International Journal of Remote Sensing, Volume 41, Issue 23, 2020, pp. 9229-9255
DOI: 10.1080/01431161.2020.1815890

Spatio-temporal analysis of land use and land cover change: a systematic model inter-comparison driven by integrated modelling techniques

Srishti Gaur1, Ateeksha Mittal1, Arnab Bandyopadhyay2, Ian Holman3, Rajendra Singh4

[1] srishtigaur76@gmail.com, +91-9547652260, Agricultural and Food Engineering, Indian Institute of Technology, Kharagpur, 721302, West Bengal, India.

ORCID iD: https://orcid.org/0000-0002-5361-4474

[1] mittalateeksha@gmail.com, +91-8006820697, Agricultural and Food Engineering, Indian Institute of Technology, Kharagpur, 721302, West Bengal, India.

[2]Associate Professor, arnabbandyo@yahoo.co.in, +91-9402275849, Department of Agricultural Engineering, North Eastern Regional Institute of Science and Technology, Nirjuli (Itanagar), 791109, Arunachal Pradesh, India.

ORCID iD: http://orcid.org/0000-0001-6653-250X

[3]Professor, i.holman@cranfield.ac.uk, +44(0)-1234758277, Cranfield Water Science Institute, Vincent Building, Cranfield, Bedfordshire MK430AL, United Kingdom.

[4]Professor, rsingh@agfe.iitkgp.ac.in, +91-9434016968, Department of Agricultural and Food Engineering, Indian Institute of Technology, Kharagpur, 721302, West Bengal, India

Abstract

Currently, land use and land cover change (LULCC) is of utmost concern for global environmental change and sustainability. Among the portfolio of techniques, modelling is considered as the best approach to explore LULC dynamics. We have performed a model inter-comparison exercise (including hybrid and non-hybrid models) using multiple performance metrics to identify the best modelling approach for the subsequent projection of future LULC. The methodology is demonstrated on the Subarnarekha basin of Eastern India by utilising the LANDSAT imagery of 1989, 1994, 2006, and 2011. Before LULC modelling, the cross-tabulation and trend-surface analyses were performed to identify dominant land transitions in post-classification maps. Temporal mapping results over 1989-2011 exhibited a drastic decrease in the area under dense forest (25.7% to 19.0%), a substantial increase in the area under scrubland (21.0% to 26.1%) and a nominal reduction in the coverage of the agricultural land (51.2% to 49.0%). Four integrated models namely Multilayer perceptron-Markov Model (MLP-MC), Logistic Regression-Markov Model (LR-MC), and two hybrid models, i.e., Multilayer perceptron-Cellular automata-Markov model (MLP-CA-MC) and Logistic Regression-Cellular automata-Markov model (LR-CA-MC) were tested for their suitability for predicting future LULC for the basin. Based on the multiple model validation techniques, the MLP-MC model performed the best. MLP-MC model subsequently used a non-stationary relationship between selected explanatory variables and LULC to predict the future LULC for 2020 and 2030. The MLP-MC model projected that relative to the level of 2011, agricultural land, dense forest, and barren land may decrease by 8.3%, 28.2%, and 23.5%, respectively, and scrubland, built-up area, and water bodies may increase by 22.5%, 87.3% and 13.3%, respectively, by 2030. Our findings contradict the prevalent view regarding the nationwide intensification of agriculture over the Indian subcontinent but are consistent with the national decreasing trend in the dense forest. The study provides a transferable methodology for the systematic comparison of LULC models (including hybrid and non-hybrid) against multiple performance metrics. The outcomes of the study may help land-use planners, environmentalist, and policymakers in framing better policies and management recommendations

Keywords: Land use and Land cover, multilayer perceptron, logistic regression, cellular automata, Markov, and explanatory variables.

Published by Taylor & Francis. This is the Author Accepted Manuscript issued with: Creative Commons Attribution Non-Commercial License (CC:BY:NC 4.0). The final published version (version of record) is available online at DOI:10.1080/01431161.2020.1815890. Please refer to any applicable publisher terms of use.

1. Introduction

Land use and land cover change (LULCC) is an outcome of human-driven modification to the Earth's terrestrial surface, in particular the alteration of land to feed and shelter the increasing human population (DeFries and Eshleman 2004). Land 'cover' (the physical and biological cover over the Earth's surface, e.g., water bodies, vegetation, wetlands, barren land, and artificial structures) and land 'use' (the anthropogenic utilisation of land, e.g., agriculture, forestry and building construction) are utterly different but intimately linked to each other (Lambin and Geist 2006; Behera et al. 2012; Ellis 2007; Islam, Rahman, and Jashimuddin 2018). LULCC is a locally pervasive phenomenon having a globally significant ecological trend (Agarwal et al. 2010), which may alter biogeochemical cycles and surface-atmosphere energy trade-offs and influence regional climate (Turner 2006).

Global changes to forest cover, intensification of agricultural activities, expansion of urban areas, and industrial expansion are the outcomes of rapid LULC changes over the last 30-years (Agarwal et al. 2010). India occupies only 2.3% of the global land area yet harbours 17% of the global population; therefore; the pressure on Indian landmass is 4-6 times the global average (Roy and Roy 2010), leading to substantial spatial and temporal changes in LULC patterns (Tian et al. 2014; Moulds, Buytaert, and Mijic 2018). Consequently, India has the most extensive national bare ground loss (34%) and the second-largest gain in short vegetation due to the agricultural intensification and expansion of the 'Green revolution' (Song et al. 2018).

However, national-scale assessments of LULC can obscure substantial regional and local differences due to the intrinsic variability in topography, hydrology, climate, and anthropogenic influences. These consequences raise the science question: How far do national LULCC trends agree with regional LULCC? Modelling LULC could help answer this question, which is the underlying motivation of the current study. The analysis is performed at a local scale (i.e., at river basin scale) as local LULC underpins the healthy functioning of ecosystems on which the human enterprises so critically depend.

Modelling LULC through a spatially explicit technique provides a platform to perceive the knowledge regarding the underlying processes and explore alternative future scenarios (Veldkamp and Lambin 2001). Such models aim to establish mathematical relationships between past land-use change and a set of explanatory variables to project locations that might experience a change in the future. Such explanatory variables are multifaceted,

spatially and temporally variable and emanate from the linkage between human-environment systems, socioeconomic, policy and institutional factors (Overmars and Verburg 2005; Kamwi et al. 2018).

Many LULC modelling approaches are available, including machine learning-based models (multilayer perceptron: e.g., Grekousis, Manetos, and Photis 2013) and regression-based models (e.g., Hu and Lo 2007). The machine learning-based and regression-based models assume the stationarity of the relationship with the explanatory variables, so that these may be applied to subsequent intervals. Transition-based models (Liu and Seto 2008), e.g., Cellular Automata (CA) models, use a set of transition rules, neighbourhood effects and expert elicitation to analyse the spatial aspect (i.e., location) of change, and so are often integrated with Markov Chain models (MC) to quantify the temporal changes. Each of these model types has its pros and cons. Therefore, hybrid approaches (Wu 2002) are recommended for bringing machine learning-based/regression-based, transition-based, and change detection modelling approaches together to overcome the limitations of individual approaches (e.g., Poelmans and Rompaey 2010; Arsanjani et al. 2013, Munshi et al. 2014). The hybridisation of modelling theories permits the development of novel approaches that can better characterise the complexity of reality. The relative performance of different hybrid models and their benefits over non-hybridised models is, however, uncertain. The selection of the model is primarily driven by the research question (modeller's expectation) that need to be answered as no modelling approach is capable of answering all questions. Sometimes the research question itself imposes limitations on the applicability of a specific model due to its governing modelling process or spatial and temporal scale. Often, the selection of the model is constrained by data availability and accessible resources. A model that can address the research question is often subjectively considered to be the best model for the specific purpose/region. Therefore, developing objective approaches to select the best model for a study area, based on its relative performance, is crucial. The best performance of any model over a particular region, however, does not establish its universal superiority over other models**.** The current study provides the transferable methodology for the systematic comparison of LULC models (including hybrid and non-hybrid) against multiple performance metrics.

Consequently, here, we have applied and tested four different LULC models, namely, Multilayer perceptron Markov model, Logistic regression Markov model, and two hybrid models, i.e., Multilayer perceptron CA-Markov model and Logistic regression CA-Markov

model, to assess their suitability to predict the LULC dynamics of the Subarnarekha river basin in Eastern India. This river basin has been selected due to past and ongoing severe deforestation, rapid urbanisation, large-scale mining and industrialisation that is leading to severe soil degradation in the upper catchments and increasing susceptibility to flash floods (Singh and Giri 2018). Hence, it is essential to understand how recent trajectories of land use may manifest into the future, as this knowledge may prove crucial for land-use planners, resource managers, stakeholders and conservation practitioners in their decision making. The present study is a maiden attempt to inter-compare the relative performance of hybrid and non-hybrid models for the prediction of LULC using multiple performance metrics. Unlike previous studies, we have considered the non-stationary relationship between the explanatory variables and LULCC. The study also presents the generalised methodology for spatial and temporal analysis of LULC growth patterns. A new hybrid technique 'Multilayer perceptron-Cellular automata-Markov model' is introduced in the present work**.**

2. Data and Methodology

2.1 Study area

The Subarnarekha River, within the 19296 km^2 Subarnarekha basin (Figure1), is the prime source of irrigation and urban water supply in three Indian states, i.e., Jharkhand, Odisha, and West Bengal. The climate in the basin is typically tropical with hot summers and mild winters (average temperature varying from 9 to 41°C), and the mean annual rainfall is 1,800 mm. Agricultural activities occupy the major area of the basin. There are two major reservoirs, i.e., Getalsud and Chandil. The Getalsud dam is operational since 1971, and the construction of the Chandil dam was initiated during 1982-83 but only completed in 2011. Construction of this dam has led to the displacement of communities in the flooded area and deforestation (due to irrigated agricultural expansion and supporting constructions) over the Chandil region.

The 2011 total population of the basin is 7426016 (Census of India), with the population density being higher in the lower part of the basin. Two major cities of Jharkhand state, i.e., Jamshedpur and Ranchi, are situated in the basin, plus the towns of Ghatshila and Muri. The urban population of Jharkhand increased rapidly during 1971-1981 due to the establishment and growth of several industrial complexes in Jamshedpur, Ranchi, and Bokaro. This population growth supported infrastructural and commercial developments and expansion of

the road, rail, and air transport networks (Figure 1), leading to further expansion of the urban population.

2.2 Data

Table 1 lists of the geospatial and attribute data used in this study, including the source and resolution. Remote sensing imagery is commonly used to obtain geo-referred maps for the evaluation and monitoring of LULC. Landsat TM and Landsat ETM imageries have been obtained from the United States Geological Survey (USGS) data portal. These years were selected to : i) maintain a gap of 5-6 years between successive maps to be able to monitor significant LULC changes, and ii) to incorporate LULC information at various stages during and after the construction of water storage structures (reservoirs and barrages) in the basin.

Topographic details (Digital elevation map; DEM) were obtained from the Shuttle Radar Topography Mission (SRTM). Data for the socioeconomic variables (population, population density, and literacy rate) from the Census of India (http://censusgis.org/india/) were obtained from Indiastat.

2.3 Systematic approach for analysis of LULCC

Figure 2 presents the step-wise workflow of the detailed LULCC analysis. The first step is the 'LULC temporal mapping,' which includes the systematic interpretation of the classified LULC maps. Subsequent steps focus on LULC modelling and include calibration, simulation, and validation stages. Calibration (Step 2) uses the past LULC maps and the explanatory variables to identify the locations that may change in the future (Pontius and Schneider 2001). The simulation phase (Step 3) predicts the LULC by using the information obtained during the calibration. Finally, the validation phase (Step 4) uses the LULC maps (of independent time interval than calibration) to justify the fidelity of the prediction method through some diagnostic measure, e.g., ROC. Here, the model calibration and validation are performed on 1994-2006 and 2006-2011 intervals, respectively. LULC-1989 is not considered in the calibration/validation process as some reservoirs and water bodies were constructed later. The 'Land Change Modeler' module of IDRISI Selva (Eastman 2012) software was utilised for the LULCC modelling.

2.3.1 LULC temporal mapping

LULC temporal mapping consists of image pre-processing (image registration and image correction) followed by image classification to obtain the LULC map of any area at time t.

For the Subarnarekha basin, Landsat images were geometrically registered to UTM coordinate system with zone 45 north and WGS 84 (World Geodetic System) datum. Dark Object Subtraction (Chavez 1996) was employed for the atmospheric corrections through the ATMOS module of IDRISI software. Land use maps of 1989, 1994, 2006, and 2011 were derived from the registered and atmospheric corrected images through the unsupervised classification in ERDAS Imagine. For classification, the reference (ground truth) data was obtained from Google Earth imagery, and 250 random control points were used to examine the classification accuracy. Table S1 presents the details of the six LULC classes, viz., water bodies, dense forest, scrubland, and built-up area, identified in Subarnarekha basin. The classification accuracy for maps of 1989, 1994, 2006, and 2011 are 90%, 88%, 87.3%, and 89.2%, respectively.

2.3.1.1 Spatial dynamics of LULC changes

The spatial identification of major LULC transitions is performed on the post-classified maps through Trend Surface Analysis (TSA) (Chorley and Haggett 196).

