

Spatio-Temporal Classification and Prediction of Land Use and Land Cover Change for the Vembanad Lake System, Kerala – a Machine Learning Approach

Parthasarathy Kulithalai Shiyam Sundar (✉ parthas1993@gmail.com)

National Institute of Technology Karnataka <https://orcid.org/0000-0003-0936-4065>

Paresh Chandra Deka

National Institute of Technology Karnataka

Research Article

Keywords: Google Earth Engine, Random Forest, Support Vector Machines, Machine Learning, Land use and land cover, CA-Markov chain analysis, LULC prediction

Posted Date: June 15th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-581788/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

1 **Spatio-temporal classification and prediction of land use and land cover**
2 **change for the Vembanad Lake System, Kerala – A machine learning**
3 **approach**

4

5 Parthasarathy Kulithalai Shiyam Sundar* and Paresh Chandra Deka

6 Department of Water Resources and Ocean Engineering, National Institute of Technology
7 Karnataka, Surathkal, Mangalore, India

* Contact: Parthasarathy K.S.S.
Mail id: parthas1993@gmail.com
Phone: +91-8973533057

8 **Abstract**

9 Land use and land cover (LULC) change has become a critical issue for decision planners and
10 conservationists due to inappropriate growth and its effect on natural ecosystems. As a result, the
11 goal of this study is to identify the LULC for the Vembanad Lake System (VLS), Kerala in the
12 short term, i.e., within a decade, utilizing two standard machine learning approaches, Random
13 Forest (RF) and Support Vector Machines (SVM), on the Google Earth Engine (GEE) platform.
14 When comparing the two techniques, SVM is classified at an average accuracy of around 84.5%,
15 while RF is classified at 89%. The RF outperformed the SVM in almost identical spectral classes
16 such as barren land and built-up areas. As a result, RF classified LULC is considered to predict
17 the Spatio-temporal distribution of LULC transition analysis for 2035 and 2050. The study was
18 conducted in Idrisi TerrSet software using the Cellular Automata (CA)-Markov chain analysis.
19 The model's efficiency is evaluated by comparing the projected 2019 image to the actual 2019
20 classified image. The efficiency was good with more than 94% accuracy for the classes except for
21 barren land, which might have resulted from the recent natural calamities and the accelerated
22 anthropogenic activity in the area.

23 **Keywords:** Google Earth Engine, Random Forest, Support Vector Machines, Machine Learning,
24 Land use and land cover, CA-Markov chain analysis, LULC prediction

25 **1. INTRODUCTION**

26 The natural and anthropogenic activities worldwide influence the land cover, resulting in
27 modifying its landscapes and the subsequent dynamics of natural processes (Silva et al. 2020).
28 Monitoring and assessing urban growth aid in the planning and utilization of natural resources for
29 the foreseeable future. Anthropogenic processes have altered almost half of the Earth's land
30 surfaces (Tayyebi and Pijanowski 2014). These changes are called land use and land cover (LULC)
31 changes. Over the last few decades, economic prosperity and population growth have resulted in
32 unplanned urbanization and industrialization to meet livelihood and job needs. These enormous
33 increases in need boosted the demand for critical infrastructures such as water supplies, sewage
34 services, and recreational activities. It also causes road congestion, pollution, climate change-
35 related problems, urban floods, and Urban Heat Island (UHI) impacts (Saxena et al. 2021). As a
36 result, LULC change is regarded as a critical environmental issue of global significance. Major
37 causes of biodiversity loss and related habitat loss are human-induced effects such as urbanization,
38 erosion, overgrazing, and resulting land degradation. Besides, nature also contributes to this
39 alteration (Halmy et al. 2015; Lambin 1997). The complex interaction of the factors like policy
40 management, human needs, environment, culture, and economics results in changing LULC.
41 Alteration in the LULC can significantly alter the water quality as the increase in urban and
42 agriculture results in nitrates and phosphates in the freshwater (Álvarez-Cabria et al. 2016;
43 Krishnaraj and Deka 2020).

44 Rapid urbanization in different parts of cities has been a severe threat to developed countries such
45 as India, Indonesia, Malaysia, and Sri Lanka. India is the second-most populous country in the
46 world, with 1.38 billion people. According to the Composite Water Management Index (CWMI)
47 report released by Niti Aayog in August 2019 under the Government Of India (GOI) stated that

48 many cities in India, such as Delhi, Bengaluru, and Chennai, may face "Day Zero" in the upcoming
49 years (Abijith et al. 2020). Due to the country's rapid population growth and subsequent
50 urbanization, policies for planning, analyzing, and tracking land-use transitions are required to
51 meet the people's basic needs. As a result, enormous areas of forest cover are being converted to
52 other land uses, causing severe soil erosion. Such rapid soil erosion will lead to a catastrophic
53 flood that will impact the residents downstream. Consequently, sustainable LULC is critical for
54 long-term livelihood and environmental improvement (Mishra et al. 2020).

55 **1.1 Role of Geospatial Technology**

56 Geospatial technology is a significant finding of humanity that has evolved since prehistoric times.
57 The main types of geospatial technology involve Remote Sensing (RS), Geographical Information
58 System (GIS) and Global Positioning System (GPS). RS is an excellent tool for regular monitoring
59 and assessing the LULC change in the natural ecosystem, the evolution of landforms by analyzing
60 the geomorphological changes. Likewise, it has played a major role in geography, geology, and
61 environment for the researchers and scientists (Ghosh et al. 2017). It is also more cost-effective
62 and less time-consuming than the traditional method. It has very long spectral bands, and good
63 spatial resolution helps distinguish a significant change in land cover (Abijith et al. 2020). RS and
64 GIS gained much attention from the government and the public due to their ability to map the
65 vulnerabilities during the disaster on a larger scale. The primary goal of GIS is to make it useful
66 to the public and improve their community's needs (Parthasarathy and Deka 2019).

67 One of the most extensively utilized sources for analysis is satellite images. In 1972 Earth
68 Resources Technology Satellite (ERTS-1) was launched and renamed "Landsat" in 1975. Landsat
69 has launched eight series of satellites and has data contiguity for almost 50 years. As of now,
70 Landsat 7 and 8 are functional. This series of satellites has become one of the important long-term

71 freely available data for the civilian purpose and used widely in fields like coastal monitoring
72 (Parthasarathy et al. 2020), LULC (Shi and Yang 2015), Vegetation phenology (Senf et al. 2017),
73 and hydrology (Abijith et al. 2020). Thus, Landsat offers a deeper understanding of LULC changes
74 for better decision-making and resource management. Several methods for detecting change using
75 remotely sensed data have been established during the last three decades (Hua et al. 2014; Jat et
76 al. 2017; Rienow and Goetzke 2015; Saxena and Jat 2020; Serasinghe Pathirana et al. 2018).

77 **1.2 Characteristics of Google Earth Engine**

78 Google earth engine (GEE) is a multi-petabyte collection of geospatial data co-located with high
79 performance and intrinsically parallel computation service. In recent years it has been in RS big
80 data processing spotlight. GEE is an interactive development environment (IDE) to enable rapid
81 visualization and analysis controlled through an application program interface (API) accessed via
82 the internet. It contains large publicly available datasets of satellite and aerial imagery in both
83 optical and non-optical wavelengths. It consisted of the most freely available RS imagery of entire
84 Landsat archives and Sentinel-1 and Sentinel-2 (Gorelick et al. 2017). The major advantage of
85 GEE is that it provides large sets of data, including land cover, environmental variables, weather,
86 and climate forecast, which is already pre-processed to access the data. Besides, the raw data is
87 also pre-processed, cloud removed, and mosaicked in the GEE to reduce the computational time.
88 Earth engine code editor is handled by the client libraries handled via JavaScript and Python
89 (Gomes et al. 2020). GEE utilizes the parallel processing technique using MapReduce architecture.
90 It is a method to process a large amount of data into several smaller chunks in several machines.
91 Thus, the data processed as several chunks were recompiled for the final result. As the data can be
92 accessed through the API, it is less labor-intensive and requires less storage space to save the data,
93 and also, its simple yet effective architecture does not require high power computing machines

94 (Noi Phan et al. 2020). Thus, the use of GEE has significantly increased within the RS community
95 (Tamiminia et al. 2020).

96 **1.3 CA-Markov chain analysis**

97 As the population grows, so does the need for land. As the demand for land increases, it eventually
98 increases urban. This further increases urban areas on LULC, causing disturbance in the ecosystem
99 affecting sustainability (Aburas et al. 2018). Thus, Cellular Automata (CA) – Markov chain model
100 is used to understand the factors affecting the Spatio-temporal distribution of LULC and to predict
101 the future LULC changes. It is one of the most widely used models to predict the LULC change
102 (Ozturk 2015). This model was created using Idrisi TerrSet's Land Change Modeler (LCM). This
103 software aids in analyzing and developing models in the event of stable land cover as opposed to
104 a rapidly changing environment. LCM facilitates the comparison of LULC categories and the net
105 change observed by every class and the contributor to the net change observed by every other
106 LULC category (Hamad et al. 2018). Predicting LULC transition has been used in several
107 applications, including environmental planning by modeling rural development and urban growth,
108 identifying conservation target areas and establishing alternate conservation strategies, analyzing
109 the dynamics of changing agriculture, and simulating rangeland dynamics under various climate
110 change scenarios (Halmy et al. 2015).

111 To summarise, it was found that GEE aids in the study of LULC shift in a cost-effective and time-
112 consuming manner, and it is commonly used in the literature (Agarwal and Nagendra 2019; Gomes
113 et al. 2020; Noi Phan et al. 2020; Sidhu et al. 2018; Tamiminia et al. 2020; Tassi and Vizzari 2020;
114 Xing et al. 2021). CA – Markov chain model helps in determining the future land cover change
115 and their land-use patterns for decision-makers to provide sustainable development (Aburas et al.

