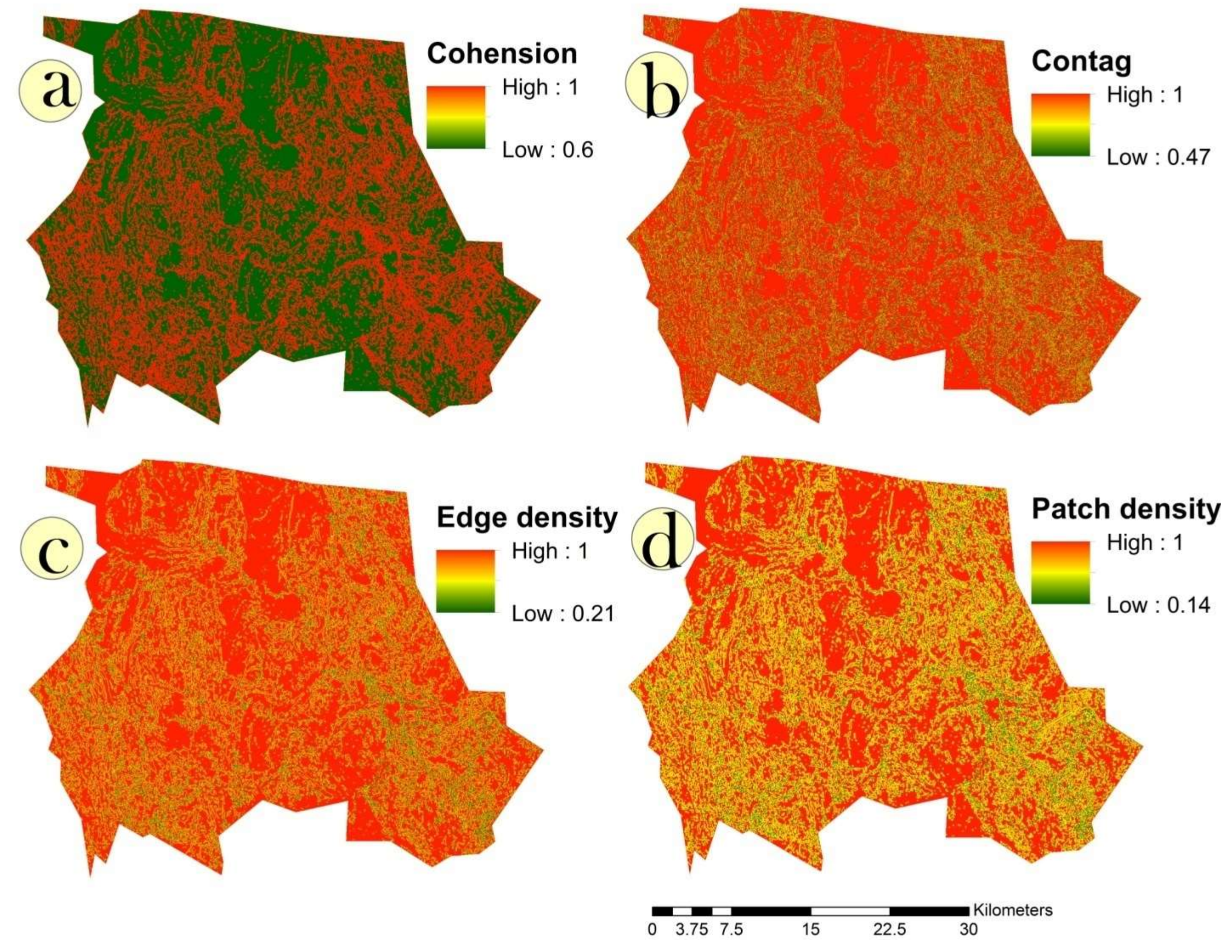
Supplementary figure 7 parameters for modeling ecosystem health conditions in 2000, such as (a) ESV, (b) vegetation status (NDVI), (c) resilience



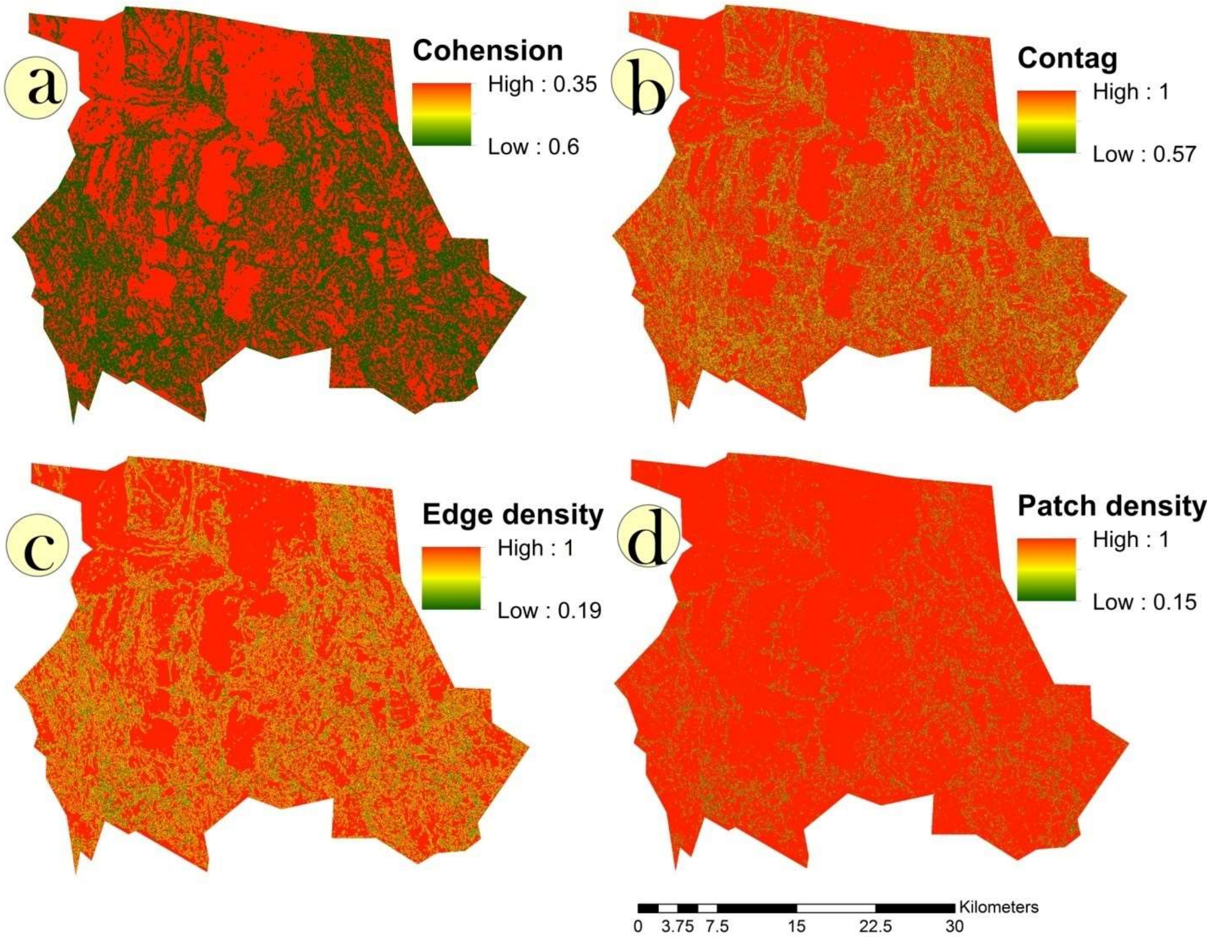
Supplementary figure 8 parameters for modeling ecosystem health conditions in 2018, such as (a) ESV, (b) vegetation status (NDVI), (c) resilience