In TSA, the coordinates of the 'n' number of points (X_i, Y_i) for the areal distribution of attribute Z (LULC transition in our case), are taken as independent variables, with attribute Z being the dependent variable. TSA is mathematically represented as

$$Z(X,Y) = a + bX + cY + dX^2 + eXY + fY^2 + gX^3 + hX^2Y + iXY^2 + jY^3 \quad (1)$$

Where, Z is the areally distributed variable that represents the major transition between the LULC categories, (a, b, …, j) represents the polynomial coefficients, and X and Y are the locational coordinates. The difference between the real and fitted values of Z is minimised to obtain the polynomial coefficients (a, b,…, j). TSA surfaces are calculated by assigning the pixels of a particular transition value of '1' and pixels of no transition '0' and considering them as if they were continuous values. To identify the general locations of LULC transition during 1989-2011, we have employed the third-order polynomial. More details of TSA can be assessed from Chorley and Haggett (1965) and Václavík and Rogan (2009).

2.3.1.2 ***Temporal dynamics of LULC changes***

Cross-tabulation analysis (Pontius, Shusas, and McEachern 2004) deals with the temporal analysis of change. A cross-tabulation matrix quantifies temporal changes in the LULC maps over two periods in terms of gains, losses, net change, total change (sum of gain and losses), persistence, and swapping (exchanges between land cover classes). Persistence represents the amount of unchanged land cover between two periods.

2.3.2. Calibration phase

Calibration phases utilises LULC maps at time t_0 and t_1 along with explanatory variables to identify the location of change.

We employed three categories of explanatory variables, namely biophysical, proximity, and socio-economic variables (Figure 3). Biophysical variables define the environmental conditions influencing land-use change through biotic and abiotic factors (i.e., topography, soil, climate, vegetation, and lithology) that vary with location and across times (Lambin et al. 2001). Due to the significant variation in topography in Subarnarekha basin, two biophysical variables, slope and altitude, are selected. Likewise, six proximity variables, i.e., distance to major roads, distance to the rail network, distance to the reservoir, distance to major cities, distance to airports, and distance to main streams are selected as the areas situated near these places are highly susceptible to LULCC. Only the major roads are considered here and are digitized from the Survey of India toposheets of 1994, 2006, and 2011. The distances of the proximity variables (other than main streams) are considered for the later year of the historical time interval, e.g., for LULCC during 1994-2006, distance to major cities is calculated in 2006.

The socio-economic variables, viz., population, population density, and literacy rate are selected to represent the increasing demographic pressure, rapid urbanisation, and industrialisation in the basin. The influence of socio-economic variables could be advantageous or detrimental, e.g., an increase in population may lead to overexploitation of natural resources, but may spur technological innovations leading to intensification in agriculture. Likewise, the literacy rate may either provoke or alleviate the land-use changes, e.g., educated persons may migrate to cities for better career opportunities, but this may necessitate urban expansion at the cost of deforestation. A higher literacy rate may also create a favourable condition for higher income generation from land, i.e., diversification to cash crops or non-agricultural activities (Jose and Padmanabhan 2015). Enhancements in adult

literacy may raise environmental awareness, which is liable to increase recreational demand from forests (Basnyat 2009). The socio-economic data of 1991, 2001, and 2011 are collected at district levels and interpolated for 1994, 2006, and 2011. The cell size is converted to 30 m to make the explanatory variables compatible with the land cover maps.

Cramer's V, which measures the strength of association between explanatory variables and land cover maps based on Chi-square statistics, is used to test the utility of the selected explanatory variables to LULC change. Cramer's V takes values from 0 to 1, with 1 showing a perfect association. Cramer's V values of ≥ 0.4 are considered as good, values between 0.15 and 0.4 are useful, and values < 0.15 are not useful (Hamdy et al. 2016). It uses contingency table analysis of two inputs, i.e., the boolean maps of the area that has undergone the transition and the explanatory variable that is binned into 256 classes. The p-value expresses the probability of Cramer's V not being significantly different from 0.

2.3.3 Simulation phase

The locations of individual LULC transitions are identified by analyzing the past land-use changes and the current information from explanatory variables (Kolb, Mas, and Galicia 2013; Mas et al. 2014). The models used for LULC analysis generate transition maps by utilising the relationship between the LULC and set of spatial explanatory variables (known as drivers of change). Identification of the probable locations of change is referred to as transition potential. Transition potential maps (TPMs) present the values of probability for the areas that participate in a specific LULC transition (e.g., spreading built-up areas in case of urban sprawl). An accurate model identifies the areas having high and low transition potentials precisely (Kolb, Mas, and Galicia 2013). Here, we have used two regression models, MLP and LR, to generate the TPMs.

The simulation phase utilises the TPMs along with Markov chain (MC) for LULC prediction. MC is a stochastic model having processes described by the set of states, i.e., $S = \{S_0, S_1, S_2, S_3 \ldots S_n\}$, such that a process initiates in one state and changes subsequently to another; any state at time t is dependent only on the immediately previous state at a time (t-1) (Zhang, Narteau, and Rozier 2010). For LULC prediction, MC model analyses the two LULC maps of different periods to calculate the transition probability matrix (P_{ij}). The P_{ij} illustrates the

probability with which each state changes to other states (Ghosh et al. 2017). Based on the Markov process and probability formulae of Bayes theorem (Guan et al. 2011), the predicted MC is given by,

$$P_{(t+1)} = P_{ij} \times P_t \quad (1)$$

Where $P_{(t)}$ and $P_{(t+1)}$ represent LULC status at time t and t+1, respectively, P_{ij} is the transition probability of the i^{th} type land cover into the j^{th} type land cover, and m is the number of LULC classes in the study area.

The details of the integrated modelling techniques used for the LULC prediction are as follows:

2.3.3.1 Multilayer perceptron - Markov model (MLP-MC)

MLP-MC is an integrated model, in which MLP is utilised to generate the TPMs for all transitions, and the MC is used to carry out the temporal analysis.

MLP is a feed-forward neural network having three layers, i.e., input, output, and hidden layers (Mas and Flores 2008). MLPs adjust the weights between the input and output layers based on the mathematical relationships by training the neural network (Tu 1996). MLP utilises backpropagation algorithm for training, and the trained model is used for testing purposes. The performance of MLP depends on the training parameters (i.e., learning rate, momentum, and the number of iterations performed) and the number of hidden layers. In land-use predictions, MLP employs the explanatory variables as input and land-use change as output. IDRISI software uses the accuracy rate to estimate the efficiency of TPMs. The maps having an accuracy rate higher than 80% are used for LULC predictions (Aguejdad, Houet, and Hubert-Moy 2017).

2.3.3.2 Logistic Regression- Markov model (LR-MC)

LR-MC is an integrated model, in which the LR is used to generate the TPM for each transition, and the MC is used to carry out the temporal analysis. LR defines the statistical relationship between the set of independent variables (e.g., explanatory variables) and categorical binomial dependent variables (e.g., LULC). The dependent variables must be dichotomous, i.e., it takes only two values, either 0 (absence) or 1 (presence). LR is based on the assumption that the probability of the dependent variable is 1, and follows the logistic (or sigmoidal) curve (Mahiny and Turner 2003). Mathematical expression for logistic regression is:

$$\ln\left(\frac{P}{(1-P)}\right) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots\ldots\ldots\ldots + \beta_n x_n + \varepsilon \quad (3)$$

Where P presents the probability of dependent variable, β_0 is an intercept variable, { β_1, β_2 ,..., β_n } are the coefficients associated with the independent variables { x_1 , x_2 ,…, x_n }, and ε is the residual.

While incorporating the LR models, the heterogeneity and autocorrelation of the spatial data must be removed to avoid unreliable or inefficient parameter estimation and false conclusions. Therefore, the autocorrelation of spatial data is eliminated by stratified random sampling (Xie et al. 2000; Arsanjani et al. 2013). IDRISI software uses ROC (explained in section 2.3.4) to estimate the accuracy of transition potential maps. The maps having higher than 0.65 accuracy are used for LULC predictions.

2.3.3.3 Hybrid Models

Modelling outcomes from MC models do not have any spatial components, i.e., the outcomes are valid for one land cover, but do not provide any information regarding the other land cover classes in the basin (Mahiny and Turner 2003). Cellular automata (CA) can add the spatial interaction effects to the outcomes but lags in the incorporation of socio-economic variables to change dynamics (Hu and Lo 2007; Arsanjani et al. 2013). The details of the CA can be referred from Ghosh et al. (2017).

CA are characterised by a group of grid cells, a set of states specifying the grid cells. It consists of transition rules that decide the state transition of each grid cell and further update the configurations and compositions of each grid cell at different time steps. Mathematically, it is expressed as

$$\{S_{t+1}\} = f(\{S_t\}\{I_t^h\}) \quad (4)$$

$\{S_{t+1}\}$ is the state of the cell at a time (t+1), $\{S_t\}$ is the state of the cell at time t, f() is a transition rule, $\{I_t^h\}$ refers to the neighbourhood and h denotes the size of the neighbourhood.

Regression-based models, like LR, incorporate the impact of drivers (viz., socio-economic) to change dynamics by setting up a mathematical relation between the two but often suffer from the inability to quantify the spatial and temporal changes (Wu 2002; Polmans and Rompaey

2010). Therefore, 'Hybrid models' overcome the limitations of individual models while utilising their advantages.

The hybrid modelling framework integrates explanatory variables, neighbourhood interaction rules, and MC processes in a single platform. As per Polmans and Rompaey (2009), the transition probability of the hybrid model is given by

$$P_{hybrid} = P_{LR} \times N_{neigh}^{\sigma} \quad (5)$$

Where P_{hybrid} is the transition probability for the hybrid model based on explanatory variables and neighbourhood effect, P_{LR} is the transition probability map generated from logistic regression, and N_{neigh} is the neighbourhood effect from the CA.

Here, we have used two hybrid models, the first one combining the LR, CA, and MC, i.e., LR-

CA-MC, and the second one by combining the MLP, CA, and MC, i.e., MLP-CA-MC.

2.3.4 Validation phase

For the observed LULC maps at t_0, t_1, and t_2, the maps at time t_0 and t_1 are utilised to parameterise the model for calibration. The calibrated model then predicts the LULC for time t_2 to validate the simulated map at time t_2 and the map of actual change between [t_1 t_2].

The validity of any LULCC model can be established if the model can simulate the amount of land cover change and capture the persistence in each category, besides capturing the spatial distribution (locations) of land cover change between the actual maps.

Several techniques are available for model validation, with the Kappa index being the most commonly used technique. Pontius and Millones (2011), however, identified the unsuitability of the Kappa index for accuracy assessment in remote sensing applications. Therefore, a quantitative measure, ROC, is used to validate a LULC model that can generate a series of transition potential maps. ROC analysis is applied on the binary simulated map at time t_2 and the map of actual change (showing change and persistence) between time [t_1-t_2].

For the estimation of ROC, the map of modelled suitability is divided into several groups by a user-defined threshold, e.g., for a 10% threshold, 10% of the cells having the highest suitability are segregated as group-1, the next 10% cells as group-2 and so on. Each division is called a scenario and is compared with the map of actual change. A cell is considered as true positive if the simulated change in the model scenario agrees with the actual map.

Conversely, a cell is considered as false-positive if a change is simulated in the modelled scenario but there is no change in the actual map. Plotting true positive on the ordinate and false-positive on abscissa constitutes the ROC curve (Pontius and Schneider 2001). The area under the ROC curve, AUC is used as the diagnostic metric. AUC lies between 0 and 1, and in case of random allocation of cells in suitability map, AUC is considered as 0.5. The same transition potential maps generated from the MLP and LR methods are utilised in hybrid models and, therefore, ROC metric is applied for the MLP-MC and LR-MC models.

For the analysis of the amount of change between the simulated and actual LULC map, the Chi-Square goodness of fit test is performed at 5% significance level using the null hypothesis that actual and simulated areas of distinct land cover categories in 2011 are the same. The calculated value of the Chi-Square goodness of fit test is compared with the tabulated value, and the null hypothesis is rejected if the calculated value is found higher than the tabulated value,

The amount of persistence between the simulated and actual maps of 2011 is obtained by the cross-tabulation analysis (Section 2.3.1). Ideally, for the two maps of the same time, the amount of persistence should be 100% for a given land cover category. Hence, the model incorporating the highest persistence (land cover wise) is considered as the most suitable model.

For the analysis of the spatial distribution of LULCC, the pixel-wise comparison is carried out. For a given land cover, if any pixel in the simulated map can capture the same land cover in the observed pixel, then it is recognised as 1 (true value) otherwise 0 (false value). The model covering the largest amount of true values is considered as the most suitable model.

In a nutshell, to select the best model for LULC prediction in Subarnarekha basin, the ROC measure is applied for the validation of non-hybrid models. Besides, the Chi-Square goodness of fit test, cross-tabulation analysis for persistence, and pixel-wise comparison are applied for all models (hybrid and non-hybrid).

2.3.5 Prediction phase

After successful validation, the best model is used to extrapolate LULC for the future period. Most of the LULCC models use the machine learning/regression techniques to extrapolate the

change; however, these techniques are only valid within the range of LULC change on which they are based, which is typically the order of one or two decades (Serneels, Said, and Lambin 2001; Lambin and Geist 2006).