116 2018; Ansari and Golabi 2019; Bose and Chowdhury 2020; Faichia et al. 2020; Fu et al. 2018;
117 Ghosh et al. 2017; Gidey et al. 2017; Halmy et al. 2015; Hamad et al. 2018; Ozturk 2015).

118 In this work, LULC was classified in the GEE platform using two standard machine learning
119 approaches, the Random Forest (RF) classifier and the Support Vector Machines (SVM). We
120 attempted to understand the performance of both models using the same collection of training and
121 validation points. Then, using the CA – Markov chain model, forecast the future Spatio-temporal
122 LULC transition analysis and comprehend the changes in LULC for the years 2035 and 2050.

123 **2. MATERIALS AND METHODS**

124 **2.1 Study Area**

125 The field of study comprises six watersheds, including Periyar, Muvattupuzha, Meenachil,
126 Manimala, Pamba, and Achenkovil, draining into the Vembanad Lake System (VLS) in Kerala,
127 India (Fig. 1). It falls in six districts, namely Ernakulam, Idukki, Kottayam, Alappuzha,
128 Pathanamthitta, and Kollam. It is located in between latitude of 9° 1' 9" N to 10° 20' 22" N and
129 longitude of 76° 16' 47" E and 77° 24' 43" E comprising of area 12,183 Km². An area of 398.12 Km²
130 is situated below the MSL, and a total of 763.23 Km² is located below 1 meter MSL. The eastern part
131 of the study consists of Western Ghats, whereas the Arabian Sea binds the western part. The area in
132 and around the VLS flooded recently in the 2018 and 2019 Kerala floods. Lots of built-up and
133 agricultural land flooded completely, causing lots of damage to the resources. In which the state took
134 quite a lot of damage due to this unprecedented rainfall. It has a moist, tropical climate, with an
135 average of 150 rainy days each year. The southwest summer monsoon, as well as the northeast
136 winter monsoon, brings rain to this state. From May to August, the southwest monsoon season
137 takes place. Whereas between September and November, the Northeast winter monsoon arrives,

138 bringing cool winds with it. As a result, the state receives around 3000 mm of rainfall on average.
139 (CWC 2018). The mean annual temperature in the state's coastal lowlands is around 25°C-27.5°C,
140 compared to 20°C-22.5°C in the eastern highlands. The geology of the study is characterized by
141 charnockites, charnockitic gneisses, and pyroxene-bearing granulite in the Western Ghats and
142 central areas of the study. The sedimentary deposits of the Neogene and Quaternary periods
143 dominate the western sections of the sample. Alluvial deposits from the recent past can be found
144 along the coast.

145 <Fig. 1 Keymap of the study>

146 **2.2 Data Sources and Preparation**

147 The data prepared from various sources have shown in Table 1. Landsat 7 and Landsat 8 images
148 were mainly used in the GEE environment for LULC classification. Landsat 7 was chosen for
149 2009, while Landsat 8 was chosen for 2013, 2015, 2017, and 2019. A cloud mask was used to
150 choose the image covering the months of January to December for their respective years. Digital
151 Elevation Model (DEM) is prepared using the ASTER data in the ArcGIS environment. Slope and
152 stream networks are derived from the DEM using hydrology tools in the ArcGIS environment.
153 OpenStreetMap is an open licensed map to create a free editable map worldwide that allows users
154 to download all the features like road, river, streams to the desired vector file. Hence road layer is
155 downloaded from the OpenStreetMap. The built-up area is extracted from the LULC 2019 using
156 the Raster Tool in ArcGIS.

157

<Table 1 Data Source>

158 **2.3 Land Use and Land Cover classification**

159 GEE helps analyze the data at a planetary scale and helps the researchers and developers detect
160 changes in the map trends and quantity differences in the Earth's surface. In recent years, machine
161 learning techniques for high precision classification have developed in the area of RS. LULC is
162 categorized using two machine learning techniques in the GEE environment: Random Forest (RF)
163 classifier and Support Vector Machines (SVM). The RF classifier has attracted the interest of the
164 RS community (Belgiu and Drăgu 2016; Gislason et al. 2006; Hamad et al. 2018; Mahdianpari et
165 al. 2018). Whereas the SVM, a novel algorithm based on statistical learning theory, has not
166 exploited much within the RS community (Adam et al. 2014; Adelabu et al. 2014; Shi and Yang
167 2015). As a result, we aimed to test, investigate, and compare the performance of these machine
168 learning algorithms in the categorization of LULCs.

169 *2.3.1 The Random Forest classifier*

170 An RF classifier is a multi-decision tree ensemble classifier that creates many decision trees using
171 a random selection of training samples and variables. In recent years, these ensemble learning
172 approaches have been frequently applied in the RS. It is introduced by Breiman (2001) that
173 combines K binary Classification And Regression Trees (CART) as it is a non-parametric
174 classifier, so no statistical assumptions have to be made before the distribution of the data. Input
175 sample subgroup used to build each tree obtained from the bootstrapping. The input variable splits
176 into subgroups based on the testing from the individual algorithm in each tree. RF trees are
177 different from the Decision Trees by selecting each node in a subgroup of the input variables in a
178 random manner and built without pruning (Pelletier et al. 2016). Hence this is a recursive process
179 till the samples are similar or the splitting no longer enriches the model in each subgroup.

180 2.3.2 *Support Vector Machines*

181 SVM, introduced by Cortes and Vapnik (1995), is based on statistical learning theory. By
182 minimizing the empirical risk and confidence interval of learning derived from the systemic risk
183 minimization hypothesis, it aimed to achieve strong generalization capability. It is an efficient and
184 robust algorithm for both classification and regression. The SVM concept uses the support vectors
185 at the class domain's edges to create hyperplanes between classes in feature space. The model seeks
186 optimal hyperplane to separate the classes at maximum margins (Shi and Yang 2015). SVM was
187 designed to handle linearly separable classes by bilinear classification, but hyperplane may not be
188 located between the two cases in most cases. In such cases, the model converts the highly
189 inseparable data into a higher dimension or even infinite-dimensional feature space to separate it
190 linearly (Raghavendra and Deka 2014). In theory, the error penalty, which allows for
191 misclassification, substantially impacts SVM classification accuracy (Ustuner et al. 2015).

192 The classification, once done using the above machine learning techniques, error matrix has been
193 generated for each year to identify the accuracy of the classifications. The degree to which the
194 results are close to values accepted as true is referred to as accuracy. Error matrix allows
195 calculating overall accuracy, consumer accuracy, producer accuracy, and kappa statistics. The
196 overall accuracy (OA) can be derived using the equation (1)

$$197 \quad \text{OA} = \left(\frac{P_c}{P_n} \right) \times 100 \quad (1)$$

198 Where P_c is the number of correctly classified pixels, and P_n is the total number of pixels.

199 Consumer accuracy of each class depends on the number of correctly classified pixels in the class
200 to the number of pixels belonging to this class in the classification. Simultaneously, the producer

201 accuracy depends on the number of correctly classified pixels to that of the number of pixels
202 belonging to each class in the reference data. The notation similar to Cohen (1960), the kappa
203 coefficient κ is estimated. The proportional reduction of errors by the classification classes to the
204 error of the completely random class (Forghani et al. 2007; Tassi and Vizzari 2020). The
205 magnitude of κ usually lies between -1 to +1. The values more than +0.5 indicate it is in good
206 agreement with the classification (Taati et al. 2015). The best-performing model is analyzed. Then
207 the classified model is further future predicted to analyze the Spatio-temporal change in the model.

208 **2.4 CA-Markov model**

209 CA (Cellular Automata)-Markov is one of the most extensively utilized models predicting LULC's
210 future spatial and temporal changes. To forecast the change in LULC, this hybrid model combines
211 the CA with the Markov Chain model (Aburas et al. 2018). As the Markov Chain model is
212 stochastic, it predicts one cell transformation to another, i.e., the transition probability of the cell.
213 However, the disadvantage of the Markov Chain model is that it does not consider the effect of the
214 neighboring cells over the other. Thus, it lacks spatial modeling capability (Ozturk 2015). CA
215 model considers only the neighboring cell of interest for the estimation of the future. Thus, both
216 the model combined to make the CA-Markov model analyze the Spatio-temporal changes in the
217 land cover. It takes two time periods of LULC (i.e., earlier and later dates) as the input to analyze
218 the change in trend. The LULC class distribution was predicted using the transition probabilities
219 between class i and j using P_{ij} . It was determined over a period from time t to time $t+1$.

$$220 \quad P_{ij} = \frac{n_{ij}}{n_i} \quad (2)$$

$$221 \quad \sum_{j=1}^k P_{ij} = 1 \quad (3)$$

222 Where n_i is the total number of pixels transformed from class i over the transition period and n_{ij} is
 223 the number of pixels transformed from class i to j . k is the total number of classes. Using the
 224 transition probability P and the LULC distribution at the beginning of time t (T_t), the change in
 225 LULC over time $t+1$ (T_{t+1}) using the equation (4).

226
$$T_{t+1} = P \times T_t \quad (4)$$

227 CA assesses the contiguity configuration as well as transition probabilities (Hamad et al. 2018).
 228 The CA-Markov model, which consists of the basic LULC layer, transition potential areas for
 229 future change formed by the Markov Chain model, and transitional potential layers for LULC such
 230 as road network, built-up regions, and stream network, is used to determine the appropriate
 231 transition from one class to the other (Halmy et al. 2015). In this study, a typical 5x5 pixel
 232 contiguity filter had a considerable impact on the cellular center. As a result, each cell is
 233 surrounded by a 5x5 matrix space that defines the neighborhood of each land class. The standard
 234 contiguity filter is given as

235
$$\text{contiguity filter } 5 \times 5 = \begin{bmatrix} 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 1 & 1 & 0 \\ 1 & 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 \end{bmatrix}$$

236 The LULC layer of 2009 and 2015 is given as the input to predict the LULC for 2019. This is
 237 performed to prove the model's accuracy under the given conditions. Then the model is further
 238 used to predict the future Spatio-temporal changes of LULC change.