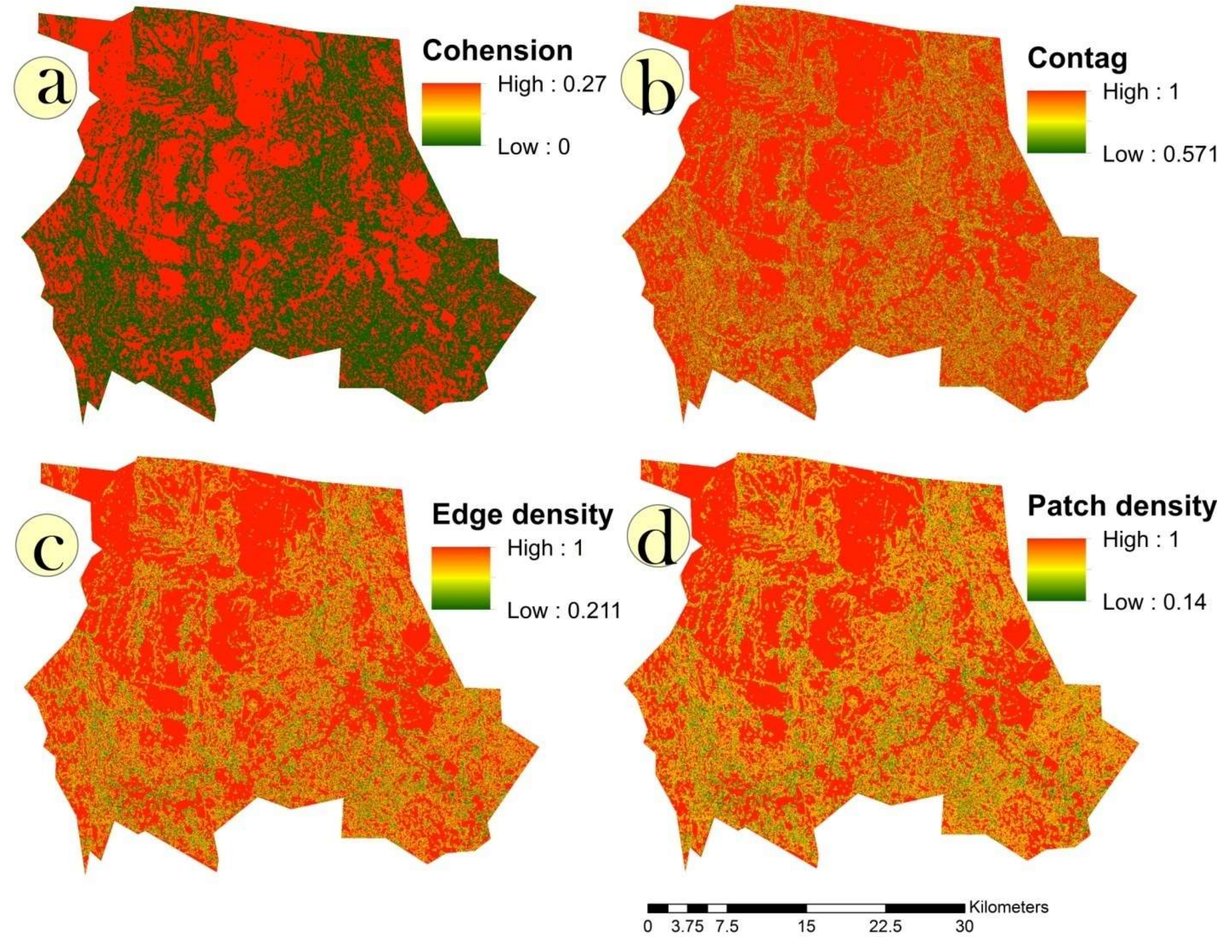
Supplementary figure 9 parameters for modeling ecosystem health conditions in 2028, such as (a) ESV, (b) vegetation status (NDVI), (c) resilience