For the future prediction of LULC, the TPMs are generated from the validation interval (2006-2011) using the LULC maps of 2006 and 2011 and the explanatory variables of 2011. We have considered the non-stationary relationship between the LULC and the explanatory variables (i.e., the TPM generated from [t_1-t_2], i.e., from the most recent datasets, are used for the future prediction). Most of the earlier studies, however, did not consider the spatial and temporal variability of the explanatory variables (i.e., the TPM generated from [t_0-t_1], i.e., datasets used for calibration were used for the future prediction). For the socio-economic variables, the most recent data are available for 2011 (from Census of India); therefore, TPMs (2006-2011) driven by the explanatory variables from 2011 are used for the extrapolation of future LULC maps to 2020 and 2030.

3. Results

3.1 LULC temporal mapping for Subarnarekha basin

The classified LULC maps over the interval 1989-2011 (Figure 4) illustrate that agricultural land is the dominant land cover class in the basin. A drastic decrease is seen in the area under dense forest (25.69% to 18.97%); while a substantial increase is found in the area under scrubland (21.06% to 26.14%). The coverage of the agricultural land reduced from 51.20% to 49%, while barren land has increased from 0.71% to 1.61%. The built-up area increased threefold (0.98% to 3.63%), while the area of water bodies has declined from 1.26% to 0.83%.

3.1.1 Spatial dynamics of change

TSA was performed for the time interval 1989-2011 to identify the spatial trends of change for the significant transitions from dense forest to scrubland, dense forest to water bodies, scrubland to the built-up area, scrubland to agricultural land, agricultural land to scrubland, and agricultural land to the built-up area (Figure 5). The upper portion of the basin shows the emergence of the scrubland, while the built-up area shows an increasing trend in the middle reaches of the basin, i.e., around Jamshedpur city. The transition from scrubland to agricultural land also takes place in the middle portion of the basin.

3.1.2 Temporal dynamics of change

The cross-tabulation analysis was performed to explain the persistence, losses, and gains, and swapping in land cover categories during 1989-1994, 1994-2006, and 2006-2011. Table 2 presents the results for 1994-2006 which shows that agricultural land has the highest persistence of 75.54%, with only 24.46% of the pixels under the agricultural land tending to change to other classes. Most of the agricultural land has been converted into the scrubland (1638.94 km^2). Water bodies have the second-highest persistence (74.42%) and show a considerable gain in the surface area from the dense forest (28.83 km^2). The dense forest has a moderate persistence (63.76%) as it experiences a substantial conversion to scrubland (1070.31 km^2).

The built-up (urban) area would be expected to show the highest persistence as once an area is urbanised, it is expected to remain so, but we have considered both urban and sub-urban areas under the built-up area (land cover category). Consequently, a reasonable degree of persistence (68.50%) is obtained for the built-up area. Although scrubland has an overall gain in the surface area, it shows low persistence (44.17%) and greater susceptibility to change. The barren land shows a low persistence of 27.49% and gains the majority of surface area (106.34 km^2) from the agricultural land.

The cross-tabulation analysis was performed for 1989-1994, 1994-2006 and 2006-2011, with Table 3 presenting the summary of cross-tabulation in terms of gains, losses, net change, total change, and swap for 1994-2006. It is evident from Table 3 that among all land cover classes, scrubland experienced the maximum amount of total change and swapping with other land covers in the basin due to its lowest persistence. Swapping for the water bodies is the lowest. Figure 6 (a) maps the locations of gain, loss, and persistence in each major land cover categories during 1989-2011. Figure 6 (b) presents the summary of the cross-tabulation analysis results in terms of net contribution to major classes from other land cover classes for the three-time intervals. For agricultural land and dense forest, the amount of area lost to other classes is more than the amount of area gained from other classes during 1989-2011 (Figure 6 (a)). Throughout the basin, the scrubland gained a relatively high area from other land cover classes as compared to the area lost to other categories (Figure 6 (a)). The amount of area gained from the other land cover classes is very high for the built-up area in comparison to the area lost to other land cover classes (Figure 6 (a)).

The spatial and temporal changes in each land cover category obtained from TSA and cross-tabulation analysis provided the details of a significant transition of LULCC dynamics that can explain the LULCC dynamics from 1989 to 2011 shown in Figure 4. Consistently decreasing trends in dense forest from 1989 (25.69%) to 2011 (18.97%) (Figure 4) arise, in part, from the construction of the Chandil dam and other barrages in the forested area and increased demographic pressures, extension of rail and road networks, and infrastructural development in the basin, especially in and around cities. Despite the construction of reservoirs, the coverage of water bodies has reduced in the basin from 1.26% to 0.83% during 1989-2011 (Figure 4) due to conversion into other classes, e.g., to agriculture due to higher fertility of the soils near the water bodies.

Consistently increasing trends are obtained for scrubland during 1989 (21.06%) to 2011 (25.94%) due to conversion of dense forest and agricultural land (Figure 4). Agricultural land decreases between 1989 (51.27 %) to 2011 (49%) due to its transition to scrubland, built-up areas and barren land (Figure 4). Development of infrastructure, communication networks, and technological advancement have resulted in socio-economic awareness and development, leading to migration from rural to urban areas, and resulting in the abandonment of marginal agricultural lands (Singh and Giri 2018) and an increasing trend in barren land between 1989 (0.71%) and 2011 (1.61%). Consequently, the rapidly growing population, increasing literacy rate, migration and urbanisation are factors responsible for the increase in the built-up area in the Subarnarekha basin between 1989 (0.98%) and 2011 (3.63%).

3.3 Calibration

3.3.1 Selection of explanatory variables

The association level between the explanatory variables and LULC, measured through the Cramer's V for all three-time intervals, i.e., 1989-1994, 1994-2006, and 2006-2011, is presented in Table 4. The coefficient of Cramer’s V is tested with the p-value less than 0.05. The bio-physical variables (i.e., elevation and slope) are found to be the most influential. Only two proximity variables (distance to cities and distance to the reservoir) and two socio-economic variables (population and population density) have shown a good degree of association with the LULC, which shows the prominent contribution of urban processes, development of artificial water bodies, and demographic variables to the LULCC in the basin.

3.3.2 Generation of transition potential maps

Two models, multilayer perceptron (MLP) and logistic regression (LR) are employed with the chosen explanatory variables to generate the transition potential maps. In MLP, the training parameters are set as 0.1 for learning rate, 0.04 for momentum factor, and 10000 number of iterations. All possible transition maps are modelled using the chosen explanatory variables, and the maps having accuracy greater than 80% were selected for LULC predictions. For the transition potential maps generated using logistic regression, those having ROC higher than 0.65 are used for LULC prediction. Figure 7 (a-b) presents the transition potential maps generated using the two models during the calibration period (1994-2006). The colour variation in Figure 7 shows the areas having high to low suitability of change in the basin.

3.4 Simulation

The MC model provided the transition probability matrix for 2011 by utilising the information from the LULC maps of 1994 and 2006. This transition probability matrix is combined with the probability surfaces generated using the MLP and LR models for the prediction of LULC for 2011 through the integrated approaches of MLP-MC and LR-MC. Further, for the implementation of hybrid approaches, LR-CA-MC and MLP-CA-MC, CA function is applied to the probability surfaces obtained from MLP and LR methods. Figure 8 (a-d) presents the predicted LULC map for 2011 generated using the four models, i.e., MLP-MC, LR-MC, LR-CA-MC, and MLP-CA-MC. The validity of these models is tested through distinct validation techniques described in the next section.

3.5 Validation of predicted land cover

ROC analysis is performed on the map of modelled suitability (2011) using the map of actual change (2006-2011) at a 1% threshold for model validation for MLP-MC and LR-MC models. Figure 9 (a) presents the ROC plots for MLP-MC and LR-MC methods. As evident from Figure 9 (a), MLP-MC results in AUC of 0.91 while LR-MC results in AUC of 0.58. Thus, MLP-MC performs better than LR-MC method for the Subarnarekha basin.

Table 5 presents the statistics of Chi-square goodness of fit test between the actual and predicted LULC for 2011 from the different models. Table 5 shows that the tabulated chi-square value is only higher than the calculated one for the MLP-MC model,; therefore, for

this model, the null hypothesis cannot be rejected. The Chi-square test signifies the goodness of fit of MLP-MC model for the prediction of LULC for the Subarnarekha basin.

The amount of persistence for each land cover category is obtained by comparing the actual and predicted images of 2011 (Figure 9 (b)). Ideally, for the same period, the observed and simulated images should show 100% persistence for each land cover category in the cross-tabulation analysis. Figure 9 (b) shows that the MLP-MC model can capture the relatively high persistence among all land cover categories shown in Table 2. Hence, in terms of replicating the amount of persistence, the MLP-MC model performs best.

Figure 9 (c) presents the statistics of pixel to pixel comparison of all models. The predicted map from MLP-MC model can capture the pattern of actual LULC map of 2011 better than other models; however, both hybrid models perform well in capturing the spatial patterns of actual LULC map of 2011. Therefore, from this analysis of different validation techniques, MLP-MC model is selected for subsequent use for the prediction of LULC for future periods.

3.6 Prediction of future LULC

The future LULC for 2020 and 2030 is predicted using MLP-MC method (Figure 10 a-b). The predicted LULC follows the trend of the historical LULC, i.e., the built-up area, scrubland and barren land are found to increase at the expense of dense forest, agricultural land, and water bodies. Urbanization is predicted to increase around towns like Muri and Ghatshila also.

The heat map (Figure 11), a 2D/3D representation of data in which colours display values, presents the percentage change for each class between the future and historical LULC. Relative to the level of 2011, agricultural land, dense forest and barren land may decrease by 8.3%, 28.2%, and 23.5% respectively by 2030. The scrubland, the built-up area and water bodies, on the other hand, may increase by 2030 by 22.5%, 87.3%, and 13.3%, respectively.

4. Discussion

4.1 Selection of the best model for LULC prediction

To predict the future LULCC, we have used a range of modelling approaches (including machine learning and statistical technique, proximity techniques, and the hybrid of these approaches), namely, MLP-MC, LR-MC, LR-CA-MC, and MLP-CA-MC. Based on our

comparison, MLP-MC model provided the best results for the study area on a range of spatial and temporal performance metrics and was subsequently used for the prediction of future LULC. The ability of neural networks to model the nonlinear relationship between the dependent and independent variables may be responsible for this (Schultz, Wieland, and Lutze 2000; Mahiny and Turner 2003). Our finding that MLP-MC model outperforms LR-MC/CA-MC models is consistent with Ozturk, (2015); Aguedad, Houet, Hubert-Moy (2017), and Islam, Rehman, and Jashimuddin (2018); but these researchers only used non-hybrid approaches. Our model inter-comparison also shows that LR-CA-MC model also outperforms the LR-MC model, which is consistent with the findings of Polemans and Rompaey (2009); Arsanjani et al. (2013); Munshi et al. (2014). Additionally, the present study also attempted to develop the generalised understanding of hybridisation of modelling theories dealing with LULCC. The overall finding suggests that non-hybrid approach (MLP-MC) may outperform hybrid approaches (MLP-CA-MC and LR-CA-MC) in the prediction of LULC.

4.2 Utility of explanatory variables

As per Yuzer and Yuzer (2006), the purpose of predicting the LULCC is to obtain a better understanding of potential transitions, locations, and drivers of LULCC rather than an accurate estimation of future changes. Usually, at the high elevation and steep slopes, the forest areas remain intact due to less logging activities. In the Subarnarekha basin, most of the forest areas are situated at higher elevations and steep slopes, but contrary to the expected trend, these areas suffered from deforestation during 1989-2011. Only Adhikari, Fik, and Dwivedi (2017) reported similar results for a national park in South India. Our findings are contrary to many studies (Wear and Bolstad 1998; Munroe, Southworth, and Tucker 2004; Lee, Choi, and Lee 2018; Birhane et al. 2019) that reported a negative correlation between deforestation and intense topographic factors (high elevation and steeper slopes).

Distance to a nearby city is a vital proximity variable for LULCC as the areas near cities are highly susceptible to the cultivation of commercial/cash crops and to development. The distance to the reservoir is an essential driver to LULCC as the probability of deforestation is expected to be higher near reservoirs due to regular access to water for different applications. Furthermore, proximity to reservoirs is treasured for plantation crops like eucalyptus. Our findings are consistent with many studies that utilised the impact of distance to nearby cities and distance to water bodies on multidimensional LULCC (Poelmans and Rompaey 2010;

Arsanjani et al. 2013; Nourqolipour et al. 2016; Hamdy et al. 2016; Peng et al. 2017). We found two socio-economic drivers, i.e., population and population density, contributing to LULCC, which is consistent with Fei et al. (2015), Li et al. (2016), Munthali et al. (2019) who reported the significant contribution of demographic variables to LULC dynamics. The importance of socio-economic explanatory factors provides future research opportunities to incorporate long-term projections of these (e.g. Riahi et al. 2017) into LULCC modelling, given the higher sensitivity of integrated assessment models results of land-use change to socio-economic rather than climate change (Harrison et al. 2016; Holman et al. 2017).

4.3 Regional Vs National Scale

Since the same land use planning policies are applicable at regional and national scales, we performed the study at the regional level. We attempted to compare the dynamics of each land cover category at the regional level with those reported at the national level, e.g., Tian et al. (2014); Roy et al. (2015); Moulds, Buytaert, and Mijic (2018). Indian LULCC statistics at the national level have shown a significant increase in agricultural lands and built-up areas during 1985-2010 at the expense of forest lands, scrublands, grassland and barren lands (Tian et al. 2014; Roy et al. 2015; Moulds, Buytaert, and Mijic 2018).