239 **3. RESULTS AND DISCUSSION**

240 **3.1 LULC using GEE**

241 The LULC for 2009, 2013, 2015, 2017, and 2019 were classified using RF and SVM in GEE, as
242 shown in Fig. 2 and Fig. 3. For providing a better region of interest (ROI), 15-35 pixels in each
243 sample were selected. Both models were given the same collection of training and validation points
244 as input to examine their results closely. It is classified into 5 major classes like waterbody, built-
245 up, vegetation, barren land, and forest. Every class was trained with 75-95 ROI for classification
246 and 60-75 ROI for validation. It also ensured that the information was normally distributed and
247 spectrally pure. According to Lillesand and Kiefer (1979) minimum of 50 samples per class is
248 required as a rule of thumb. Random stratified sampling is used to determine accuracy, with the
249 minimal number of observations put in each segment at random.

250 <Fig. 2 Classification using RF>

251 <Fig. 3 Classification using SVM>

252 The consumers and producers accuracy for RF classification is shown in Table 2. Overall accuracy
253 and kappa statistics of RF classification are shown in Table 3. Previous studies of Cánovas-García
254 et al. (2017); Ghimire et al. (2012); Noi Phan et al. (2020) stated that ntree of 100 gives better
255 results. However, the current analysis found that the ntree of 30 worked better and had a higher
256 accuracy than the 100. During the classification, ntree values greater than 100 failed to do better.
257 For 2009, this model accurately categorized the waterbody, built-up, and barren land, but there
258 was a misclassification between the forest and vegetation type. In 2015, there was a very close
259 misclassification of built-up and barren land, with the former being listed as the latter. Using
260 Landsat 8 images, the differentiation between forest and vegetation was clear and precise. This
261 may be attributed to the Landsat 8's (level 1 products) higher radiometric resolution, i.e., twice the
262 predecessor Landsat 7.

263 <Table 2 Consumers accuracy and Producers accuracy in RF for each year of LULC>

264 <Table 3 Overall accuracy and Kappa statistics in RF for each year of LULC>

265

266 Over the image, an SVM model is trained with the same collection of ROI and validation points.

267 The importance of the SVM settings has a significant impact on classification results (Huang et al.

268 2002; Shi and Yang 2015). The kernel type, cost parameter, and gamma value are the critical

269 parameters in the model. Hence kernel type is assigned with radial basis function (RBF), cost

270 parameter (C) is assigned with the value of 100, and gamma value is given as 0.143 (Kavzoglu

271 and Colkesen 2009; Yang 2011). Thus, the consumers and producers accuracy is shown in Table

272 4. The overall accuracy and kappa statistics were shown in Table 5.

273 <Table 4 Consumers accuracy and Producers accuracy in SVM for each year of LULC>

274 <Table 5 Overall accuracy and Kappa statistics in SVM for each year of LULC>

275

276 Table 3 and Table 5 show that the RF algorithm performs more accurately than that of the SVM.

277 RF is more suited to multiclass problems so that it can handle small differences in classification.

278 In contrast, SVM misclassified between the built-up and the barren land. Their reflectance remains

279 slightly identical, and the SVM model shows some class misclassification in the region around the

280 water bodies. This is evident in Table 4, resulting in very less consumers accuracy of the barren

281 land. The RF algorithm performed poorly in images obtained during the rainy season (2015),

282 misclassifying built-up areas as barren land. During the dry season, the classification is accurate

283 despite RF's outperforming SVM in terms of efficiency. As a result, the RF algorithm-based

284 classification is being proposed for further analysis.

285 Table 6 shows the change in area from the year 2009 to 2019. The waterbody area decreased

286 steadily from 2013 to 2017, but after the flooding, the area of the waterbody increased significantly

287 in the study. It has also been observed that the vegetation has declined by 8% between 2017 and
288 2019. There is a boom in the built-up area from the year 2009 to 2019. This is also the main reason
289 for the study, as Kerala cities have the highest rate of urbanization. In a decade, the urban area has
290 grown by almost 100%. Fig. 4 shows the rapid increase in barren land in the year 2019. This may
291 be attributed to the combined effect of the anthropogenic activities and the state's floods in 2018
292 and 2019, which caused landslides. It accounts for a 117% increase in barren land from 2009 to
293 2019, resulting in decreased vegetation cover and forest cover combined. Thus, the RF classified
294 image in GEE gives very effective insight into the LULC study transition over a decadal period.

295 <Table 6 Areal change in the LULC from 2009 to 2019>

296 <Fig. 4 Graphical visualization of areal change in the LULC from 2009 to 2019>

297 The classified image based on the RF algorithm is then used to forecast and interpret the LULC
298 change in 2019, 2035, and 2050 using CA-Markov analysis in LCM.

299 **3.2 Future prediction using CA-Markov**

300 CA-Markov analysis has been carried out using the Idrisi TerrSet software in the LCM module.
301 The LULC map 2009 and 2015 is given as the earlier land cover image and later land cover image.
302 The transition probabilities from 2009 to 2015 and the 2015 LULC base map were used to model
303 the 2019 LULC using the CA-Markov approach. The potential for transition probability is
304 calculated based on the area of each LULC class (preferably barren land, forest, and vegetation)
305 that can be converted into the built-up (Fig. 5). The transition of these areas was divided into the
306 number of timesteps (as an iteration) given by the simulation, which provided the areas to be
307 transitioned each iteration. Land suitability maps show the most suitable pixel to each LULC class.
308 Down-weighted pixels are those that are further apart from the pixel to be converted. These
309 neighborhood rules were defined using the contiguity filter of 5x5 pixels. Assignment of the pixels

310 in the future depends on the pixel's suitability to a specific LULC class. This simulation continues
311 till each pixel of the LULC class is iterated.

312 <Fig. 5 Potential for the transition from barren land (1), forest (2), and Vegetation (3) to built-up>
313 LULC class changes are influenced by various driver variables such as DEM, slope, road network,
314 distance from the road, distance from built-up areas, and distance from stream network (Fig. 6).
315 The DEM and slope reveal that the built-up class has a higher probability in areas with a lower
316 elevation and a lesser degree of slope. Distance from the road, distance from the built-up, and
317 distance from the stream are the driver variables that signify the likelihood of a rise in the built-up
318 classes that are very close to the road, current built-up, and stream, respectively those that are
319 farther apart. In this analysis, the distance from the road, DEM, slope, and distance from built-up
320 areas was generated and used to predict the increase in built-up areas. The resulting map is poor
321 and has much less accuracy. The inclusion of distance from the stream to the model increased the
322 forecast map's accuracy for 2019. Thus, it reveals that urban area is heavily reliant on the river
323 valley, as many residents depend on agriculture.

324 The association between the driver variables and the distribution of LULC classes in later year
325 image (2015) has been derived by finding the Cramer's V. Although it is an imprecise fishing tool,
326 higher Cramer's V shows that the variable's possible explanatory value is good. However, it does
327 not guarantee good results because it does not account for the mathematical constraints of the
328 modeling technique utilized and the relationship's complexity. The Cramer's V for the driver
329 variables is given in Table 7. $V < 0.3$ shows the weak association, $V \in [0.3-0.5]$ shows the medium
330 association, and $V > 0.5$ shows the strong association. Distance from built-up and slope shows the
331 medium association whereas DEM, distance from road, and stream shows potentially high
332 Cramer's V.

333 <Fig. 6 Driver variables for CA-Markov model>

334 <Table 7 Cramer's V for driver variables>

335 Table 8 compares the predicted map of 2019 to the classified map of 2019, and the accuracy for
336 each class in LULC is evaluated. The LULC classes have a high correlation, with more than 95%
337 confidence precision in the waterbody, built-up area, and forest cover. The vegetational cover is
338 94% compared to the barren land, which is the least accurate than the actual 2019 LULC map,
339 48.99%. This is because the study area was one of the most severely damaged and devastated areas
340 during the 2018 and 2019 Kerala floods. These heavy floods caused landslides and erosion along
341 the river banks in the Western Ghats section. This may be one of the key explanations for more
342 barren land than predicted by the model.

343 <Table 8 Accuracy assessment of the predicted 2019 LULC with actual 2019 LULC>

344 The areal change in the predicted LULC of 2019, 2035, and 2050 are shown in Table 9. Following
345 2019, there is a gradual growth in built-up and barren land in 2035 and 2050. The estimated barren
346 land area for 2019 is 391.7 Km², while the total area is 799.5 Km². For the same forecast rate, the
347 annual rise in barren land accounts for just 422.67 and 458.84 Km², respectively. There is a slight
348 reduction in built-up area from 549.16 Km² in 2019 (predicted) to 564.14 Km² in 2019 (actual).
349 The estimated growth in a built-up area for 2035 and 2050 is 657.64 and 773.01 Km², respectively.
350 In the long run, there is a considerable decrease in vegetation cover. Thus, the graphical
351 representation with the LULC 2019 (actual), predicted 2019, 2035, and 2050 is shown in Fig. 7.

352 <Table 9 Areal change in the predicted LULC 2019, 2035 and 2050>

353

354 Fig. 8 shows the Spatio-temporal prediction of the LULC maps for 2019, 2035, and 2050. The
355 map depicts the rise in the built-up area along the stream until 2035; in 2050, only the density of
356 built-up rises rather than moving spatially. Thus, the Spatio-temporal trend indicates that by 2050,

357 almost 773 Km² of land would be transformed to a built-up area, with a density significantly higher
358 than the current LULC.