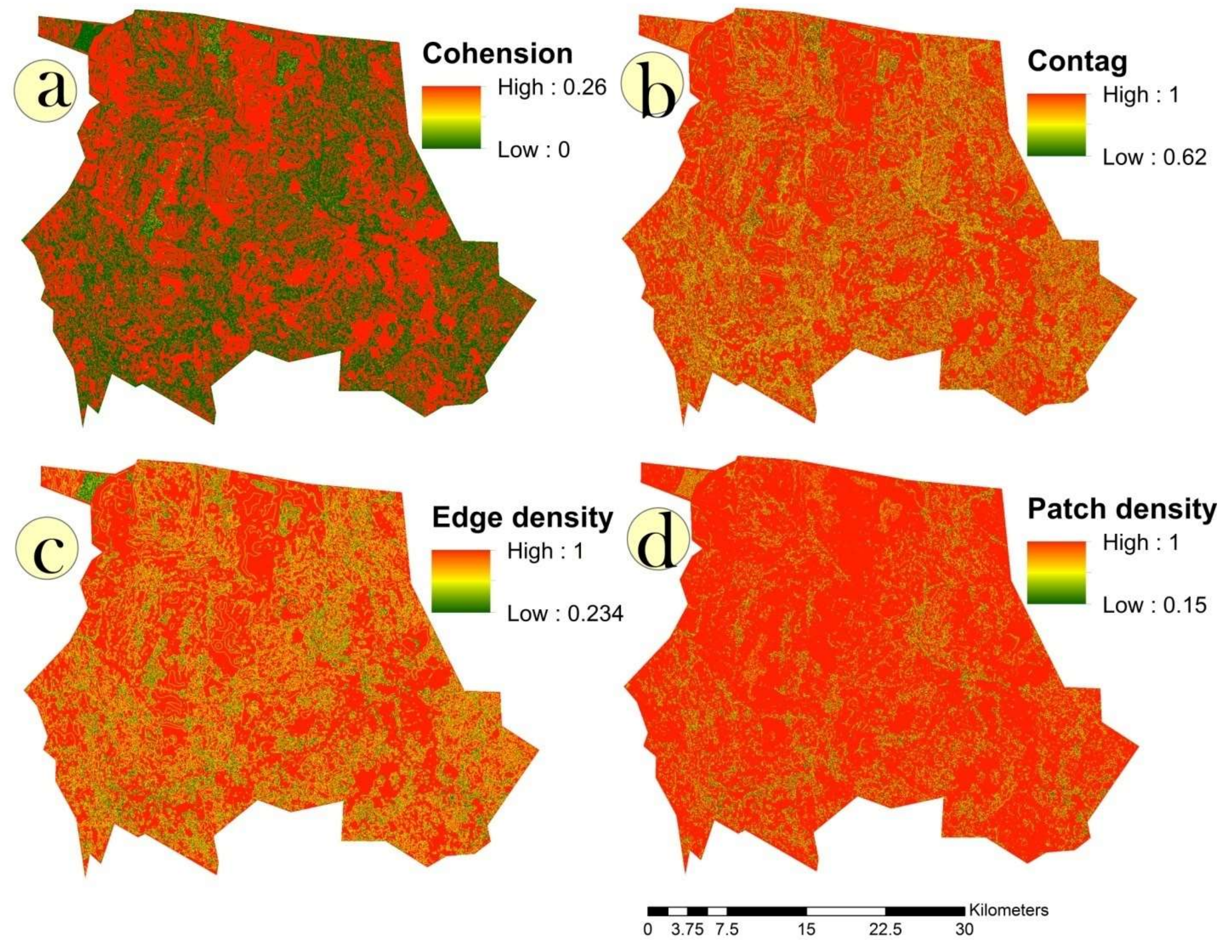
Supplementary figure 10 modeling of fragmentation conditions in 1990 using (a) cohesion, (b) patch contagion, (c) edge density, (d) patch density



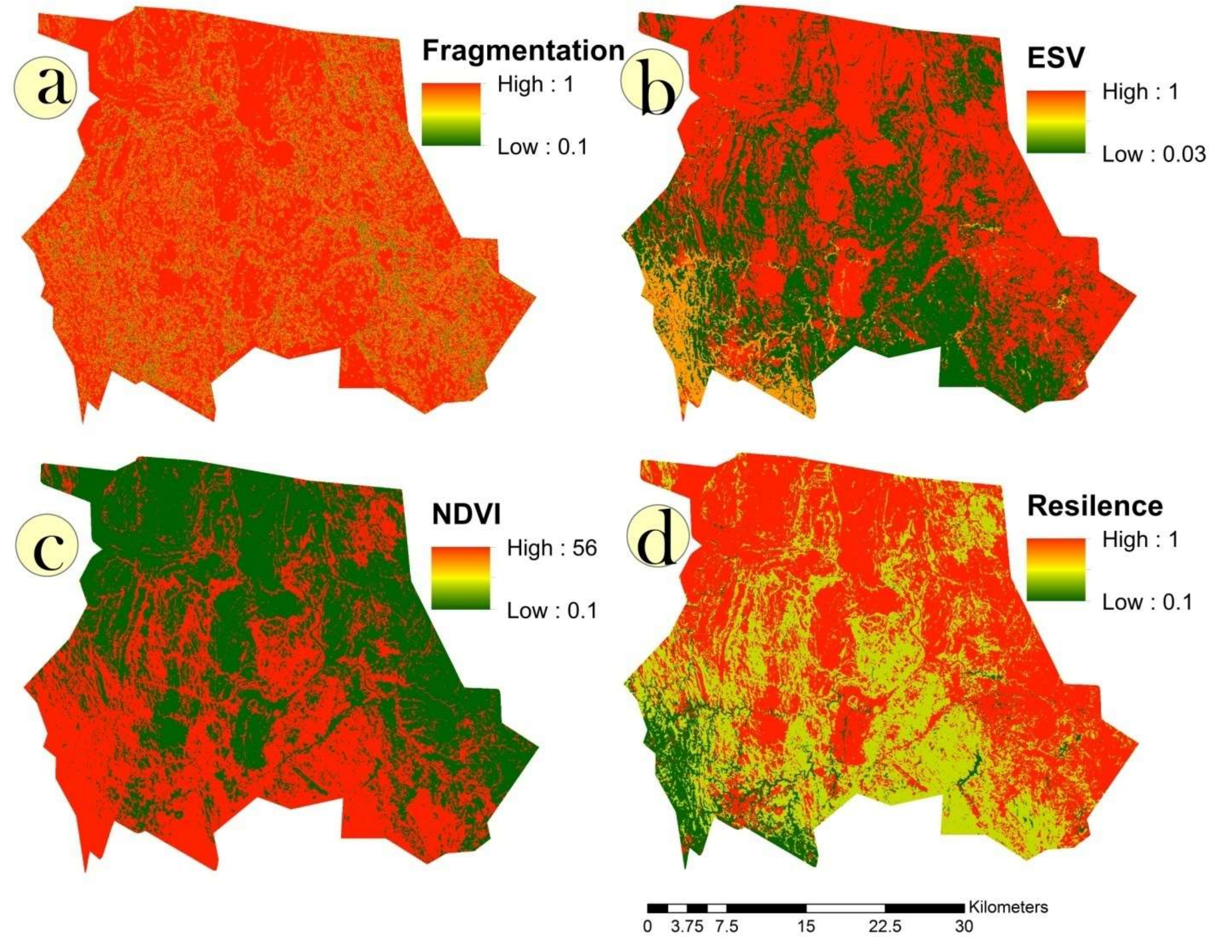
Supplementary figure 11 modeling of fragmentation conditions in 2000 using (a) cohesion, (b) patch contagion, (c) edge density, (d) patch density



