Simulation of Land Use and Land Cover Change and Urban Sprawl Prediction in Lucknow Metropolitan Area Using Markov Chain Model

Vijay Kumar Singh Rathore[1,], Sudhir Kumar Singh[2], Anjani Kumar Singh[1]*

[1] Department of Geography, K.S. Saket P.G. College, Ayodhya, Uttar Pradesh. India, vksrathore707@gmail.com, anjanikumarsingh358@gmail.com

[2] K. Banerjee Centre of Atmospheric and Ocean Studies, IIDS, Nehru Science Centre, University of Allahabad, Prayagraj, Uttar Pradesh, India, sudhirinjnu@gmail.com

* corresponding author

doi: 10.5281/zenodo.11657555

Abstract: This study aims to simulate LULC and predict urban sprawl. Predictive modeling was conducted using the Land Change Modeler (LCM) from Clark Labs, employing advanced algorithms like Multi-Layer Perceptron-Neural Network (MLP-NN) and Markov Chain (MC). LCM was trained with 12 variables including elevation and distances from roads, rails, water bodies, forests, built-up areas, etc. The model prediction accuracy has been accessed by evaluating Receiver Operating Characteristic (ROC) values. The ROC/AUC values for agricultural land, vegetative cover, built-up, waterbody, scrub land and sodic land has been recorded as 0.62, 0.65, 0.91, 0.71, 0.79 and 0.81, respectively. The findings highlight a significant increase in built-up areas, indicative of urban sprawl, alongside decreases in agricultural land, wasteland, and tree cover from 2020 to 2030.

Keywords: land change modeller (LCM); multi-layer perceptron-neural network (MLP-NN); Markov chain (MC); Urban sprawl; physical drivers.

1 Introduction

In recent years, academic interest in local and regional Land Use and Land Cover change (LULCC) processes has surged. Assessing LULCC is crucial for addressing environmental issues like unplanned development, loss of arable land, and habitat destruction. The escalating population in major metropolitan cities of the world causes urban area expansion, altering Land Use and Land Cover (LULC) and inducing urban sprawl. Predicting LULCC and urban sprawl aids in planning for future infrastructure, healthcare, education, and service needs, essential for resource management. The LULCC research community has developed numerous spatially explicit models integrating geospatial data, aiding in better management of land resources and providing precise assessments, evaluations, and future scenario projections. (Sohl et al. 2016). LULC models are employed to analyse and forecast land use dynamics, offering insights into the socio-economic and environmental ramifications. These models, utilizing techniques like Markov chain, Artificial Neural Network, Cellular Automata, and Logistic Regression.

The Land Cover Change Model (LCM) of TerrSet is an advanced framework merging GIS and remote-sensing technologies to analyse and predict LULCC, employing intricate algorithms like Multi-Layer Perceptron-Neural Network (MLP-NN) and Markov Chain (MC). MLP-NN, a machine learning method, discerns complex data relationships, identifying drivers like land-use policies, economic dynamics, and population growth (Onate-Valdivieso and Sendra 2010).

The present study aims to simulate the LULCC in the study area and to know the future urban sprawl. The outcomes of the study will provide an insight for the planners to achieve the goals of sustainable development and for the management of natural resources effectively.

2 Materials and methods

Area under Master plan boundary – 2031 of Lucknow Municipal Corporation has been selected for this study (Figure 1).

The research utilized LANDSAT satellite data sourced from the United States Geological Survey (USGS) website. Decade- spanning satellite imagery from 2000, 2010, and 2020 was employed. All three satellite images were corrected radiometrically and atmospherically for classification process which performed using the pixel-based maximum likelihood supervised classification technique. In the present study, images were classified into 6 LULC such as Agricultural Land (AL), Vegetative Cover (VC), Built-up (BU), Waterbody (WB), Scrub Land (SCL) and