359 <Fig. 7 Graphical visualization of areal change in the actual LULC 2019 with that of the predicted LULC for 2019,
360 2035 and 2050>

361 <Fig. 8 Predicted LULC for the year 2019, 2035 and 2050>

362 Only the trend of anthropogenic presence on the land affects the expected LULC transition. It can
363 vary mainly due to natural disasters such as the 2018 and 2019 floods, which increased the amount
364 of barren land. The current pattern of LULC transition is primarily due to development practices,
365 and it may result in land deterioration, resulting in pollution, a decline in groundwater quality,
366 tension in coastal areas, and so on. These predicted maps can be seen as a prospectus for
367 stakeholders to better understand the impact of land use patterns on land cover. It can help them
368 demonstrate an interest in environmental planning and decision-making for potential land use
369 management and sustainable land cover utilization.

370 **4. CONCLUSION**

371 As urbanization in developed countries becomes uncontrollable and unsustainable, determining
372 the spatial pattern and patterns of urban development is a crucial problem to solve to achieve a
373 sustainable environment. As a result, there is a need to raise public consciousness about the
374 dramatic improvements and degradation of natural environments. In this regard, the research was
375 carried out to examine the transition in the urban environment. Since the study area was seriously
376 impacted by the Kerala floods in 2018 and 2019, this study will help society understand the
377 changes in the LULC due to anthropogenic and flood-related factors. This will assist planners in
378 doing the requisite precautionary plans for the current and potential urban growth trend for
379 sustainable development.

380 GEE has a powerful scripting language that works in tandem with its cloud infrastructure and user-
381 friendly API. It aided in the efficient study of the classification based on machine learning. This
382 open-source tool offers an outstanding opportunity for users to explore time series data over a
383 longer span of time to observe shift trends such as LULC, shoreline-related problems, and urban
384 heat, among others. The availability of free accessible Satellite data like Landsat, ASTER DEM,
385 and Sentinel series of satellites allows the RS community to explore the vast possibilities that
386 remain unexplored. We found out that the average accuracy of the SVM model is 84.5%, while
387 RF accuracy is 89%. This demonstrates the RF algorithm's ability to tackle multiclass classification
388 and outperform the SVM. In diverse spatial and spectral land covers, RF performs well and
389 dramatically increases mapping precision. The transition study in the LULC groups reveals a
390 dramatic growth in the built-up area and barren land, followed by a steady decline in vegetation
391 and forest cover.

392 The machine learning application in the CA-Markov model of Idrisi TerrSet software helps to
393 analyze the Spatio-temporal change prediction of the LULC using the LCM module. Except for
394 barren land, this model had more than 94% accuracy for all LULC classes. Supporting layers such
395 as distance from the stream, road network, distance from the road, distance from the built-up area,
396 DEM, and slope were critical in forecasting future land cover. The lower prediction accuracy may
397 be influenced by an increase in both anthropogenic and natural disasters. This might be a potential
398 limitation of the analysis since the Spatio-temporal variation is limited to the current trend of
399 transition in the LULC, which can differ over time. This research can also be investigated using
400 high-resolution images such as Sentinel to observe the evolving trend. As of now, the Sentinel
401 collection of satellite data is not available for more than a half-decade; this may be the potential
402 reach of the analysis.

403 **Acknowledgements:** The authors would like to acknowledge the Department of Water Resources
404 and Ocean Engineering, National Institute of Technology Karnataka Surathkal, India, for
405 providing infrastructural support.

406 **Author contribution** The Participation of P.K.S.S. includes data collection, analyzing the results
407 and writing the article and the participation of PCD includes supervision, conceptualization,
408 reviewing and editing the article.

409 **Funding** There was no funding provided for the work.

410 **Availability of data and materials** The data used in this research are available by the
411 corresponding author upon reasonable request.

412 **Declarations**

413 **Ethics approval** The authors confirm that this article is original research and has not been
414 published or presented previously in any journal or conference in any language (in whole or in
415 part).

416 **Consent to participate** Not applicable.

417 **Consent for publication** Not applicable.

418 **Competing interests** The authors declare no competing interests.

419 **REFERENCE**

- 420 Abijith, D., Saravanan, S., Singh, L., Jennifer, J. J., Saranya, T., and Parthasarathy, K. S. S. (2020).
421 "GIS-based multi-criteria analysis for identification of potential groundwater recharge zones
422 - a case study from Ponnaniyaru watershed, Tamil Nadu, India." *HydroResearch*, Elsevier
423 BV, 3, 1–14.
- 424 Aburas, M. M., Abdullah, S. H., Ramli, M. F., Ash'aari, Z. H., and Ahamad, M. S. S. (2018).
425 "Simulating and monitoring future land-use trends using CA-Markov and LCM models." *IOP*
426 *Conference Series: Earth and Environmental Science*, 169(1), 012050.
- 427 Adam, E., Mutanga, O., Odindi, J., and Abdel-Rahman, E. M. (2014). "Land-use/cover
428 classification in a heterogeneous coastal landscape using RapidEye imagery: evaluating the
429 performance of random forest and support vector machines classifiers." *International Journal*
430 *of Remote Sensing*, Taylor & Francis, 35(10), 3440–3458.
- 431 Adelabu, S., Mutanga, O., and Adam, E. (2014). "Evaluating the impact of red-edge band from
432 Rapideye image for classifying insect defoliation levels." *ISPRS Journal of Photogrammetry*
433 *and Remote Sensing*, International Society for Photogrammetry and Remote Sensing, Inc.
434 (ISPRS), 95, 34–41.
- 435 Agarwal, S., and Nagendra, H. (2019). "Classification of Indian cities using Google Earth Engine."
436 *Journal of Land Use Science*, Taylor & Francis, 14(4–6), 425–439.
- 437 Álvarez-Cabria, M., Barquín, J., and Peñas, F. J. (2016). "Modelling the spatial and seasonal
438 variability of water quality for entire river networks: Relationships with natural and
439 anthropogenic factors." *Science of the Total Environment*.

440 Ansari, A., and Golabi, M. H. (2019). "Prediction of spatial land use changes based on LCM in a
441 GIS environment for Desert Wetlands – A case study: Meighan Wetland, Iran." *International*
442 *Soil and Water Conservation Research*, Elsevier B.V., 7(1), 64–70.

443 Belgiu, M., and Drăgu, L. (2016). "Random forest in remote sensing: A review of applications and
444 future directions." *ISPRS Journal of Photogrammetry and Remote Sensing*, 114, 24–31.

445 Bose, A., and Chowdhury, I. R. (2020). "Monitoring and modeling of spatio-temporal urban
446 expansion and land-use/land-cover change using markov chain model: a case study in Siliguri
447 Metropolitan area, West Bengal, India." *Modeling Earth Systems and Environment*, Springer
448 International Publishing, 6(4), 2235–2249.

449 Breiman, L. (2001). "Random forests." *Machine Learning*, Springer, 45(1), 5–32.

450 Cánovas-García, F., Alonso-Sarría, F., Gomariz-Castillo, F., and Oñate-Valdivieso, F. (2017).
451 "Modification of the random forest algorithm to avoid statistical dependence problems when
452 classifying remote sensing imagery." *Computers and Geosciences*, Elsevier Ltd, 103, 1–11.

453 Cohen, J. (1960). "A Coefficient of Agreement for Nominal Scales." *Educational and*
454 *Psychological Measurement*, Sage PublicationsSage CA: Thousand Oaks, CA, 20(1), 37–46.

455 Cortes, C., and Vapnik, V. (1995). "Support-Vector Networks." *Machine Learning*.

456 CWC. (2018). "Kerala Floods of August 2018." (August), 46.

457 Faichia, C., Tong, Z., Zhang, J., Liu, X., Kazuva, E., Ullah, K., and Al-Shaibah, B. (2020). "Using
458 rs data-based ca–markov model for dynamic simulation of historical and future lucc in
459 Vientiane, Laos." *Sustainability (Switzerland)*, 12(20), 1–20.

460 Forghani, A., Cechet, B., and Nadimpalli, K. (2007). "Object-based classification of multi-sensor

461 optical imagery to generate terrain surface roughness information for input to wind risk
462 simulation." *International Geoscience and Remote Sensing Symposium (IGARSS)*, 3090–
463 3095.

464 Fu, X., Wang, X., and Yang, Y. J. (2018). "Deriving suitability factors for CA-Markov land use
465 simulation model based on local historical data." *Journal of Environmental Management*,
466 Academic Press, 206, 10–19.

467 Ghimire, B., Rogan, J., Galiano, V., Panday, P., and Neeti, N. (2012). "An evaluation of bagging,
468 boosting, and random forests for land-cover classification in Cape Cod, Massachusetts,
469 USA." *GIScience and Remote Sensing*, Taylor & Francis Group , 49(5), 623–643.

470 Ghosh, P., Mukhopadhyay, A., Chanda, A., Mondal, P., Akhand, A., Mukherjee, S., Nayak, S. K.,
471 Ghosh, S., Mitra, D., Ghosh, T., and Hazra, S. (2017). "Application of Cellular automata and
472 Markov-chain model in geospatial environmental modeling- A review." *Remote Sensing*
473 *Applications: Society and Environment*, Elsevier B.V., 5(October 2016), 64–77.

474 Gidey, E., Dikinya, O., Sebego, R., Segosebe, · Eagilwe, and Zenebe, · Amanuel. (2017). "Cellular
475 automata and Markov Chain (CA_Markov) model-based predictions of future land use and
476 land cover scenarios (2015-2033) in Raya, northern Ethiopia." *Modeling Earth Systems and*
477 *Environment*, 3, 1245–1262.

478 Gislason, P. O., Benediktsson, J. A., and Sveinsson, J. R. (2006). "Random forests for land cover
479 classification." *Pattern Recognition Letters*, 27(4), 294–300.

480 Gomes, V. C. F., Queiroz, G. R., and Ferreira, K. R. (2020). "An overview of platforms for big
481 earth observation data management and analysis." *Remote Sensing*, 12(8), 1–25.