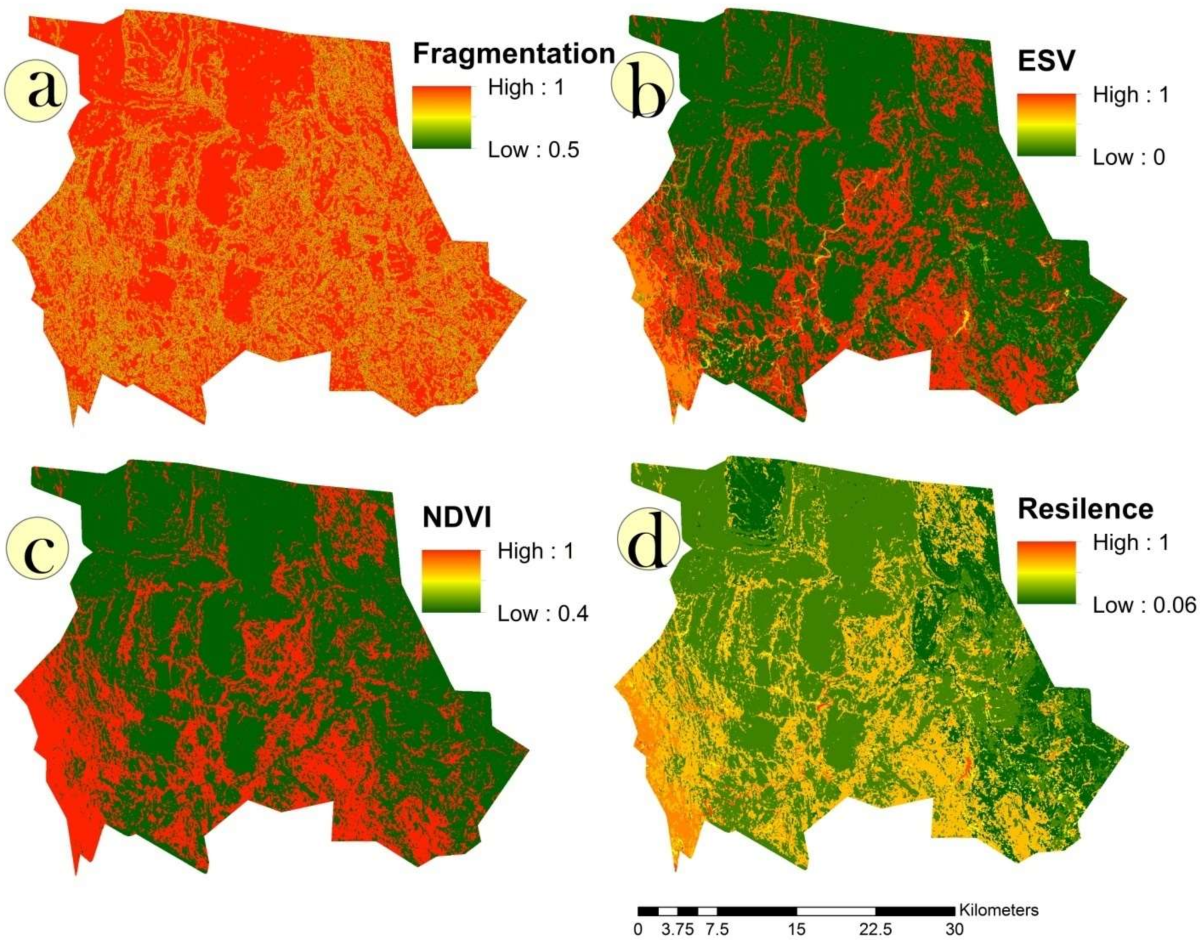
Supplementary figure 12 modeling of fragmentation conditions in 2018 using (a) cohesion, (b) patch contagion, (c) edge density, (d) patch density



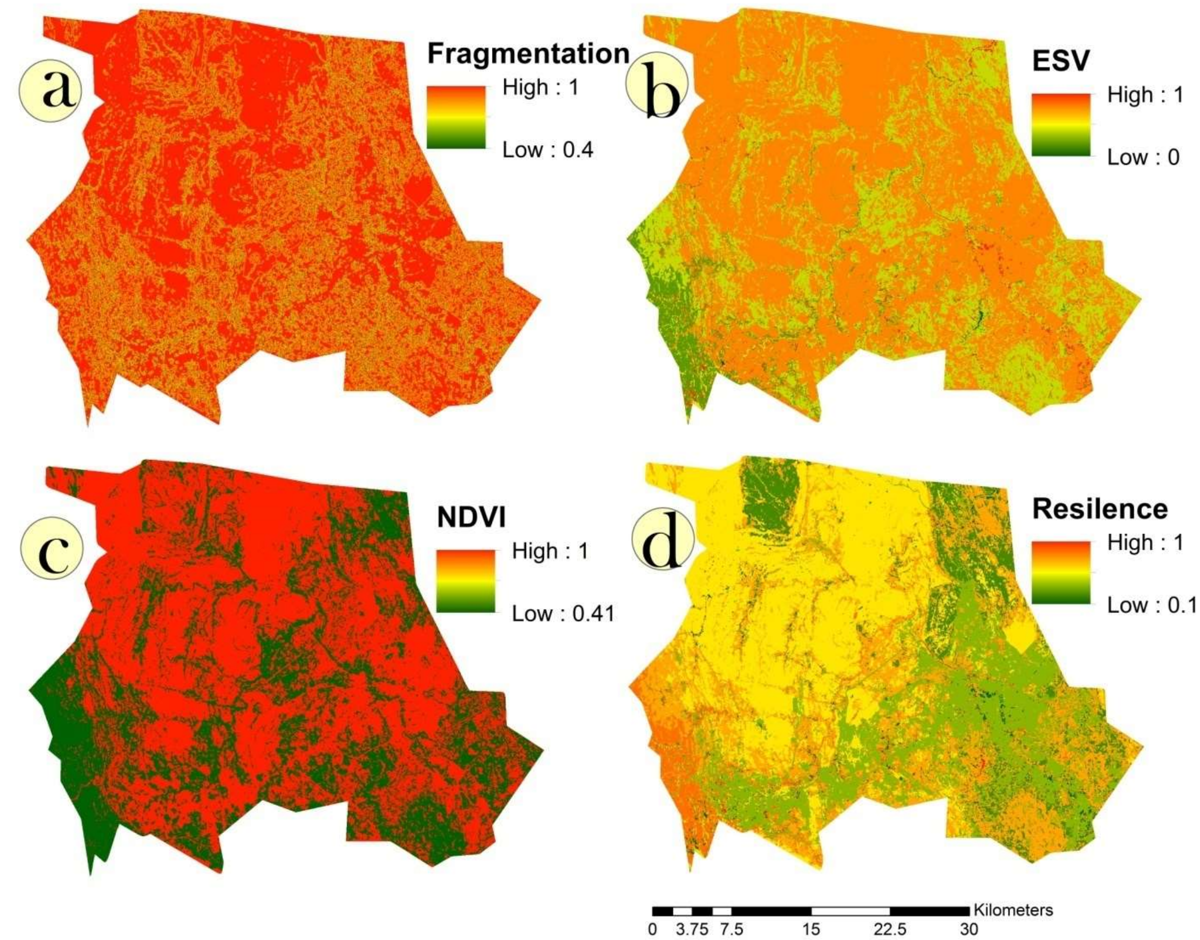
Supplementary figure 13 modeling of fragmentation conditions in 2028 using (a) cohesion, (b) patch contagion, (c) edge density, (d) patch density