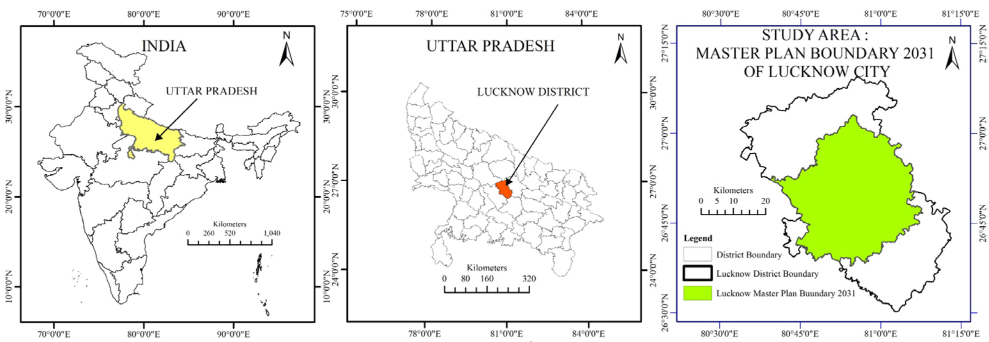

Figure 1. Study Area.

Table 1. Variable and their sources used in the study.

Sr. No.	Variables	Types of Variables	Source
1	Elevation	Static	SRTM DEM 30 M
2	Distance from Roads	Dynamic	Digitized from Google Earth
3	Distance from Railway Lines	Dynamic	Digitized from Google Earth
4	Distance from Waterbody	Dynamic	Digitized from Google Earth
5	Distance from Forest	Dynamic	Digitized from Google Earth
6	Distance from Built-up	Dynamic	Built-up layer is obtained from LULC classified map of 2010
7	Distance from Airport	Static	Digitized from Google Earth
8	Distance from Railway Stations	Static	Digitized from Google Earth
9	Distance from Bus Stands	Static	Digitized from Google Earth
10	Distance from Metro Stations	Static	Digitized from Google Earth
11	Distance from Hospitals	Static	Digitized from Google Earth
12	Population Density	Static	Population Density data of all the wards of Lucknow Municipal Corporation and villages were obtained from Census 2011.

Sodic Land (SOD). For accuracy assessment of classified images high resolution Google earth image is used as referenced data. Error matrix or confusion matrix were prepared for each classified images using these sample pints and referenced data. Kappa coefficient was calculated from confusion matrices. The majority analysis has been conducted on the classified images for better predictive results.

For Simulation of LULCC classified LULC data of 2000, 2010 and 2020 have been used. Along with 12 variables were used in the study as shown in Table 1 and Figure 2. Except elevation euclidian distance rasters were generated to all the variables and these variables were transformed to natural log scale for better prediction accuracy.

After identifying potential driver variables, the LCM multilayer perceptron (MLP) was used to generate transition potential maps. MLP, an artificial neural network capable of handling highly nonlinear variables (Sangermano et al. 2010), groups and models multiple transitions in a single sub-model

(Eastman et al. 2005). It incorporates only those driver variables with strong predictive ability into the