We found that the regional trends in LULC differ considerably from the national trends. At the regional level, e.g., Subarnarekha basin, trends in dense forest and built-up area are consistent with the national decreasing trends; however, for scrubland, agricultural lands, and barren lands, regional trends are contradictory to the national trends. This mismatch results from the substantial heterogeneity of environmental, economic and agronomic conditions across the regions in India that lead to differences in the regional drivers of LULCC. Despite increasing trends in agricultural intensification across India, decreasing trends in the Subarnarekha basin show the disparity among regions due to, amongst other reasons, to uneven development in irrigation facilities and insufficient propagation of the modern cultivation practices (Singh and Giri 2018).

Our findings suggest that future LULC scenarios for the nation should be informed by regional analyses that capture the influence of region-specific drivers, rather than directly from an assumption of nationally-homogenous drivers. Such improvements would provide an enhanced understanding of national and regional LULCC and their impact on the environment and may add to the planning and management of natural resources.

5. Conclusions

The natural resources of the Subarnarekha basin are facing rapid degradation due to increasing anthropogenic influences. Conversion of the dense forest into built-up area and scrubland is widespread in the basin. To better understand past (1989-2011) and future trends in LULC changes in the basin, we propose a transferable methodology for the systematic comparison of (hybrid and non-hybrid models) LULC models against multiple performance metrics to identify the best modelling approach for the subsequent projection of future LULC to 2020 and 2030. MLP-MC model outperformed other models in reproducing the nature, extent and location of LULCC in the study area. However, as the study focuses on providing a transferable methodological framework for LULC modelling, the results obtained here do not establish the universal superiority of the model for LULCC prediction.

We also compared the regional land cover trends in Subarnarekha basin for consistency with the national land cover trends. The LULCC trends of dense forest and built-up area were found to be consistent with the national trends. In contrast, the trends in agricultural land, scrubland, and barren land were found to differ from the national trends, demonstrating the importance of regional differences in driving forces. Our study shows that national LULCC assessments should be supplemented by an understanding of regional LULCC drivers for providing improved LULCC projections at the scales needed by policymakers, land-use planners and environmentalists to devise both national policy and management recommendations.

References

Adhikari, Sanchayeeta, Timothy Fik, and Puneet Dwivedi. 2017. “Proximate Causes of Land-Use and Land-Cover Change in Bannerghatta National Park: A Spatial Statistical Model.” *Forests* 8 (9): 1–23. doi:10.3390/f8090342.

Agarwal, Chetan, Glen M. Green, J. Morgan Grove, Tom P. Evans, Charles M Schweik, and Charles M. Schweik. 2010. “A Review and Assessment of Land-Use Change Models: Dynamics of Space, Time, and Human Choice.” *General Technical Report NE-297* 1: 1–61. doi:10.1289/ehp.6514.

Aguejdad, Rahim, Thomas Houet, and Laurence Hubert-Moy. 2017. “Spatial Validation of Land Use Change Models Using Multiple Assessment Techniques: A Case Study of Transition Potential Models.” *Environmental Modeling and Assessment* 22 (6): 591–606. doi:10.1007/s10666-017-9564-4.

Arsanjani, Jamal Jokar, Marco Helbich, Wolfgang Kainz, and Ali Darvishi Boloorani. 2013. “Integration of Logistic Regression, Markov Chain and Cellular Automata Models to Simulate Urban Expansion.” *International Journal of Applied Earth Observation and*

Geoinformation 21 (1): 265–275. doi:10.1016/j.jag.2011.12.014.

Basnyat, Bijendra. 2009. "Impacts of Demographic Changes on Forests and Forestry in Asia and the Pacific." *Agriculture*, 1–82.

Behera, Mukunda D., Santosh N. Borate, Sudhindra N. Panda, Priti R. Behera, and Partha S. Roy. 2012. "Modelling and Analyzing the Watershed Dynamics Using Cellular Automata (CA)-Markov Model - A Geo-Information Based Approach." *Journal of Earth System Science* 121 (4): 1011–1024. doi:10.1007/s12040-012-0207-5.

Birhane, Emiru, Haregeweini Ashfare, Ayele Almaw Fenta, Hadgu Hishe, Mewcha Amha Gebremedhin, Hailemariam G. wahed, and Negasi Solomon. 2019. "Land Use Land Cover Changes along Topographic Gradients in Hugumburda National Forest Priority Area, Northern Ethiopia." *Remote Sensing Applications: Society and Environment* 13: 61–68. doi:10.1016/j.rsase.2018.10.017.

Chavez, Pat S. 1996. "Image-Based Atmospheric Corrections - Revisited and Improved." *Photogrammetric Engineering and Remote Sensing* 62 (9): 1025–1036.

Chorley, R.J. and P. Haggett. 1965. "In Geographical Trend-Surface Mapping" *Transactions of the Institute of British Geographers* 37 (37): 47–67.

DeFries, R., and K. N. Eshleman. 2004. "Land-Use Change and Hydrologic Processes: A Major Focus for the Future." *Hydrological Processes* 18 (11): 2183–2186. doi:10.1002/hyp.5584.

Eastman, J.R. 2012. "IDRISI Selva. Version 17.02." Worcester, MA.: Clark Labs, Clark University.

Ellis, Erle. 2007. "Land-use and Land-cover Change and Climate Change." *The Encyclopedia of earth*,4.https://web.archive.org/web/20070503192914/http://www.eoearth.org/article/Land-use_and_land-cover_change.

Fei Li, Shuwen Zhang, Kun Bu, Jiuchun Yang, Qing Wang, and Liping Chang. 2015. "The Relationships between Land Use Change and Demographic Dynamics in Western Jilin Province." *Journal of Geographical Sciences* 25 (5): 617–636. doi:10.1007/s11442-015-1191-x.

Ghosh, Pramit, Anirban Mukhopadhyay, Abhra Chanda, Parimal Mondal, Anirban Akhand, Sandip Mukherjee, S. K. Nayak, Subhajit Ghosh, Debasish Mitra, Tuhin Ghosh, and Sugata Hazra. 2017. "Application of Cellular Automata and Markov-Chain Model in Geospatial Environmental Modeling- A Review." *Remote Sensing Applications: Society and Environment* 5: 64–77. doi:10.1016/j.rsase.2017.01.005.

Grekousis, George, Panos Manetos, and Yorgos N. Photis. 2013. "Modeling Urban Evolution Using Neural Networks, Fuzzy Logic and GIS: The Case of the Athens Metropolitan Area." *Cities* 30 (1): 193–203. doi:10.1016/j.cities.2012.03.006.

Guan, DongJie, HaiFeng Li, Takuro Inohae, Weici Su, Tadashi Nagaie, and Kazunori Hokao. 2011. "Modeling Urban Land Use Change by the Integration of Cellular Automaton and Markov Model." *Ecological Modelling* 222 (20–22): 3761–3772. doi:10.1016/j.ecolmodel.2011.09.009.

Hamdy, Omar, Shichen Zhao, Mohamed A. Salheen, and Y. Y. Eid. 2016. "Analyses the Driving Forces for Urban Growth by Using IDRISI®Selva Models Abouelreesh -

Aswan as a Case Study." *International Journal of Engineering and Technology* 9 (3): 226–232. doi:10.7763/ijet.2017.v9.975.

Harrison, Paula A, Robert W. Dunford, Ian P. Holman, Mark D.A. Rounsevell. 2016. "Climate change impact modelling needs to include cross-sectoral interactions. " *Nature Climate Change* 6(9): 885+ *10.1038/nclimate3039.*

Holman IP, C Brown, V Janes, D Sandars. 2017. "Can we be certain about future land use change in Europe? A multi-scenario, integrated-assessment analysis. " *Agricultural Systems* 15: 126–135

Hu, Zhiyong, and C. P. Lo. 2007. "Modeling Urban Growth in Atlanta Using Logistic Regression." *Computers, Environment and Urban Systems* 31 (6): 667–688. doi:10.1016/j.compenvurbsys.2006.11.001.

Islam, Kamrul, Md Farhadur Rahman, and Mohammed Jashimuddin. 2018. "Modeling Land Use Change Using Cellular Automata and Artificial Neural Network: The Case of Chunati Wildlife Sanctuary, Bangladesh." *Ecological Indicators* 88: 439–453. doi:10.1016/j.ecolind.2018.01.047.

Jose, Monish, and Martina Padmanabhan. 2016. "Dynamics of Agricultural Land Use Change in Kerala: A Policy and Social-Ecological Perspective." *International Journal of Agricultural Sustainability* 14 (3):307–324. doi:10.1080/14735903.2015.1107338.

Kamwi, Jonathan M., Moses A. Cho, Christoph Kaetsch, Samuel O. Manda, Friedrich P. Graz, and Paxie W. Chirwa. 2018. "Assessing the Spatial Drivers of Land Use and Land Cover Change in the Protected and Communal Areas of the Zambezi Region, Namibia." *Land* 7 (4). doi:10.3390/land7040131.

Kolb, Melanie, Jean François Mas, and Leopoldo Galicia. 2013. "Evaluating Drivers of Land-Use Change and Transition Potential Models in a Complex Landscape in Southern Mexico." *International Journal of Geographical Information Science* 27 (9): 1804–1827. doi:10.1080/13658816.2013.770517.

Lambin, Eric F., B. L. Turner, Helmut J. Geist, Samuel B. Agbola, Arild Angelsen, John W. Bruce, Oliver T. Coomes, et al. 2001. "The Causes of Land-Use and Land-Cover Change: Moving beyond the Myths." *Global Environmental Change* 11 (4): 261–269. doi:10.1016/S0959-3780(01)00007-3.

Lambin, Eric, F., and Geist, Helmut. 2006. "Land-use and land-cover change: Local Processes and global impacts". Springer.

Lee, Hyun Joo, Yun Eui Choi, and Sang Woo Lee. 2018. "Complex Relationships of the Effects of Topographic Characteristics and Susceptible Tree Cover on Burn Severity." *Sustainability* 10 (2). doi:10.3390/su10020295.

Li, Xiangmei, Ying Wang, Jiangfeng Li, and Bin Lei. 2016. "Physical and Socioeconomic Driving Forces of Land-Use and Land-Cover Changes: A Case Study of Wuhan City, China." *Discrete Dynamics in Nature and Society* 2016:1-11 doi:10.1155/2016/8061069.

Liu, Weiguo, and Karen C. Seto. 2008. "Using the ART-MMAP Neural Network to Model and Predict Urban Growth: A Spatiotemporal Data Mining Approach." *Environment and Planning B: Planning and Design* 35 (2): 296–317. doi:10.1068/b3312.

Mahiny, Abdolrassoul S, and Brian J Turner. 2003. "Modeling Past Vegetation Change

through Remote Sensing and GIS: A Comparison of Neural Networks and Logistic Regression Methods." *Change* 1–24.

Mas, J. F., and J. J. Flores. 2008. "The Application of Artificial Neural Networks to the Analysis of Remotely Sensed Data." *International Journal of Remote Sensing* 29 (3): 617–663. doi:10.1080/01431160701352154.

Mas, Jean François, Melanie Kolb, Martin Paegelow, María Teresa Camacho Olmedo, and Thomas Houet. 2014. "Inductive Pattern-Based Land Use/Cover Change Models: A Comparison of Four Software Packages." *Environmental Modelling and Software* 51: 94–111. doi:10.1016/j.envsoft.2013.09.010.

Moulds, Simon, Wouter Buytaert, and Ana Mijic. 2018. "Data Descriptor: A Spatio-Temporal Land Use and Land Cover Reconstruction for India from 1960-2010." *Scientific Data* 5: 1–11. doi:10.1038/sdata.2018.159.

Munroe, Darla K., Jane Southworth, and Catherine M. Tucker. 2004. "Modeling Spatially and Temporally Complex Land-Cover Change: The Case of Western Honduras." *Professional Geographer* 56 (4): 544–559. doi:10.1111/j.0033-0124.2004.00447.x.

Munshi, Talat, Mark Zuidgeest, Mark Brussel, and Martin van Maarseveen. 2014. "Logistic Regression and Cellular Automata-Based Modelling of Retail, Commercial and Residential Development in the City of Ahmedabad, India." *Cities* 39:68–86. doi:10.1016/j.cities.2014.02.007.

Munthali, Maggie G., Nerhene Davis, Abiodun M. Adeola, Joel O. Botai, Jonathan M. Kamwi, Harold L.W. Chisale, and Oluwagbenga O.I. Orimoogunje. 2019. "Local Perception of Drivers of Land-Use and Land- Cover Change Dynamics across Dedza District, Central Malawi Region." *Sustainability* 11 (3). doi:10.3390/su11030832.

Nourqolipour, Ramin, Abdul Rashid B.Mohamed Shariff, Siva K. Balasundram, Noordin B. Ahmad, Alias M. Sood, and Taher Buyong. 2016. "Predicting the Effects of Urban Development on Land Transition and Spatial Patterns of Land Use in Western Peninsular Malaysia." *Applied Spatial Analysis and Policy* 9 (1): 1–19. doi:10.1007/s12061-014-9128-9.

Overmars, Koen P., and Peter H. Verburg. 2005. "Analysis of Land Use Drivers at the Watershed and Household Level: Linking Two Paradigms at the Philippine Forest Fringe." *International Journal of Geographical Information Science* 19 (2): 125–152. doi:10.1080/13658810410001713380.