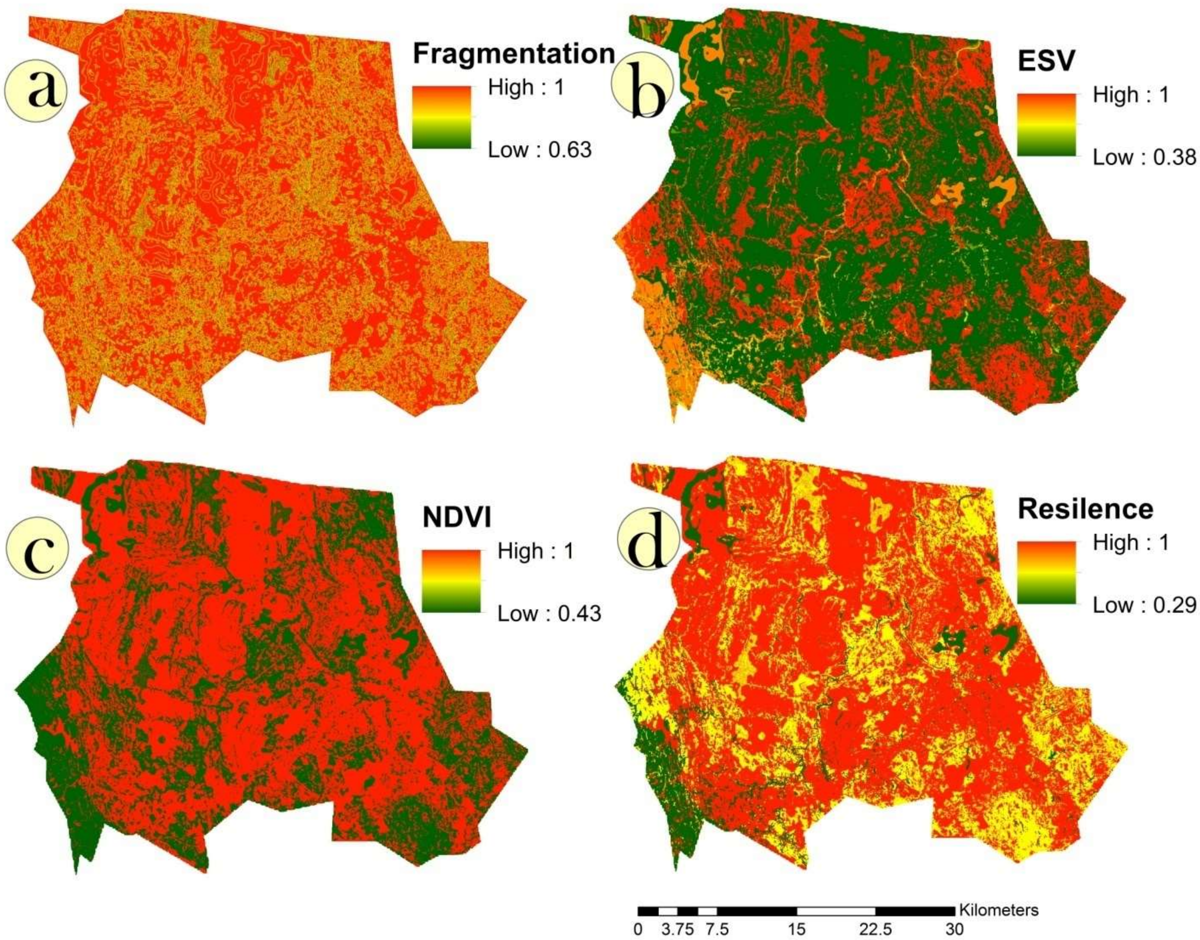
Supplementary figure 14 Final ecosystem health conditioning fuzzy crisp parameters for 1990, such as (a) fragmentation, (b) ESV, (c) NDVI, and (d) resilience



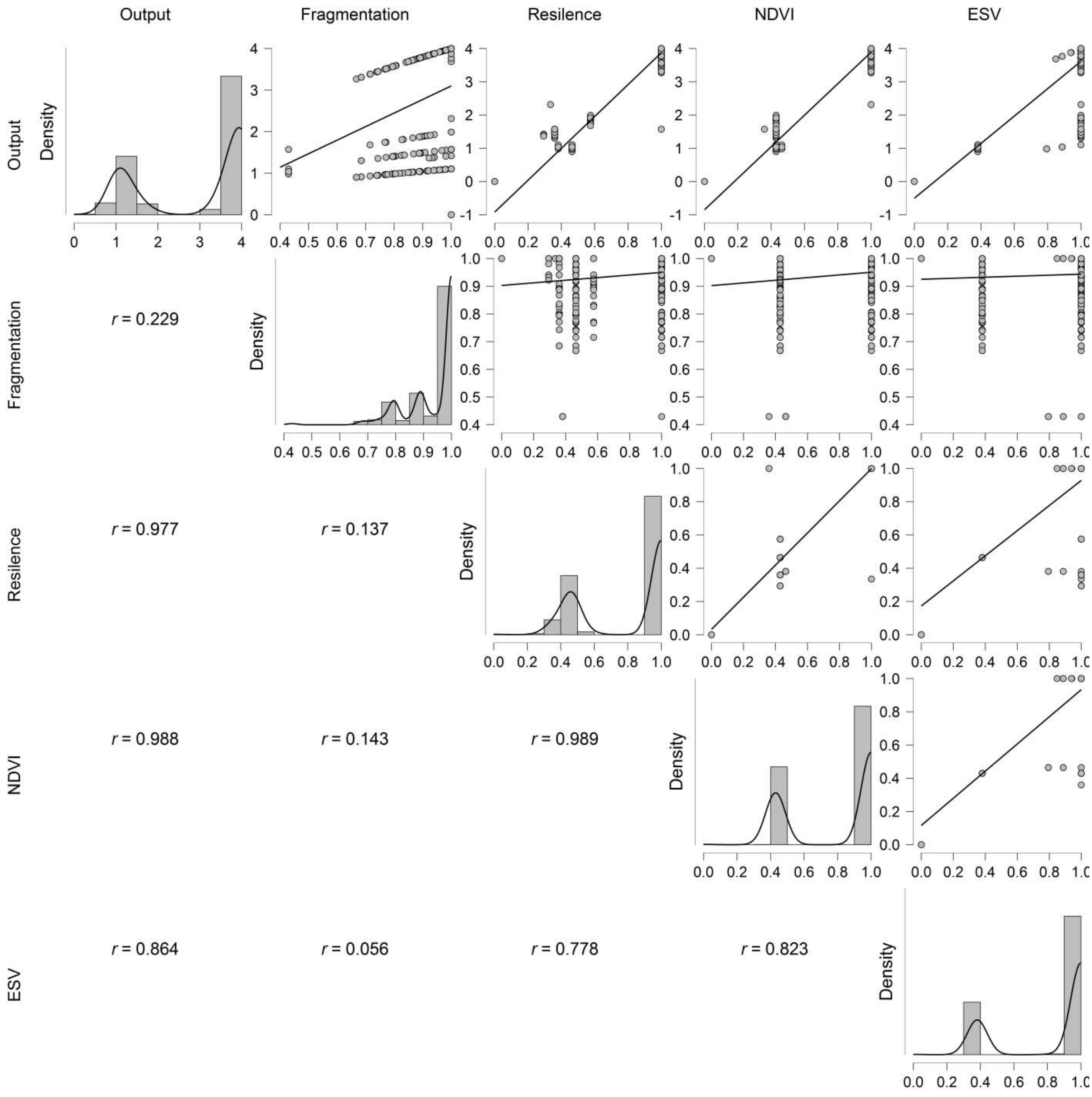
Supplementary figure 15 Final ecosystem health conditioning fuzzy crisp parameters for 2000, such as (a) fragmentation, (b) ESV, (c) NDVI, and (d) resilience