482 Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., and Moore, R. (2017). "Google
483 Earth Engine: Planetary-scale geospatial analysis for everyone." *Remote Sensing of*
484 *Environment*, The Author(s), 202, 18–27.

485 Halmy, M. W. A., Gessler, P. E., Hicke, J. A., and Salem, B. B. (2015). "Land use/land cover
486 change detection and prediction in the north-western coastal desert of Egypt using Markov-
487 CA." *Applied Geography*, Elsevier Ltd, 63, 101–112.

488 Hamad, R., Balzter, H., and Kolo, K. (2018). "Predicting Land Use/Land Cover Changes Using a
489 CA-Markov Model under Two Different Scenarios." *Sustainability*, 10(10), 3421.

490 Hua, L., Tang, L., Cui, S., and Yin, K. (2014). "Simulating urban growth using the SLEUTH model
491 in a coastal peri-urban district in China." *Sustainability (Switzerland)*, 6(6), 3899–3914.

492 Huang, C., Davis, L. S., and Townshend, J. R. G. (2002). "An assessment of support vector
493 machines for land cover classification." *International Journal of Remote Sensing*, Taylor &
494 Francis Group , 23(4), 725–749.

495 Jat, M. K., Choudhary, M., and Saxena, A. (2017). "Application of geospatial techniques and
496 cellular automata for modelling urban growth of a heterogeneous urban fringe." *Egyptian*
497 *Journal of Remote Sensing and Space Science*, National Authority for Remote Sensing and
498 Space Sciences, 20(2), 223–241.

499 Kavzoglu, T., and Colkesen, I. (2009). "A kernel functions analysis for support vector machines
500 for land cover classification." *International Journal of Applied Earth Observation and*
501 *Geoinformation*, Elsevier, 11(5), 352–359.

502 Krishnaraj, A., and Deka, P. C. (2020). "Spatial and temporal variations in river water quality of

503 the Middle Ganga Basin using unsupervised machine learning techniques." *Environmental*
504 *Monitoring and Assessment*, Springer Science and Business Media Deutschland GmbH,
505 192(12).

506 Lambin, E. F. (1997). "Modelling and monitoring land-cover change processes in tropical
507 regions." *Progress in Physical Geography*.

508 Lillesand, T. M., and Kiefer, R. W. (1979). *Remote sensing and image interpretation. Remote*
509 *sensing and image interpretation*.

510 Mahdianpari, M., Salehi, B., Rezaee, M., Mohammadimanesh, F., and Zhang, Y. (2018). "Very
511 deep convolutional neural networks for complex land cover mapping using multispectral
512 remote sensing imagery." *Remote Sensing*, 10(7).

513 Mishra, P. K., Rai, A., and Rai, S. C. (2020). "Land use and land cover change detection using
514 geospatial techniques in the Sikkim Himalaya, India." *Egyptian Journal of Remote Sensing*
515 *and Space Science*, National Authority for Remote Sensing and Space Sciences., 23(2), 133–
516 143.

517 Noi Phan, T., Kuch, V., and Lehnert, L. W. (2020). "Land cover classification using google earth
518 engine and random forest classifier-the role of image composition." *Remote Sensing*, 12(15).

519 Ozturk, D. (2015). "Urban growth simulation of Atakum (Samsun, Turkey) using cellular
520 automata-Markov chain and Multi-layer Perceptron-Markov chain models." *Remote Sensing*,
521 7(5), 5918–5950.

522 Parthasarathy, K. S. S., and Deka, P. C. (2019). "Remote sensing and GIS application in
523 assessment of coastal vulnerability and shoreline changes: a review." *ISH Journal of*

524 *Hydraulic Engineering*, 1–13.

525 Parthasarathy, K. S. S., Saravanan, S., Deka, P. C., and Devanatham, A. (2020). "Assessment of
526 potentially vulnerable zones using geospatial approach along the coast of Cuddalore district,
527 East coast of India." *ISH Journal of Hydraulic Engineering*, Taylor & Francis, 1–11.

528 Pelletier, C., Valero, S., Inglada, J., Champion, N., and Dedieu, G. (2016). "Assessing the
529 robustness of Random Forests to map land cover with high resolution satellite image time
530 series over large areas." *Remote Sensing of Environment*, Elsevier Inc., 187, 156–168.

531 Raghavendra, S., and Deka, P. C. (2014). "Support vector machine applications in the field of
532 hydrology: A review." *Applied Soft Computing Journal*, Elsevier BV.

533 Rienow, A., and Goetzke, R. (2015). "Supporting SLEUTH - Enhancing a cellular automaton with
534 support vector machines for urban growth modeling." *Computers, Environment and Urban
535 Systems*, Elsevier Ltd, 49, 66–81.

536 Saxena, A., and Jat, M. K. (2020). "Land suitability and urban growth modeling: Development of
537 SLEUTH-Suitability." *Computers, Environment and Urban Systems*, Elsevier, 81(February),
538 101475.

539 Saxena, A., Jat, M. K., and Clarke, K. C. (2021). "Development of SLEUTH-Density for the
540 simulation of built-up land density." *Computers, Environment and Urban Systems*, Elsevier
541 Ltd, 86(November 2020), 101586.

542 Senf, C., Seidl, R., and Hostert, P. (2017). "Remote sensing of forest insect disturbances: Current
543 state and future directions." *International Journal of Applied Earth Observation and
544 Geoinformation*, Elsevier B.V.

545 Serasinghe Pathiranage, I. S., Kantakumar, L. N., and Sundaramoorthy, S. (2018). "Remote
546 Sensing Data and SLEUTH Urban Growth Model: As Decision Support Tools for Urban
547 Planning." *Chinese Geographical Science*, 28(2), 274–286.

548 Shi, D., and Yang, X. (2015). "Support Vector Machines for Land Cover Mapping from Remote
549 Sensor Imagery." (December), 265–279.

550 Sidhu, N., Pebesma, E., and Câmara, G. (2018). "Using Google Earth Engine to detect land cover
551 change: Singapore as a use case." *European Journal of Remote Sensing*, Taylor & Francis,
552 51(1), 486–500.

553 Silva, L. P. e., Xavier, A. P. C., da Silva, R. M., and Santos, C. A. G. (2020). "Modeling land cover
554 change based on an artificial neural network for a semiarid river basin in northeastern Brazil."
555 *Global Ecology and Conservation*, 21.

556 Taati, A., Sarmadian, F., Mousavi, A., Pour, C. T. H., and Shahir, A. H. E. (2015). "Land use
557 classification using support vector machine and maximum likelihood algorithms by landsat 5
558 TM images." *Walailak Journal of Science and Technology*, 12(8), 681–687.

559 Tamiminia, H., Salehi, B., Mahdianpari, M., Quackenbush, L., Adeli, S., and Brisco, B. (2020).
560 "Google Earth Engine for geo-big data applications: A meta-analysis and systematic review."
561 *ISPRS Journal of Photogrammetry and Remote Sensing*, Elsevier B.V., 164, 152–170.

562 Tassi, A., and Vizzari, M. (2020). "Object-Oriented LULC Classification in Google Earth
563 Learning Algorithms." *Remote Sensing*, 2020(12), 3776.

564 Tayyebi, A., and Pijanowski, B. C. (2014). "Modeling multiple land use changes using ANN,
565 CART and MARS: Comparing tradeoffs in goodness of fit and explanatory power of data

566 mining tools." *International Journal of Applied Earth Observation and Geoinformation*,
567 Elsevier B.V., 28(1), 102–116.

568 Ustuner, M., Sanli, F. B., and Dixon, B. (2015). "Application of support vector machines for
569 landuse classification using high-resolution rapideye images: A sensitivity analysis."
570 *European Journal of Remote Sensing*, 48, 403–422.

571 Xing, H., Hou, D., Wang, S., Yu, M., and Meng, F. (2021). "O-LCMapping: a Google Earth
572 Engine-based web toolkit for supporting online land cover classification." *Earth Science*
573 *Informatics*, Earth Science Informatics, 529–541.

574 Yang, X. (2011). "Parameterizing support vector machines for land cover classification."
575 *Photogrammetric Engineering and Remote Sensing*, American Society for Photogrammetry
576 and Remote Sensing, 77(1), 27–38.