Ozturk, Derya. 2015. "Urban Growth Simulation of Atakum (Samsun, Turkey) Using Cellular Automata-Markov Chain and Multi-Layer Perceptron-Markov Chain Models." *Remote Sensing* 7 (5): 5918–5950. doi:10.3390/rs70505918.

Peng, Jian, Mingyue Zhao, Xiaonan Guo, Yajing Pan, and Yanxu Liu. 2017. "Spatial-Temporal Dynamics and Associated Driving Forces of Urban Ecological Land: A Case Study in Shenzhen City, China." *Habitat International* 60:81–90. doi:10.1016/j.habitatint.2016.12.005.

Poelmans, Lien, and Anton Van Rompaey. 2009. "Detecting and Modelling Spatial Patterns of Urban Sprawl in Highly Fragmented Areas: A Case Study in the Flanders-Brussels

Region." *Landscape and Urban Planning* 93(1):10–19. doi:10.1016/j.landurbplan.2009.05.018.

Poelmans, Lien, and Anton Van Rompaey. 2010. "Complexity and Performance of Urban Expansion Models." *Computers, Environment and Urban Systems* 34 (1):17–27. doi:10.1016/j.compenvurbsys.2009.06.001.

Pontius, Robert Gilmore, and Laura C. Schneider. 2001. "Land-cover change model validation by an ROC method for the Ipswich watershed, Massachusetts, USA." *Agriculture, Ecosystems and Environment* 85:239–248

Pontius, Robert G., Emily Shusas, and Menzie McEachern. 2004. "Detecting Important Categorical Land Changes While Accounting for Persistence." *Agriculture, Ecosystems and Environment* 101 (2–3): 251–268. doi:10.1016/j.agee.2003.09.008.

Pontius, Robert Gilmore, and Marco Millones. 2011. "Death to Kappa: Birth of Quantity Disagreement and Allocation Disagreement for Accuracy Assessment." *International Journal of Remote Sensing* 32(15): 4407–4429. doi:10.1080/01431161.2011.552923.

Riahi, Keywan, Detlef P. van Vuuren, Elmar Kriegler, Jae Edmonds, Brian C. O'Neill et al (2017). "The Shared Socioeconomic Pathways and their energy, land use, and greenhouse gas emissions implications: An overview. *Global Environmental Change* 42: 153-168 https://doi.org/10.1016/j.gloenvcha.2016.05.009

Roy, P. S., and Arijit Roy. 2010. "Land Use and Land Cover Change in India: A Remote Sensing & GIS Perspective." *Journal of the Indian Institute of Science* 90(4): 489–502.

Roy, Parth S., Arijit Roy, Pawan K. Joshi, Manish P. Kale, Vijay K. Srivastava, Sushil K. Srivastava, Ravi S. Dwevidi, et al. 2015. "Development of Decadal (1985-1995-2005) Land Use and Land Cover Database for India." *Remote Sensing* 7(3): 2401–2430. doi:10.3390/rs70302401.

Schultz, A., R. Wieland, and G. Lutze. 2000. "Neural Networks in Agroecological Modelling - Stylish Application or Helpful Tool?" *Computers and Electronics in Agriculture* 29 (1–2): 73–97. doi:10.1016/S0168-1699(00)00137-X.

Serneels, S., M. Y. Said, and E. F. Lambin. 2001. "Land Cover Changes around a Major East African Wildlife Reserve: The Mara Ecosystem (Kenya)." *International Journal of Remote Sensing* 22 (17): 3397–3420. doi:10.1080/01431160152609236.

Singh, Abhay Kumar, and Soma Giri. 2018. "Subarnarekha River: The Gold streak of India." *The Indian Rivers*, Springer Hydrogeology. doi:10.1007/978-981-10-2984-4.

Song, Xiao Peng, Matthew C. Hansen, Stephen V. Stehman, Peter V. Potapov, Alexandra Tyukavina, Eric F. Vermote, and John R. Townshend. 2018. "Global Land Change from 1982 to 2016." *Nature* 560 (7720): 639–643. doi:10.1038/s41586-018-0411-9.

Tian, Hanqin, Kamaljit Banger, Tao Bo, and Vinay K. Dadhwal. 2014. "History of Land Use in India during 1880-2010: Large-Scale Land Transformations Reconstructed from Satellite Data and Historical Archives." *Global and Planetary Change* 121: 78–88. doi:10.1016/j.gloplacha.2014.07.005.

Tu, Jack V. 1996. “Advantages and Disadvantages of Using Artificial Neural Networks versus Logistic Regression for Predicting Medical Outcomes.” *Journal of Clinical Epidemiology* 49 (11): 1225–1231. doi:10.1016/S0895-4356(96)00002-9.

Turner. 2006. “Land change as a forcing function in global environmental change”. In: Geist HJ (ed) Our Earth's changing land: An encyclopedia of land use and land cover change, 1 (A-K). Greenwood Press, Westport< London, pp xxv-xxxii.

Václavík, Tomáš, and John Rogan. 2009. “Identifying Trends in Land Use/Land Cover Changes in the Context of Post-Socialist Transformation in Central Europe: A Case Study of the Greater Olomouc Region, Czech Republic.” *GIScience & Remote Sensing* 46 (1): 54–76. doi:10.2747/1548-1603.46.1.54.

Veldkamp, A., and Lambin, E.F. Lambin. 2001. “Predicting Land Use Change.” *Agriculture, Ecosystems and Environment* 85 (2001): 1–6.

Wear, David N., and Paul Bolstad. 1998. “Land-Use Changes in Southern Appalachian Landscapes: Spatial Analysis and Forecast Evaluation.” *Ecosystems* 1 (6): 575–594. doi:10.1007/s100219900052.

Wu, Fulong. 2002. “Calibration of Stochastic Cellular Automata: The Application to Rural-Urban Land Conversions.” *International Journal of Geographical Information Science* 16 (8): 795–818. doi:10.1080/13658810210157769.

Xie, Chenglin, Bo Huang, Christophe Claramunt, Magesh Chandramouli. 2000. “Spatial Logistic Regression and GIS to Model Rural-Urban Land Conversion” *Journal of Polymer Science. Part A-1, Polymer Chemistry* 18 (11): 3133–3147.

Yüzer, M.A., and Ş.Yüzer. 2006. “Cellular Automata Tabanlı LUCAM Modeli ile İstanbul’un Gelişim ve Dönüşümüne İlişkin Makro Form Simülasyonları.” J. Kültür Univ. 4, 231–244.

Zhang, D., C. Narteau, and O. Rozier. 2010. “Morphodynamics of Barchan and Transverse Dunes Using a Cellular Automaton Model.” *Journal of Geophysical Research: Earth Surface* 115 (3). doi:10.1029/2009JF001620.

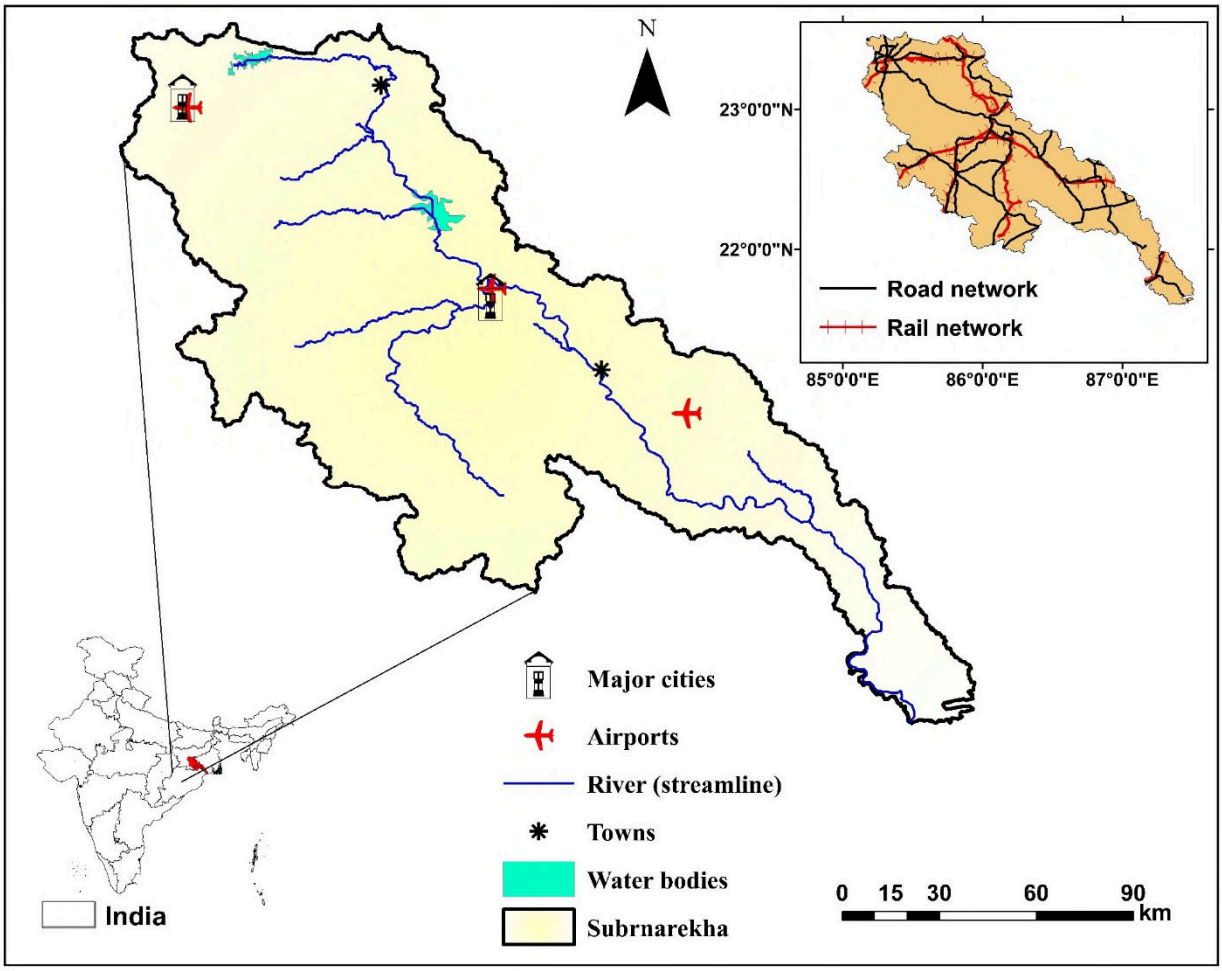

Figure 1. Index map of Subarnarekha basin

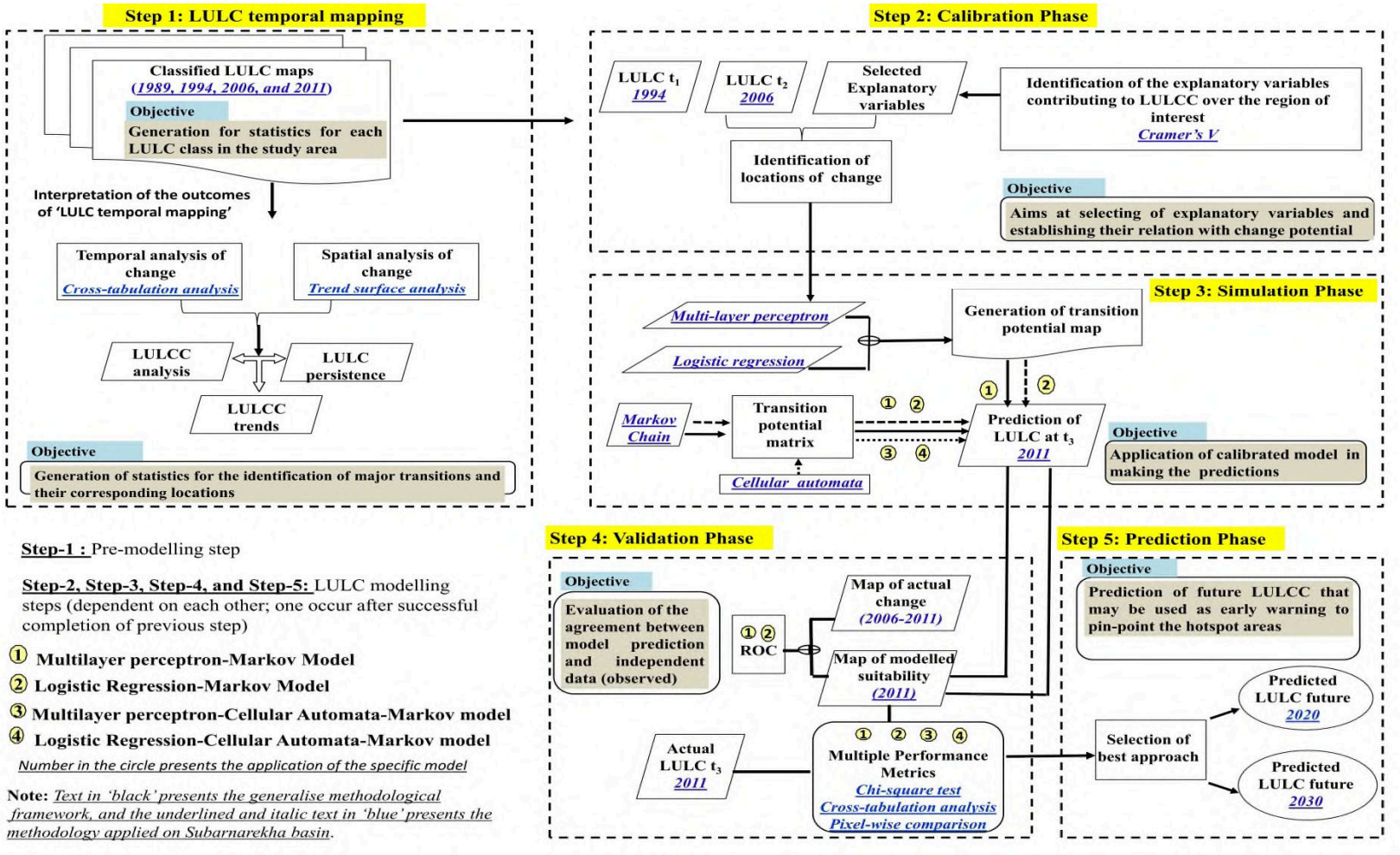

Figure 2. Overview of the five steps of the workflow for the methodology from spatio-temporal analysis of historic LULC mapping (Step 1), identification and testing of explanatory variables of LULCC (step 2), calibration and validation / evaluation of individual LULC models against historic LULC data (steps 3 and 4) and application of the best performing model for prediction of future LULC (step 5)