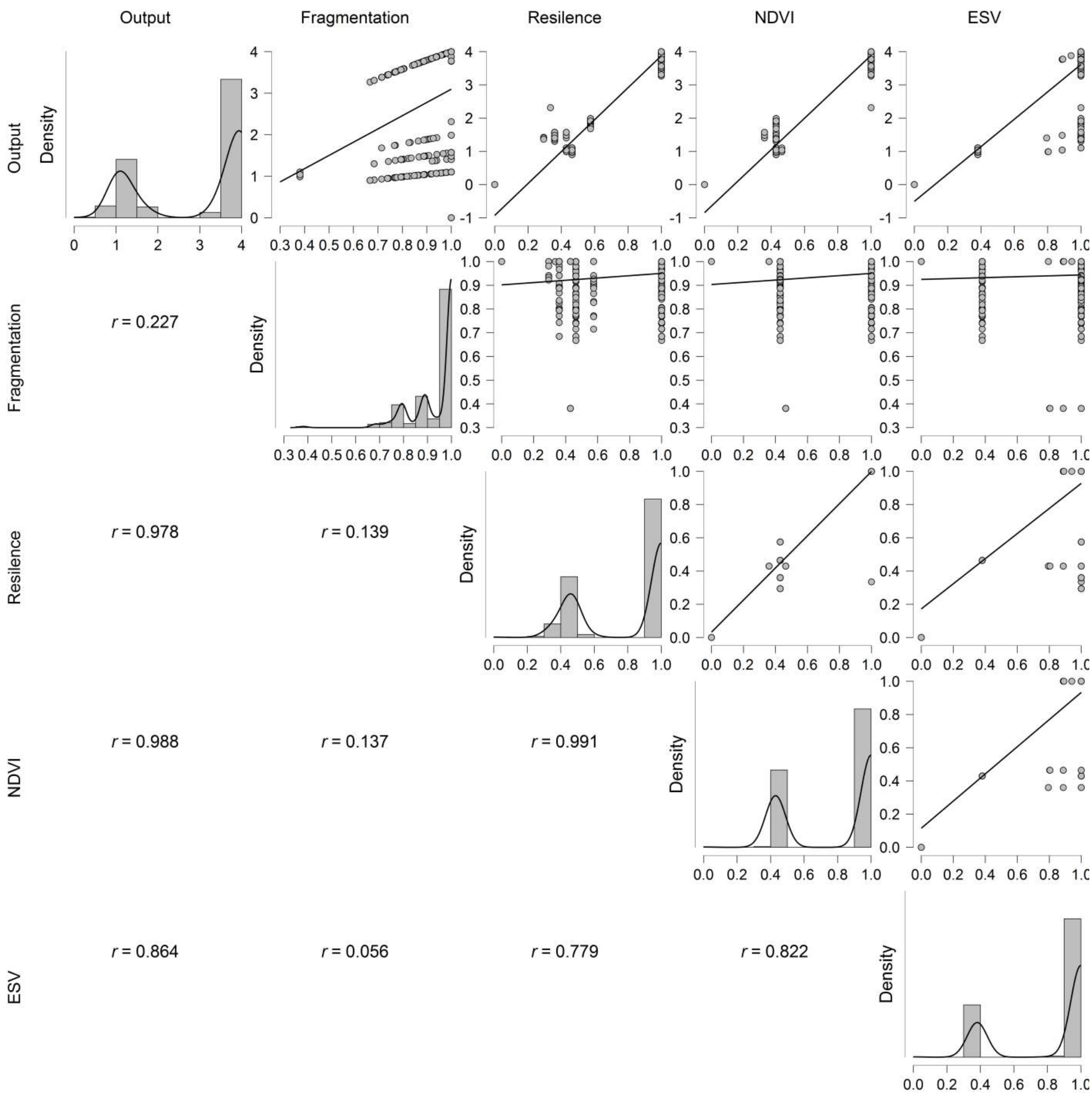
Supplementary figure 16 Final ecosystem health conditioning fuzzy crisp parameters for 2018, such as (a) fragmentation, (b) ESV, (c) NDVI, and (d) resilience