577

Figures

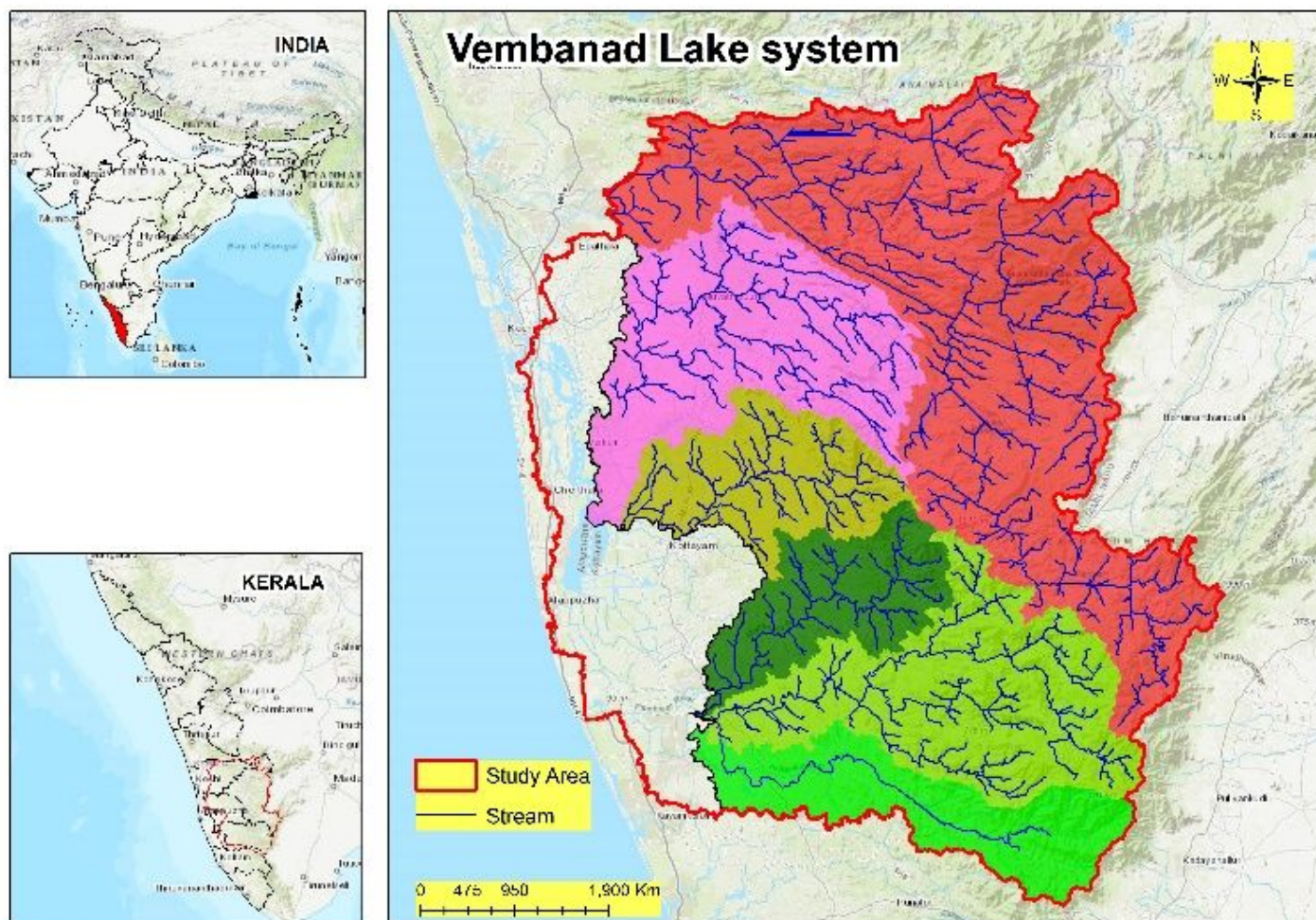


Figure 1

Keymap of the study Note: The designations employed and the presentation of the material on this map do not imply the expression of any opinion whatsoever on the part of Research Square concerning the legal status of any country, territory, city or area or of its authorities, or concerning the delimitation of its frontiers or boundaries. This map has been provided by the authors.

LULC using Random Forest

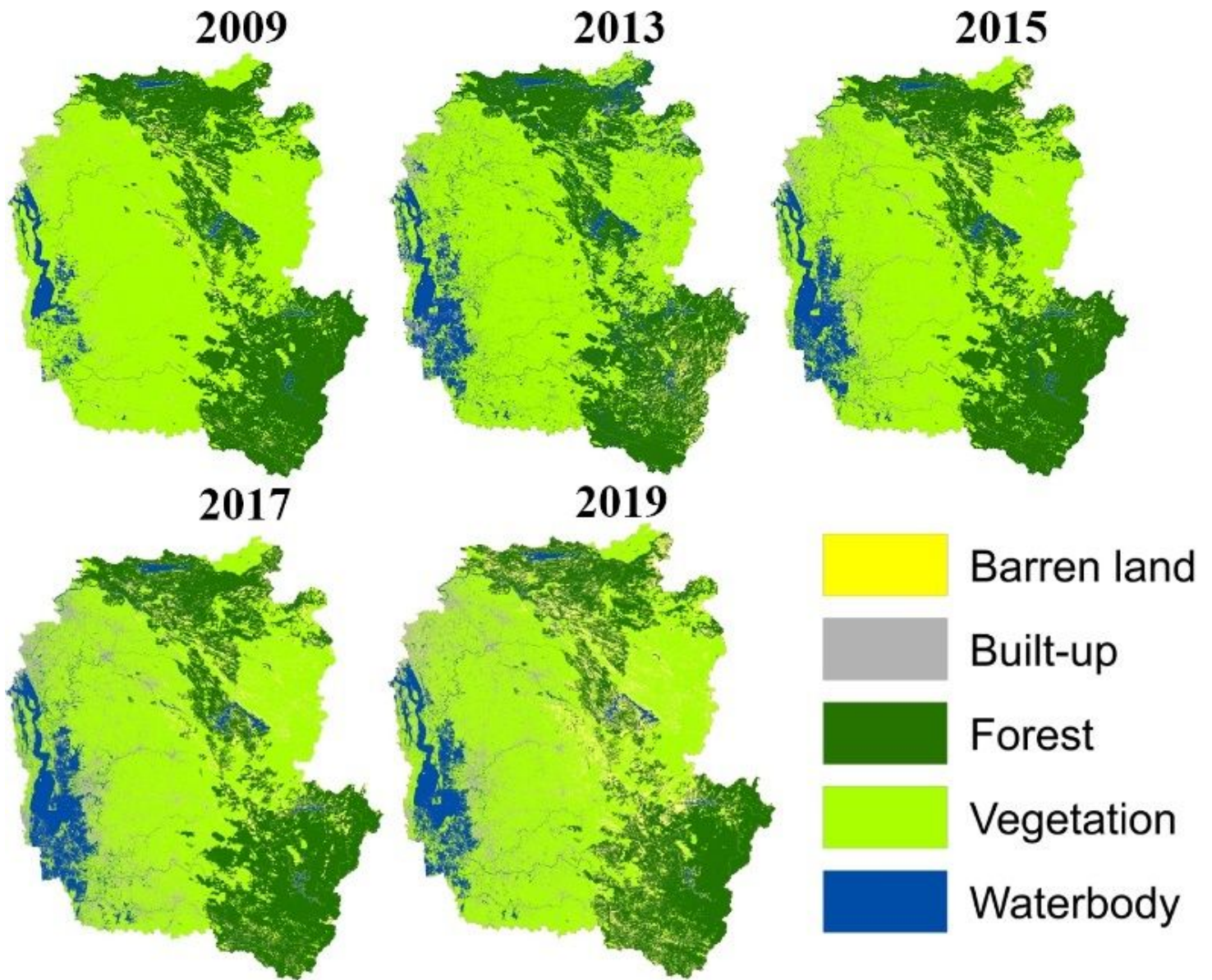


Figure 2

Classification using RF Note: The designations employed and the presentation of the material on this map do not imply the expression of any opinion whatsoever on the part of Research Square concerning the legal status of any country, territory, city or area or of its authorities, or concerning the delimitation of its frontiers or boundaries. This map has been provided by the authors.

LULC using Support Vector Machine

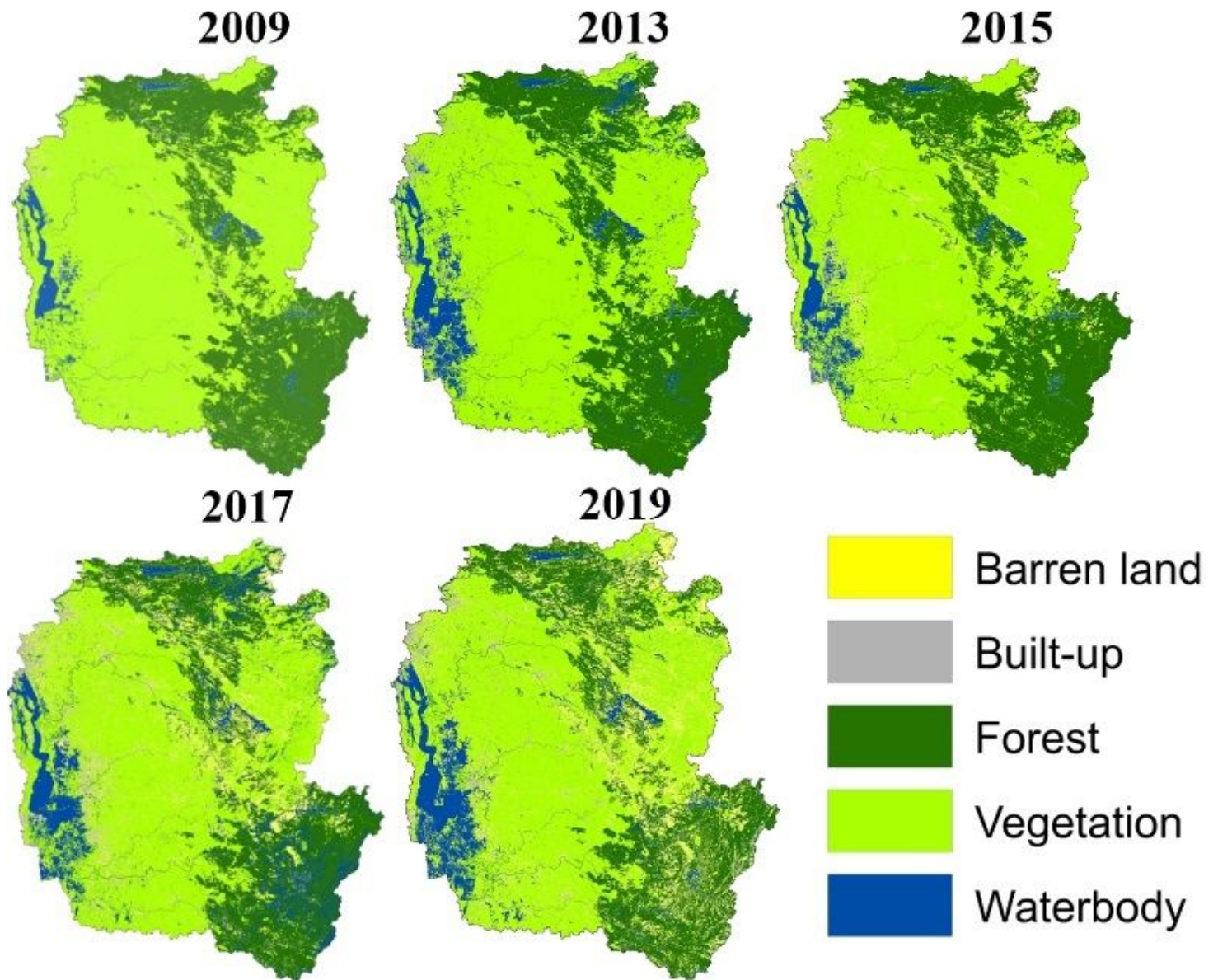


Figure 3

Classification using SVM Note: The designations employed and the presentation of the material on this map do not imply the expression of any opinion whatsoever on the part of Research Square concerning the legal status of any country, territory, city or area or of its authorities, or concerning the delimitation of its frontiers or boundaries. This map has been provided by the authors.