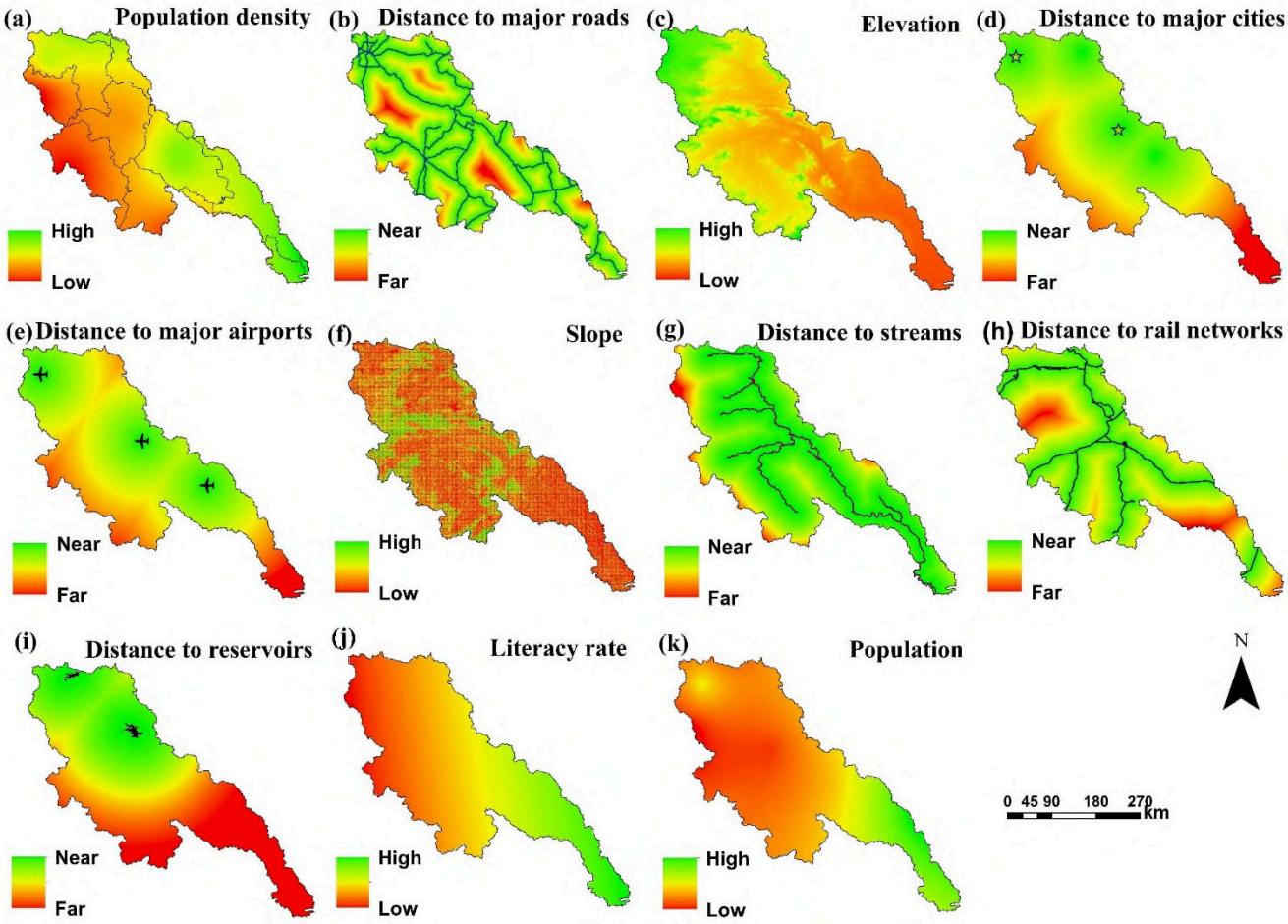

Figure 3. Explanatory variables employed for calibration by utilising 2006 data: a) population density b) distance to major roads c) elevation d) distance to major cities e) distance to major airports f) slope g) distance to streams h) distance to rail networks i) distance to reservoirs j) literacy rate k) population

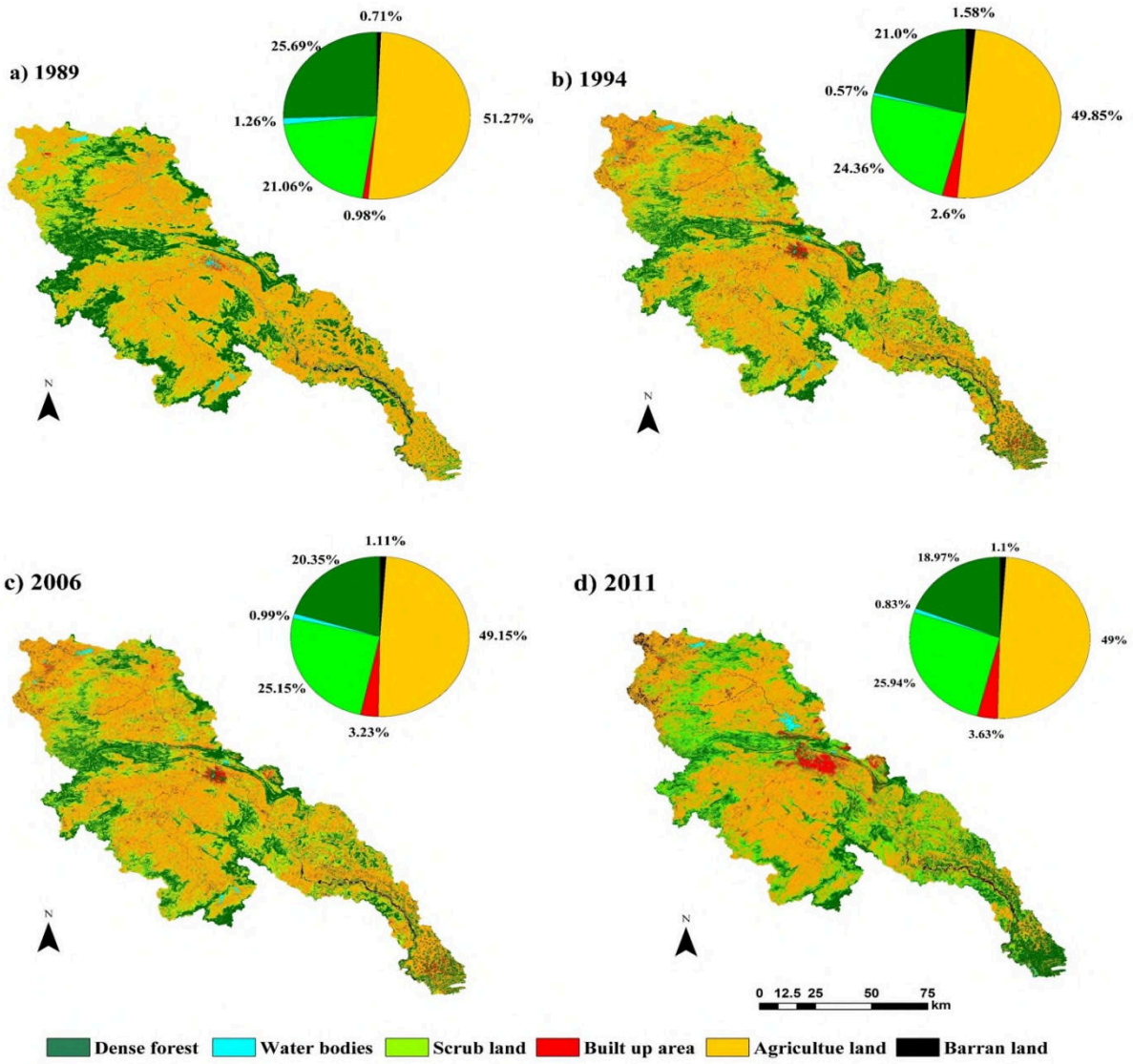

Figure 4. Maps and summary percentages of LULC in the Subarnarekha basin for a) 1989 b) 1994 c) 2006 d) 2011 based on unsupervised classification using Landsat TM and ETM imageries

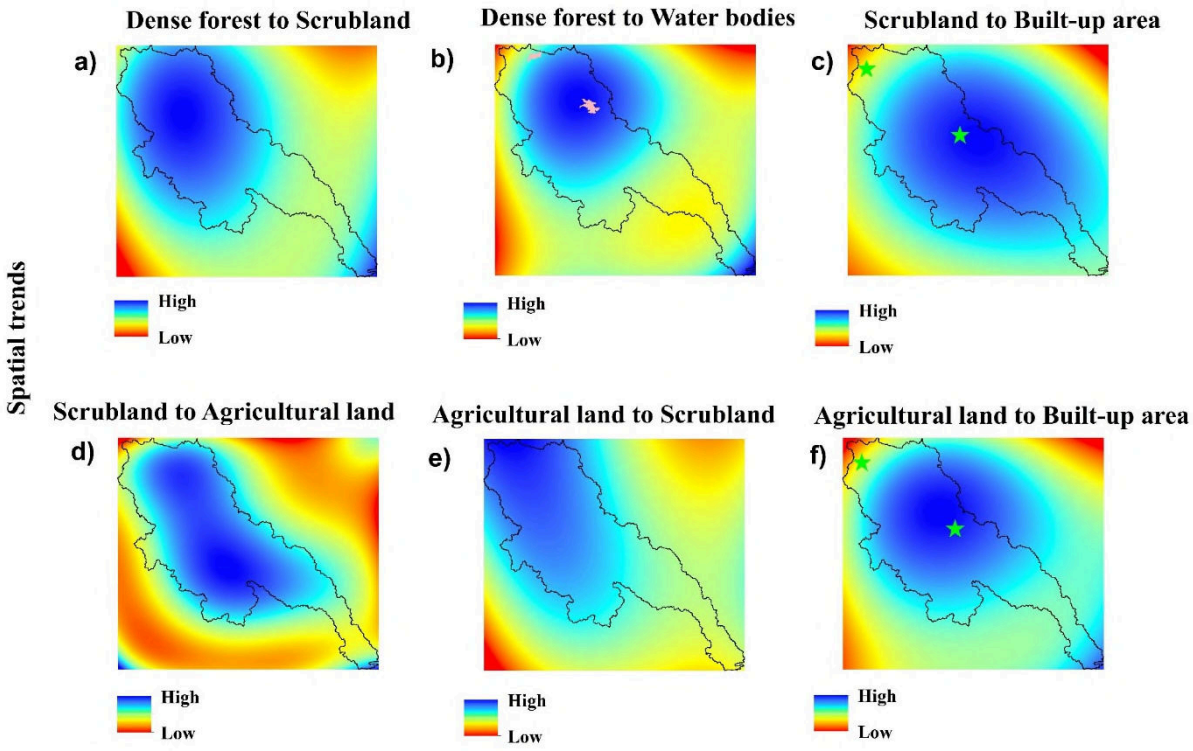

Figure 5. Spatial dynamics of change from trend surface analysis from a) dense forest to scrubland b) dense forest to waterbodies c) scrubland to built-up area d) scrubland to agricultural land e) agricultural land to scrubland f) agricultural land to built-up area