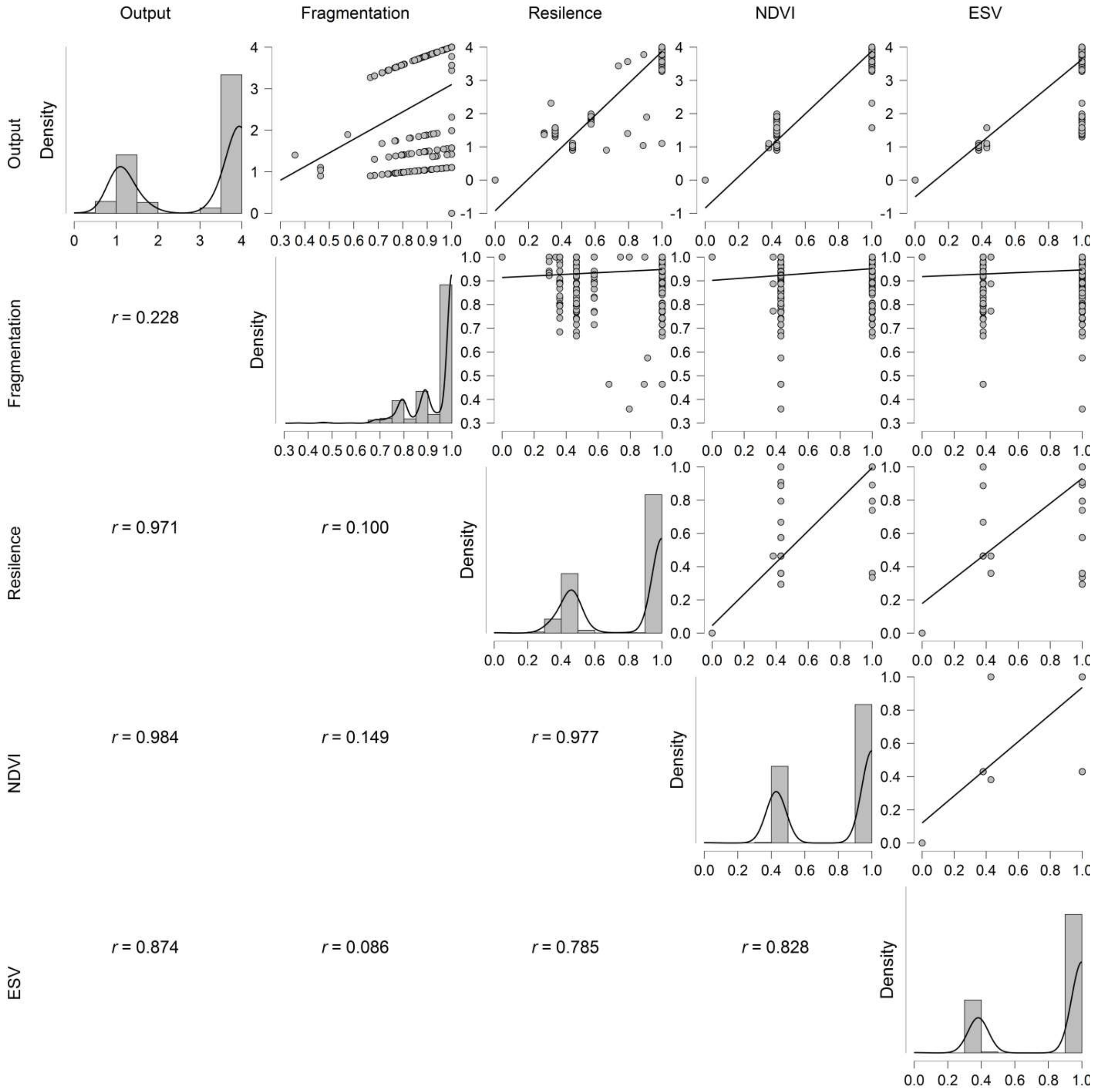
Supplementary figure 17 Final ecosystem health conditioning fuzzy crisp parameters for 2028, such as (a) fragmentation, (b) ESV, (c) NDVI, and (d) resilience



Supplementary figure 18 Sensitivity analysis for ecosystem health condition model of 1990



Supplementary figure 19 Sensitivity analysis for ecosystem health condition model of 2000



Supplementary figure 20 Sensitivity analysis for ecosystem health condition model of 2018

List of supplementary tables

Supplementary Table 1 ecosystem resilience coefficient of land use types (Xie et al., 2008; Peng et al., 2015; He et al., 2019)

Ecosystem types	Forest	HCG	MCG	LCG	water	Agricultural land	Built up	Sand	Bare land
	0.9	0.8	0.7	0.6	0.8	0.5	0.2	0.1	0.2

Supplementary Table 2 Coefficient for ESV estimation

LULC	Equivalent Biomes (Costanza)	Costanza et al. (1997) (1994 base price) [C97a] 1994 US\$ ha—1 yr—1	Costanza et al. (2014) (2007 base price) [C97b] 1997 US\$ ha—1 yr—1
Cropland	Cropland	92	5567
Sparse Vegetation	Grassland/rangelands	232	4166
Mangroves	Wetland/Tidal marsh	9990	193843
Waterbodies	/Mangroves Lakes/Rivers	8498	12512
Sandy Coast	Desert	0	0
Urban built-up	Urban	0	6661

Supplementary table 3 Transitional probability matrix between 1990-2000 LULC maps

Land use types	Built up	Water bodies	Dense vegetation	Sparse vegetation	Agricultural land	Scrubland	Bare soil	Exposed rockes
Built up	**0.7467**	0.0004	0.0008	0.0053	0.0052	0.0976	0.0360	0.1080
Water bodies	0.0218	**0.6671**	0.0033	0.0470	0.0371	0.1853	0.0172	0.0212
Dense vegetation	0.0021	0.0028	**0.2076**	0.6214	0.0239	0.0457	0.0000	0.0964
Sparse vegetation	0.0456	0.0003	0.0171	**0.4114**	0.0390	0.3053	0.0040	0.1772
Agricultural land	0.0502	0.0000	0.0091	0.0881	**0.2995**	0.3709	0.1680	0.0142
Scrubland	0.0481	0.0002	0.0009	0.0223	0.0172	**0.5827**	0.0857	0.2428
Bare soil	0.0960	0.0000	0.0001	0.0021	0.0088	0.0575	**0.8089**	0.0265
Exposed rockes	0.0305	0.0002	0.0005	0.0074	0.0011	0.0675	0.0344	**0.8585**

Explanation: Supplementary Table 3 showed the transitional probability matrix between 1990 and 2000. The LULC types, having the higher probability to be stabled or unchanged, were exposed rocks, bare soil, built-up area, and water bodies, accounting for 86%, 81%, 75%, and 67% of their original areas, respectively (Supplementary Table 3). Dense vegetation had the lowest probability to be unchanged, with significant areas transformed into sparse vegetation (62.1%) during the periods of 1990 and 2000. On the other hand, sparse vegetation was converted to scrubland and exposed rocks by 30.5% and 17.7%, respectively during this period. Similarly, agricultural land was transformed into scrubland and bare soil by 37% and 16.8% respectively. Scrubland was converted to exposed rocks by 24.3%. Therefore, based on the analysis of the transitional probability matrix, it can be stated that slight land-use changes have been taken place during these periods. Land use types, which can provide ecosystem services, were transformed into non-ecosystem services providing land-use types, which is not a good indicator. Even, the urbanization process was very insignificant, just 9% of bare soil was converted to the built-up area.