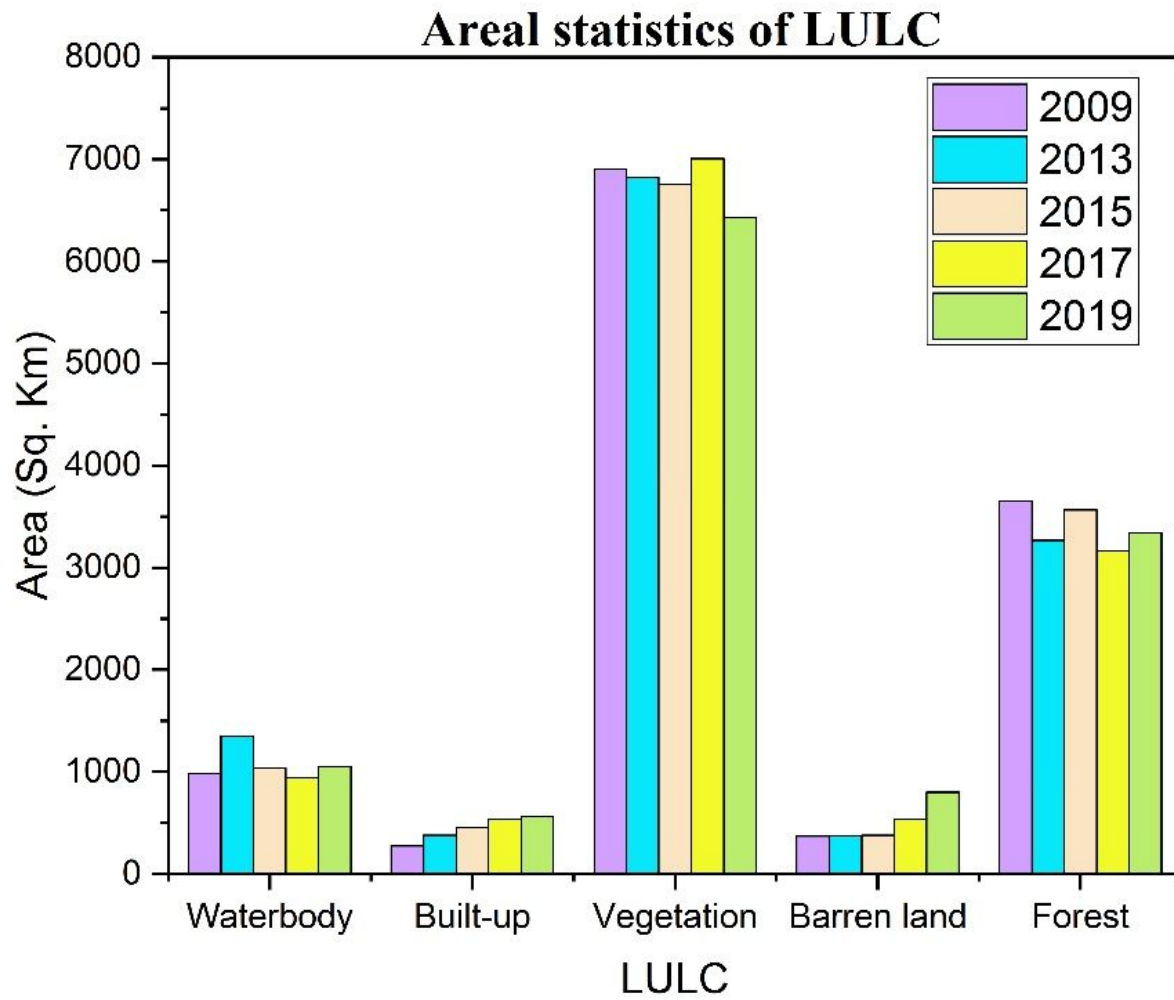


Figure 4

Graphical visualization of areal change in the LULC from 2009 to 2019

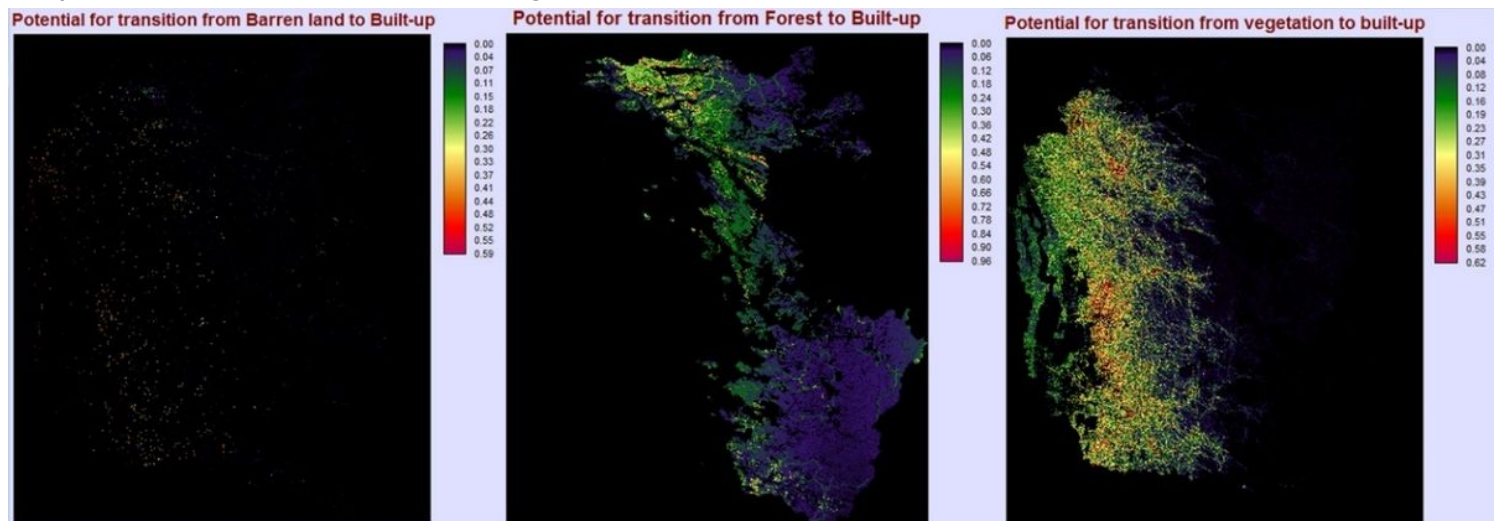


Figure 5

Potential for the transition from barren land (1), forest (2), and Vegetation (3) to built-up