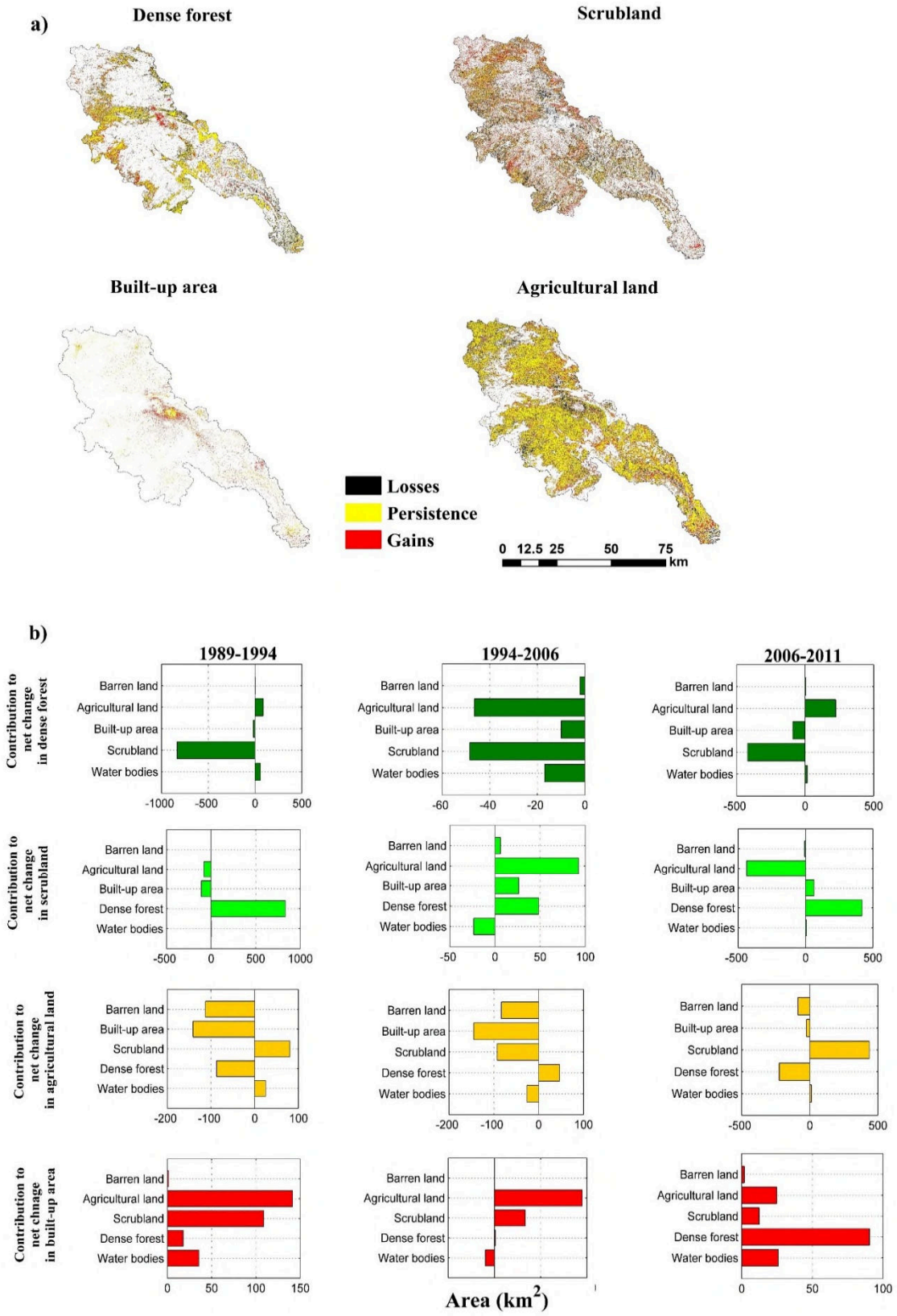

Figure 6. Past LULC change in the Subarnarekha basin between 1989-2011, showing (a) spatial maps of gains, losses, and persistence in major LULC classes and (b) net area contribution of changes in LULC classes to each major class for three sub-intervals, from cross-tabulation analysis of the LULC maps for 1989, 1994, 2006 and 2011

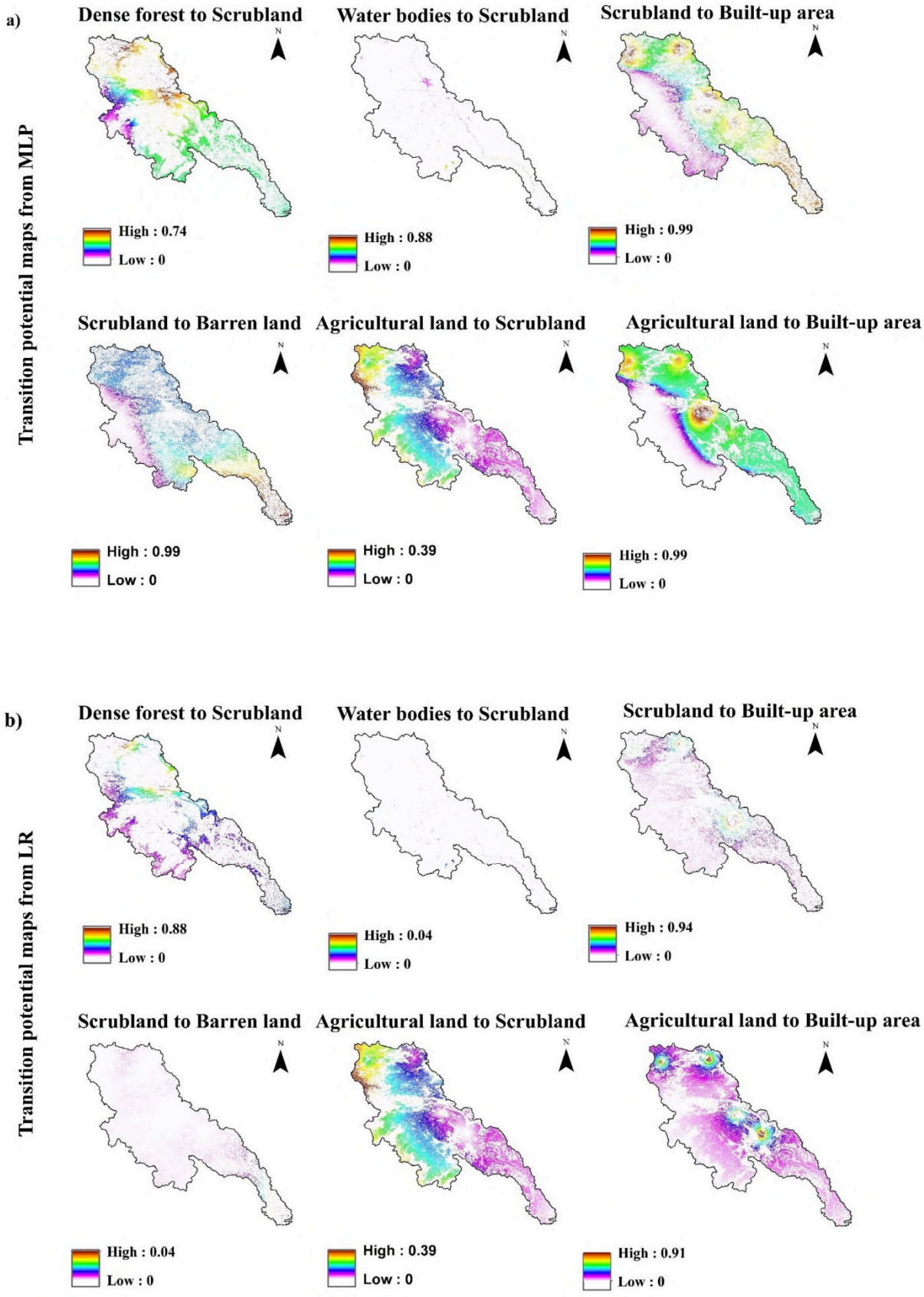

Figure 7. Transition potential maps obtained from a) multilayer perceptron b) logistic regression methods for 1994-2006

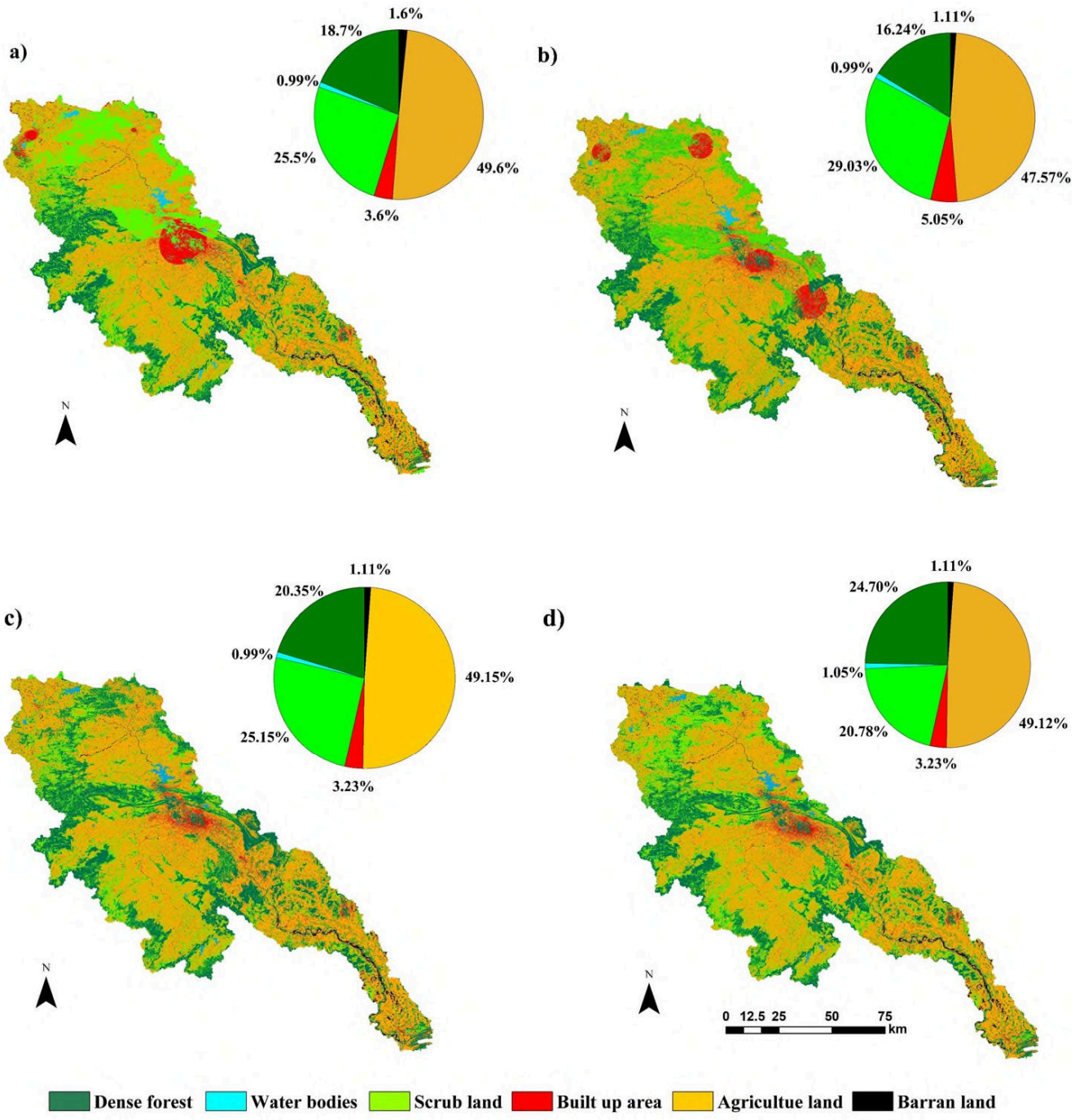

Figure 8 Maps and summary percentages of predicted LULC in the Subarnarekha basin for the 2011 using a) MLP-MC model b) LR-MC model c) LR-CA-MC model d) MLP-CA-MC model informed by LULC maps of 1994 and 2006

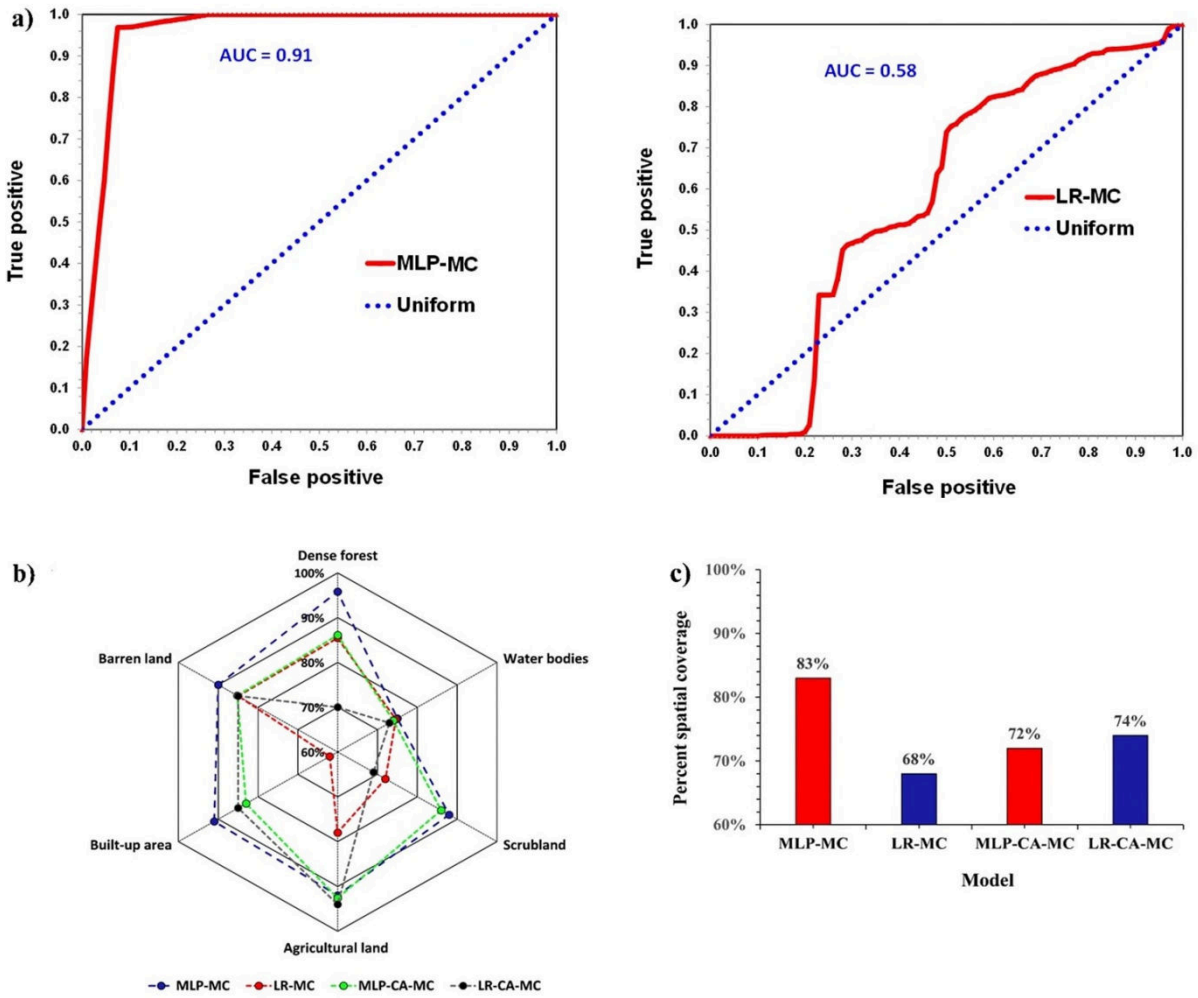

Figure 9. Validation of LULCC predictions a) ROC analysis for models generating transition potential maps b) amount of persistence between the simulated and actual LULC map of 2011 c) amount of spatial coverage between the simulated and actual LULC map of 2011