Supplementary table 4 Transitional probability matrix between 2000 and 2018 land use maps

Land use types	Built up	Water bodies	Dense vegetation	Sparse vegetation	Agricultural land	Scrubland	Bare soil	Exposed rockes
Built up	**0.8371**	0.0000	0.0033	0.0126	0.0115	0.0818	0.0151	0.0385
Water bodies	0.1238	**0.4377**	0.2412	0.0535	0.0048	0.0687	0.0040	0.0663
Dense vegetation	0.0814	0.0044	**0.3751**	0.3052	0.0952	0.0760	0.0034	0.0594
Sparse vegetation	0.0568	0.0000	0.0835	**0.5916**	0.0316	0.1910	0.0020	0.0435
Agricultural land	0.1498	0.0000	0.0477	0.1765	**0.2068**	0.3331	0.0667	0.0195
Scrubland	0.1899	0.0000	0.0039	0.0734	0.0157	**0.5367**	0.0535	0.1269
Bare soil	0.1963	0.0000	0.0015	0.0064	0.0167	0.3209	**0.4312**	0.0269
Exposed rockes	0.1287	0.0000	0.0011	0.0190	0.0017	0.0876	0.0073	**0.7546**

Explanation: Supplementary Table 4 showed the transitional probability matrix between the LULC maps of 2000 and 2018. Built-up area, exposed rocks, and sparse vegetation had the largest probability to be stabled or unchanged during this period by 83.7%, 75.4%, and 59.2% respectively. While the agricultural land had the lowest probability to be transformed by 20.7%. However, very small areas of built-up were converted to other land use types, such as exposed rocks (03.8%), scrubland (8.18%), and agricultural land (1.1%). Contrary to this, bare soil, scrubland, agricultural land, exposed rocks, and water bodies were converted to the built-up area by 19.6%, 18.9%, 14.9%, 12.8%, and 12.3% respectively. Except for the built-up area, the water bodies did not convert to any other land use types during this period. Similar to the built-up area, the scrubland has gained a huge amount of area from other land use types, such as sparse vegetation (19.1%), agricultural land (33.31%), and bare soil (32.09%). From this analysis, it can be concluded that the urbanization process has accelerated to capture other land use types in the process of urbanization and development. Therefore, this period can be called the period of urbanization in the study area.

Supplementary table 5 Transitional probability matrix between the LULC maps of 1990 and 2018

Land use types	Built up	Water bodies	Dense vegetation	Sparse vegetation	Agricultural land	Scrubland	Bare soil	Exposed rockes
Built up	**0.8367**	0.0000	0.0022	0.0138	0.0070	0.0583	0.0118	0.0702
Water bodies	0.1052	**0.3508**	0.2429	0.1681	0.0126	0.1052	0.0000	0.0152
Dense vegetation	0.0289	0.0028	**0.5341**	0.2963	0.0429	0.0169	0.0000	0.0781
Sparse vegetation	0.1528	0.0002	0.0637	**0.4780**	0.0397	0.1713	0.0011	0.0933
Agricultural land	0.1719	0.0000	0.0324	0.1572	**0.1532**	0.3995	0.0697	0.0161
Scrubland	0.2182	0.0000	0.0024	0.0469	0.0139	**0.4504**	0.0539	0.2143
Bare soil	0.2388	0.0000	0.0005	0.0061	0.0105	0.2878	**0.4296**	0.0266
Exposed rockes	0.1240	0.0000	0.0016	0.0119	0.0024	0.0839	0.0254	**0.7508**

Supplementary table 6 Transitional probability matrix between the LULC maps of 1990 and 2028

Land use types	Built up	Water bodies	Dense vegetation	Sparse vegetation	Agricultural land	Scrubland	Bare soil	Exposed rockes
Built up	**0.8380**	0.0000	0.0022	0.0132	0.0066	0.0663	0.0103	0.0635
Water bodies	0.6612	**0.2138**	0.0384	0.0212	0.0053	0.0483	0.0000	0.0119
Dense vegetation	0.0498	0.0034	**0.5080**	0.2987	0.0515	0.0203	0.0000	0.0684
Sparse vegetation	0.2236	0.0002	0.0628	**0.4378**	0.0385	0.1530	0.0010	0.0829
Agricultural land	0.2427	0.0000	0.0318	0.1417	**0.1405**	0.3686	0.0595	0.0152
Scrubland	0.2797	0.0000	0.0022	0.0441	0.0125	**0.4352**	0.0447	0.1817
Bare soil	0.2660	0.0000	0.0005	0.0152	0.0098	0.3085	**0.3794**	0.0206
Exposed rockes	0.1935	0.0000	0.0016	0.0545	0.0052	0.1256	0.0221	**0.5975**

Explanation: Supplementary Table 6 showed the transitional probability matrix between the LULC maps of 1990 and 2028. Built-up area, exposed rocks, and dense vegetation had the largest probability to be stabled or unchanged during this period by 83.8%, 59.7%, and 50.8% respectively. While the agricultural land had the lowest probability to be transformed by 14%. However, very small areas of built-up were converted to other land use types, such as exposed rocks (6.35%), bare soil (10.3%), and sparse vegetation (13.2%). Contrary to this, water bodies, sparse vegetation, agricultural land, exposed rocks, and bare soil were converted to the built-up area by 66.1%, 22.3%, 24.27%, 19.3%, and 26.6% respectively. Except for the built-up area, the water bodies did not convert to any other land use types during this period. Similar to the built-up area, the scrubland has gained a huge amount of area from other land use types, such as sparse vegetation (15.3%), agricultural land (36.86%), and bare soil (30.85%). From this analysis, it can be concluded that the urbanization process has accelerated to capture other land use types in the process of urbanization and development. Therefore, this period can be called the period of urbanization in the study area. Supplementary Table 6 represented the transitional probability matrix between the LULC maps of 2000 and 2028. The built-up area in the study area was the most consistent and stabled land-use type having the largest transitional probability to be unchanged by 96.5%. However, the built-up area has not changed to other land use types significantly. Scrubland and exposed have gained an insignificant amount of built-up area over 10 years. While the rapid urbanization process has been observed in the study area. The huge amount of other land use types, such as sparse vegetation, agricultural land, scrubland, bare soil, exposed land, and water bodies were converted into urban areas