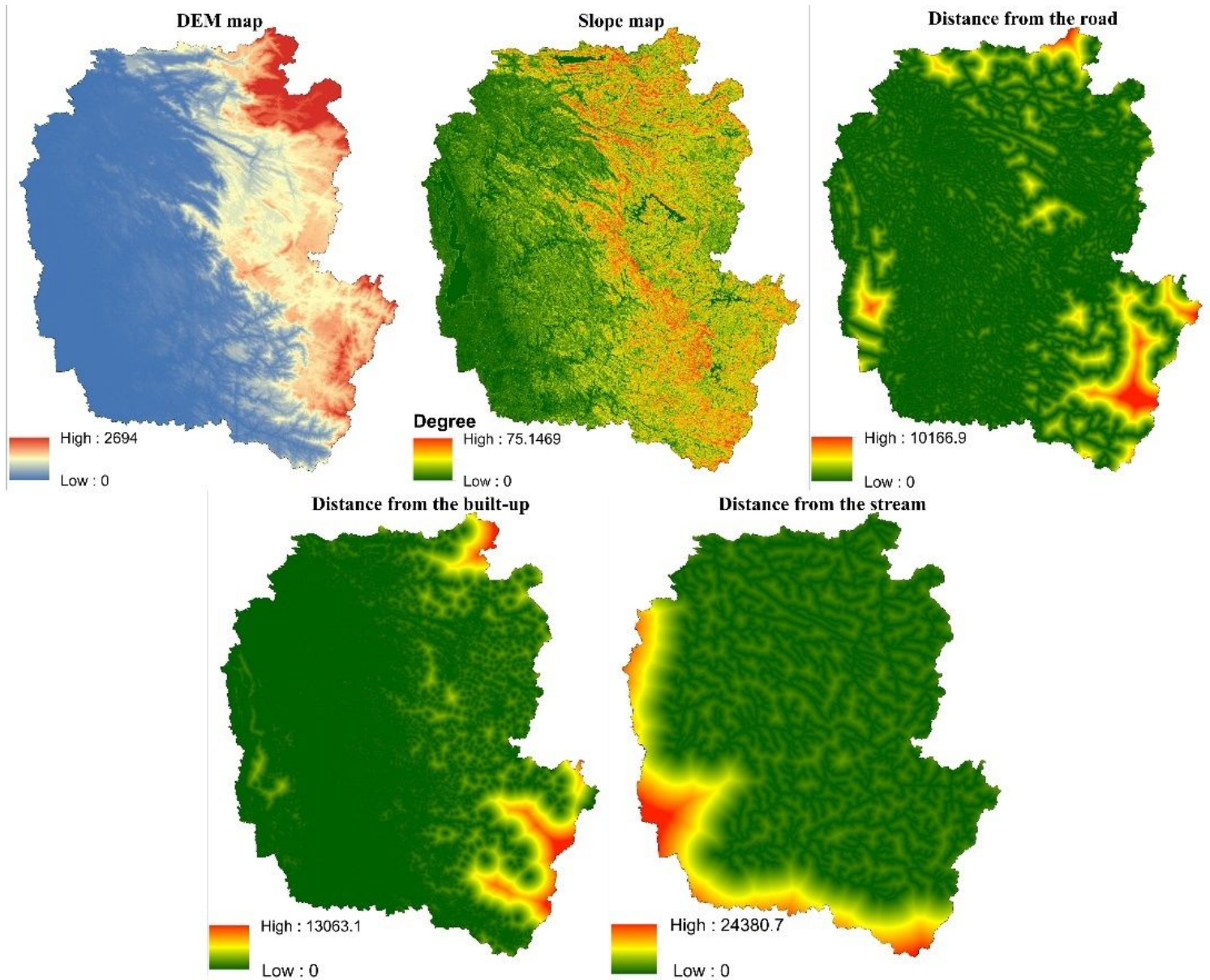


Figure 6

Driver variables for CA-Markov model Note: The designations employed and the presentation of the material on this map do not imply the expression of any opinion whatsoever on the part of Research Square concerning the legal status of any country, territory, city or area or of its authorities, or concerning the delimitation of its frontiers or boundaries. This map has been provided by the authors.

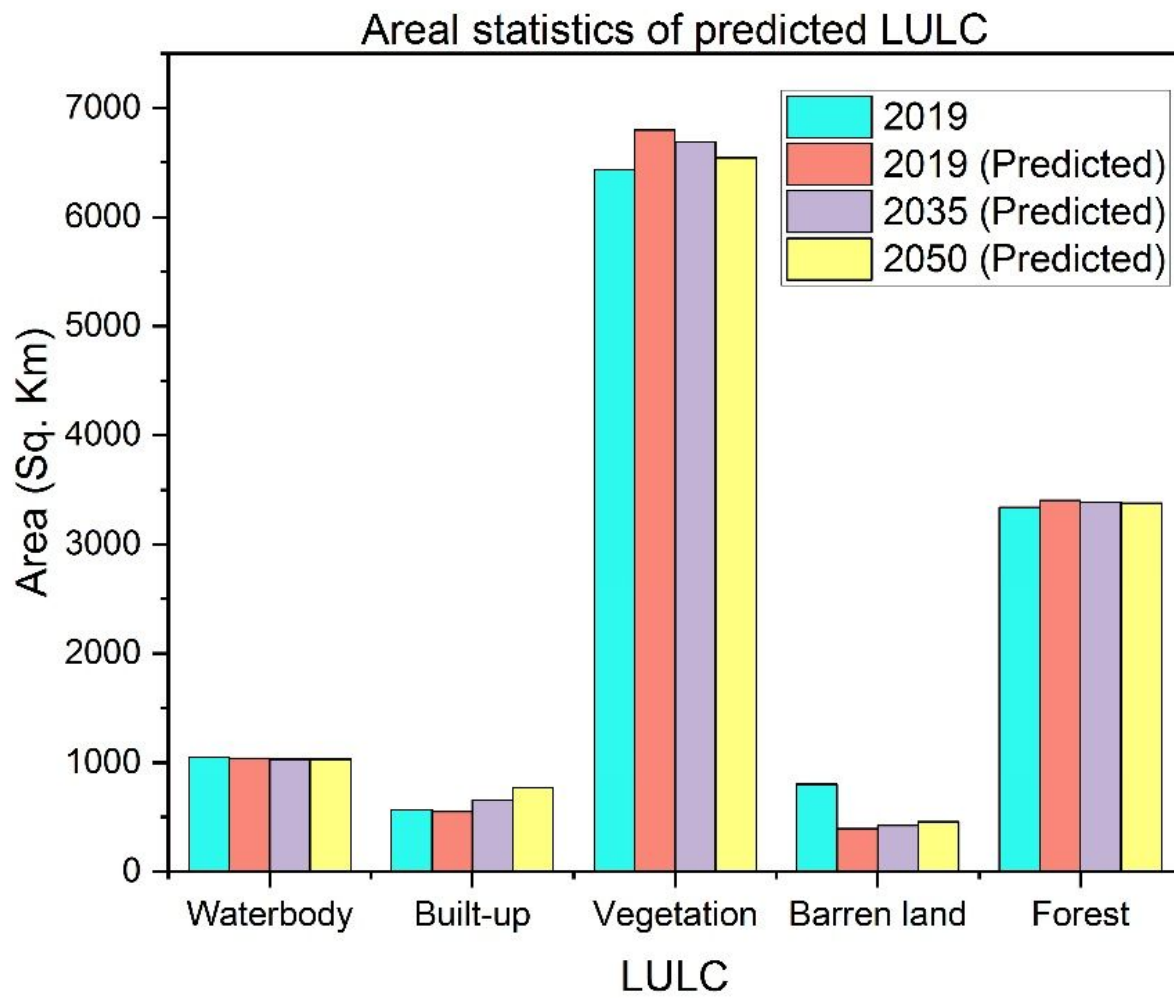


Figure 7

Graphical visualization of areal change in the actual LULC 2019 with that of the predicted LULC for 2019, 2035 and 2050

Predicted LULC for the year 2019, 2035 and 2050

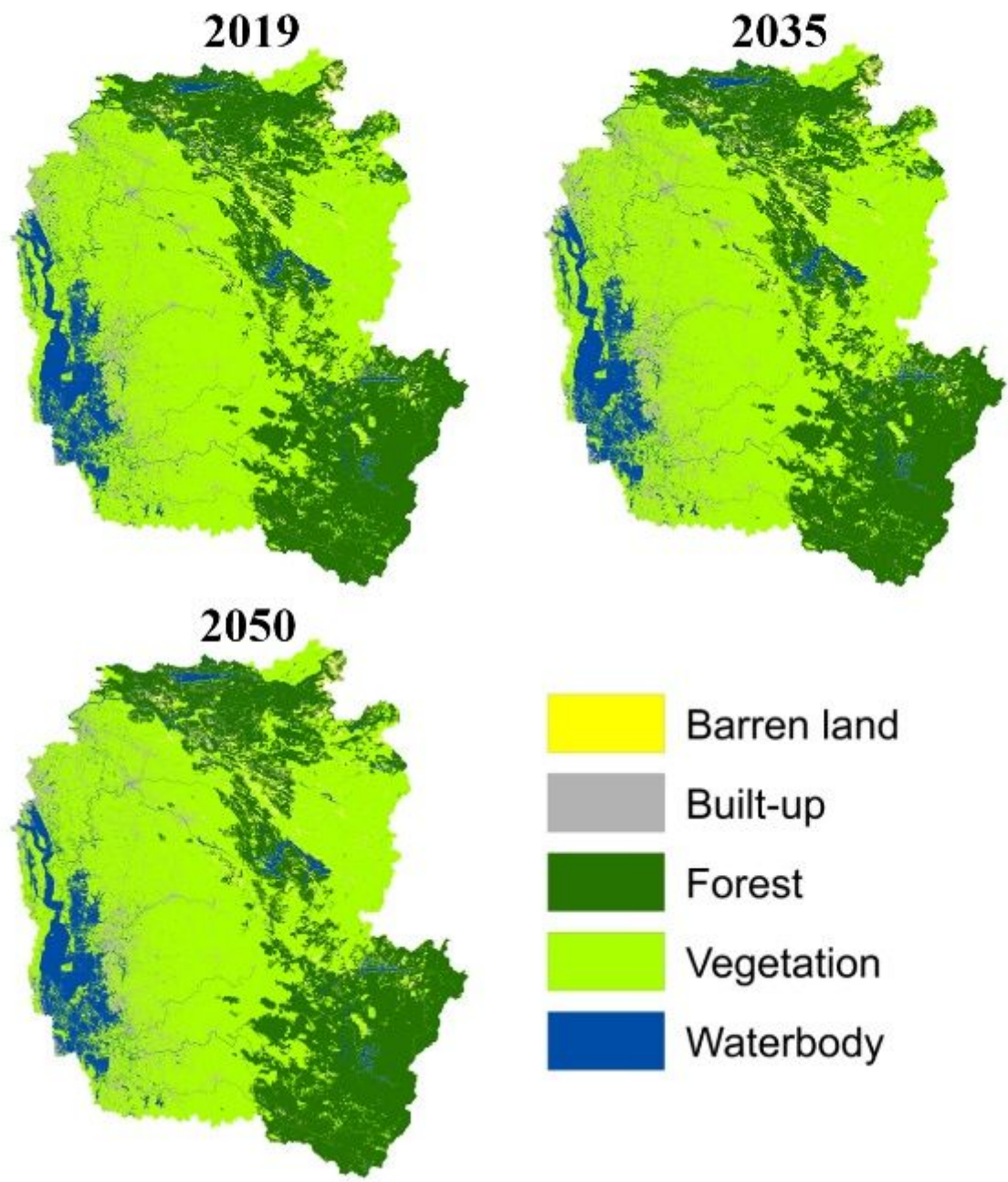


Figure 8

Predicted LULC for the year 2019, 2035 and 2050 Note: The designations employed and the presentation of the material on this map do not imply the expression of any opinion whatsoever on the part of Research Square concerning the legal status of any country, territory, city or area or of its authorities, or concerning the delimitation of its frontiers or boundaries. This map has been provided by the authors.