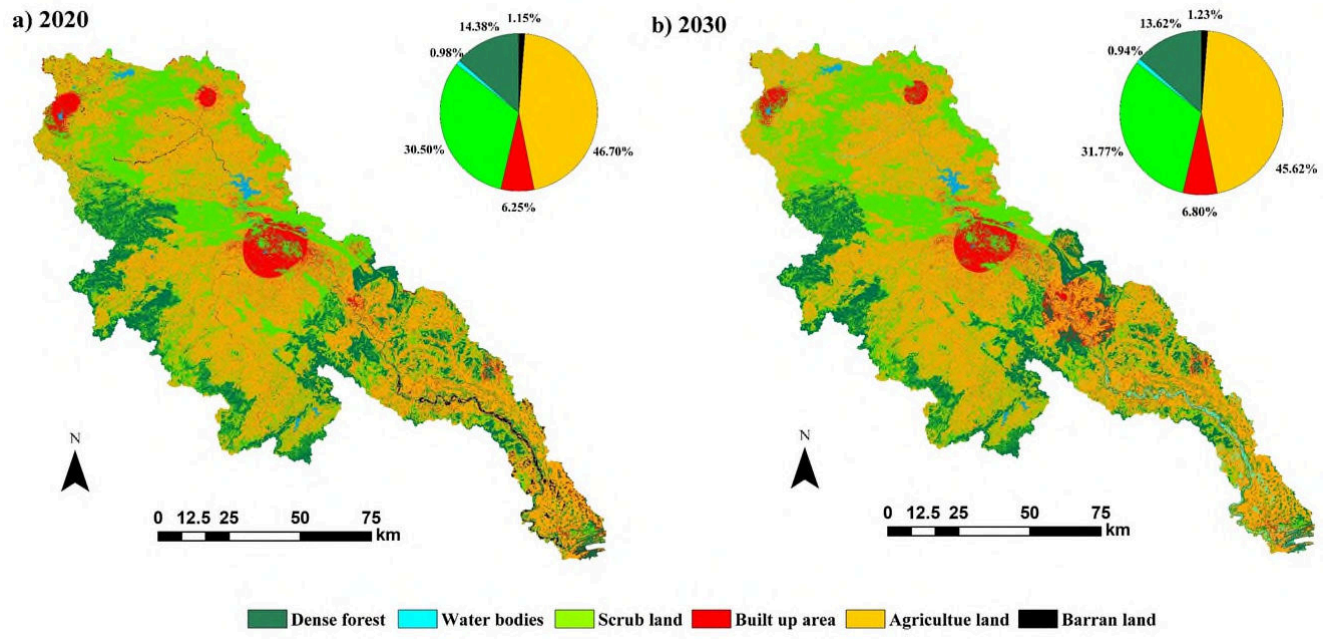

Figure 10. Maps and summary percentages of predicted future LULC in the Subarnarekha basin for a) 2020 and b) 2030 using the validated MLP-MC model

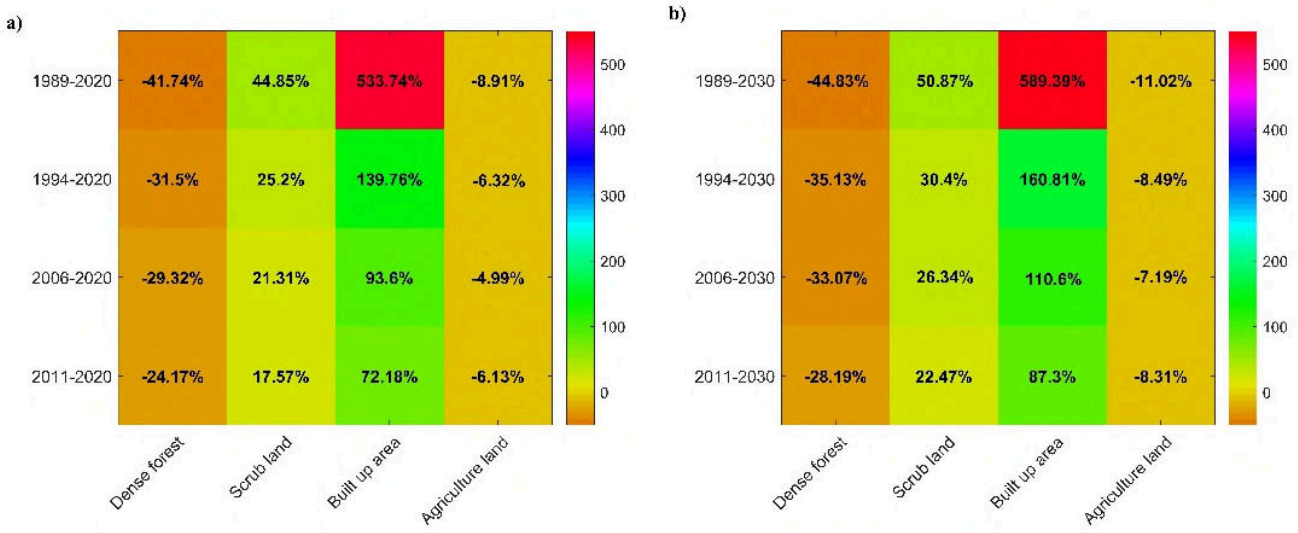

Figure 11. Heat map representing the percent change LULC classes during a) 2020 and b) 2030 from historical time periods 1989, 1994, 2006, and 2011 respectively.

Table 1

List of geospatial and attribute data used in the study

Data	**Source**	**Resolution (m)**	**Year**
Satellite imagery Landsat TM and Landsat ETM	US Geological Survey	30	1989, 1994, 2006, and 2011
DEM	US Geological Survey (SRTM)	30	
Cities	Google Earth		2006 and 2011
Airports	Google Earth		2006 and 2011
Water bodies	Google Earth		2006 and 2011
Road network	Survey of India (SOI) toposheets		
Rail network	Survey of India toposheets		
Population	Indiastat	At district level	1991, 2001, 2011
Population density	Indiastat	-do-	1991, 2001, 2011
Literacy rate	Indiastat	-do-	1991, 2001, 2011

Table 2

Cross-tabulation analysis of 1994-2006. All values are presented in km^2 with persistence (diagonal elements; highlighted) expressed in % of total area in 2006 (**D**- dense forest, **W**- water bodies, **S**- scrubland, **Bu**- built-up area, **A**- agricultural land, and **Ba**- barren land)

	D	**W**	**S**	**Bu**	**A**	**Ba**	**1994**
D	2587.03 ***(63.76%)***	11.92	991.87	42.21	257.00	2.12	3892.15
W	28.83	82.12 ***(74.42%)***	24.34	15.15	30.03	9.04	189.50
S	1070.31	0.67	2081.56 ***(44.17%)***	9.09	1638.94	8.61	4809.19
Bu	43.14	5.36	42.73	341 ***(68.50%)***	178.31	7.44	617.98
A	323.44	3.99	1576.8	83	7213.13 ***(75.54%)***	199.75	9400.11
Ba	4.22	6.29	2.53	7.34	106.34	86.11 (***27.49%)***	212.83
2006	4056.97	110.35	4719.83	497.79	9423.75	313.23	19121.76 ***(Total)***

Table 3

Summary of LULCC (1994-2006) presented as % of basin area

	Gains (%)	**Loss (%)**	**Net change (%)**	**Total change (%)**	**Swap (%)**
Dense forest	8.83	5.14	3.69	13.97	10.28
Water bodies	0.83	0.14	0.69	0.96	0.27
Scrubland	11.63	14.94	3.31	***26.57***	***23.26***
Built-up area	0.73	2.35	1.62	3.07	***1.45***
Agricultural land	11.77	10.35	1.42	22.12	20.70
Barren land	0.34	1.21	0.87	1.55	0.68

Table 4

Cramer's V value for driving variables

Explanatory Variable	**Cramer's V**		
	1989-1994	1994-2006	2006-2011
Elevation	0.49	0.51	0.54
Slope	0.49	0.50	0.53
Distance to cities	0.31	0.33	0.38
Distance to reservoir	0.15	0.20	0.28
Distance to streams	0.08	0.09	0.09
Distance to roads	0.12	0.12	0.12
Distance to rail	0.09	0.09	0.09
Distance to airports	0.10	0.10	0.10
Literacy rate	0.04	0.05	0.08
Population	0.42	0.45	0.47
Population density	0.38	0.39	0.42

Table 5

Validation of LULCC prediction of Subarnarekha basin based on all four models

Land use class	Simulated land use in 2011 (O)	Actual land use in 2011 (E)	$(O - E)^2/E$
Chi-square test (MLP-MC Model)			
Dense forest	3575.77	3627.40	0.73
Water bodies	189.31	158.70	5.89
Scrubland	4876.05	4960.20	1.43
Built-up area	688.38	694.10	0.05
Agricultural land	9484.39	9369.60	1.40
Barren land	305.95	307.80	0.01
Note: $\chi^2 = \Sigma (O - E)^2/E = 9.51$; df = 5 and $\chi^2_{0.05}(5) = 11.07$			
Chi-square test (LR-MC Model)			
Dense forest	3105.37	3627.4	75.12
Water bodies	189.31	158.7	5.89
Scrubland	5602.67	4960.2	83.23
Built-up area	965.65	694.1	106.22
Agricultural land	9096.22	9369.6	7.98
Barren land	210.34	307.8	30.89
Note: $\chi^2 = \Sigma (O - E)^2/E = 309.33$; df = 5 and $\chi^2_{0.05}(5) = 7.815$			
Chi-square test (LR-CA-MC Model)			
Dense forest	3921.87	3627.4	23.90
Water bodies	191.21	158.7	6.65
Scrubland	4803.38	4960.2	4.95
Built-up area	598.51	694.1	13.17
Agricultural land	9398.34	9369.6	0.08
Barren land	212.25	307.8	29.70
Note: $\chi^2 = \Sigma (O - E)^2/E = 78.45$; df = 5 and $\chi^2_{0.05}(5) = 7.815$			
Chi-square test (MLP-CA-MC Model)			
Dense forest	4723.07	3627.4	330.96
Water bodies	200.77	158.7	11.15
Scrubland	3973.50	4960.2	196.27
Built-up area	617.63	694.1	8.42
Agricultural land	9392.60	9369.6	0.05
Barren land	212.25	307.8	29.69
Note: $\chi^2 = \Sigma (O - E)^2/E = 807.77$; df = 5 and $\chi^2_{0.05}(5) = 7.815$			

Supplementary material

Table S1
Details of land use/land cover classes

Classes	Description
Dense Forest	It includes tree canopy density of 70% and above.
Agriculture Land	It includes all cropland and pasture land.
Scrub Area	It include all the degraded forest lands with canopy density less than 10%.
Water Bodies	It include all rivers, lakes, ponds and canals.
Built-up Area	It include all residential, commercial and industrial areas and suburban areas
Barren land	It include all deserts, dry salts, flats, beaches, sand dunes.

https://dspace.org/

Spatio-temporal analysis of land use and land cover change: a systematic model inter-comparison driven by integrated modelling techniques

Gaur, Srishti; Mittal, Ateeksha; Bandyopadhyay, Arnab; Holman, Ian P.; Singh, Rajendra

2020-10-04
Taylor and Francis
https://doi.org/10.1080/01431161.2020.1815890
Gaur S, Mittal A, Bandyopadhyay A, et al., (2020) Spatio-temporal analysis of land use and land cover change: a systematic model inter-comparison driven by integrated modelling techniques. International Journal of Remote Sensing, Volume 41, Issue 23, 2020, pp. 9229-9255

Downloaded from a DSpace repository. https://dspace.org/