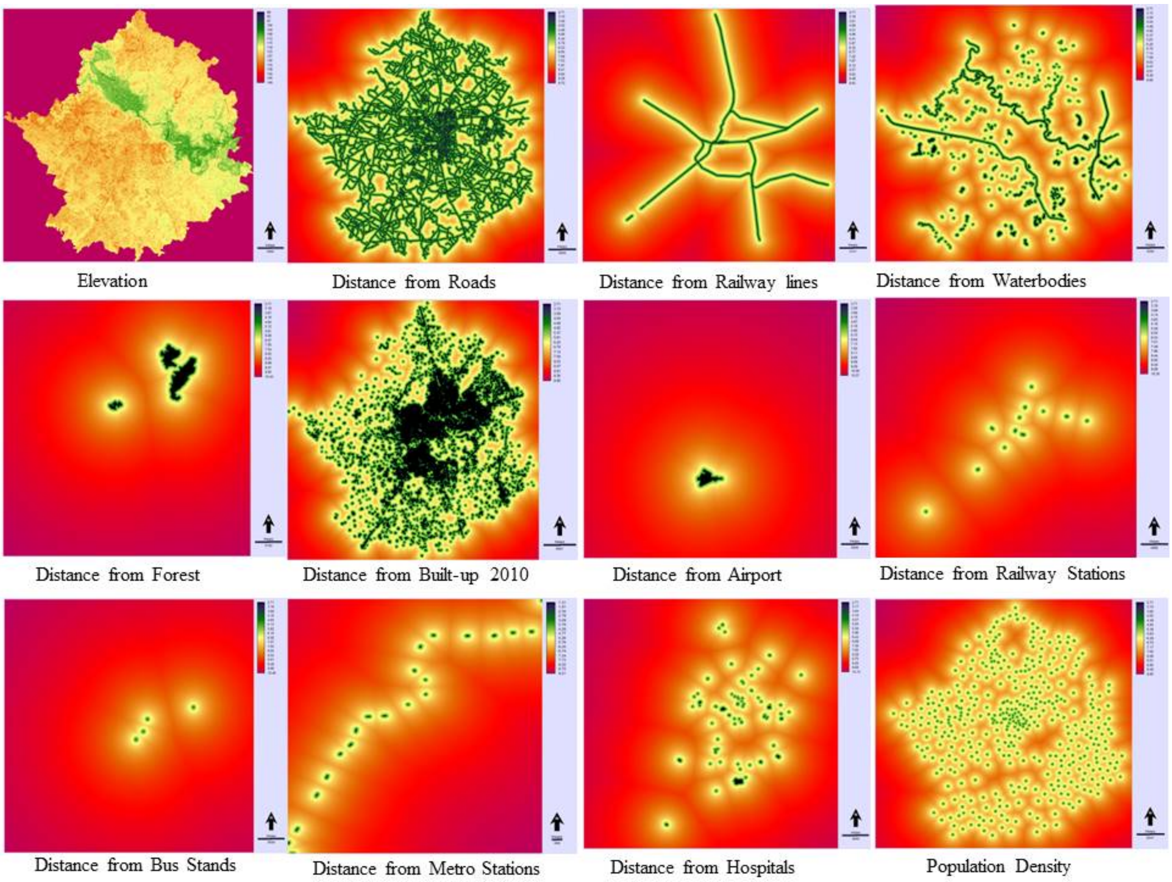

Figure 2. Variables used in the study.

computational process to produce various transition potential maps for each submodel, automatically representing the suitability for changes between LULC classes (Eastman 2012). In this study, the MLP was trained using predefined driver variables and LULC maps from 2000 (start) and 2010 (end) for model development. The Markov Chain model in LCM was applied to estimate future LULC changes based on previous land cover data and conversion probabilities. LCM generates hard and soft predicted maps using Markov Chain analysis, which provides a transition probability matrix quantifying each transition's change (Eastman 2015, Singh 2022). Validation of the prediction/simulation is an important aspect of any modelling as it assesses the model's predictive power and therefore statistical validation procedures has been applied to validate the predicted LULC map of the year 2020. Validation is performed through two ways - Figure of Merit (FOM) method and ROC method (Gidey et al. 2017).

3 Results

The LULC maps for 2000, 2010, and 2020 are given in Figures 3. The Kappa coefficients for 2000, 2010, and 2020 are 0.91, 0.90, and 0.92, respectively, indicating good classification performance with all values above 0.90.

A transition map has been created using LULC maps from 2000 and 2010. Over this period, 29 transitions were identified. All these transitions were used for modelling of transition potentials. To run transition sub models MLP neural network technique was adopted. The transition potential maps for each transition have been produced. For the prediction of future LULC which includes the built-up areas as well, Markov chain method is used. In this method initial LULC layer of 2000 and final LULC layer of 2010 were used to generate the transition probability matrix of each LULC classes from 2000 to 2010. Based on these probabilities a predicted map of LULC of 2020 was generated (Figure 4).

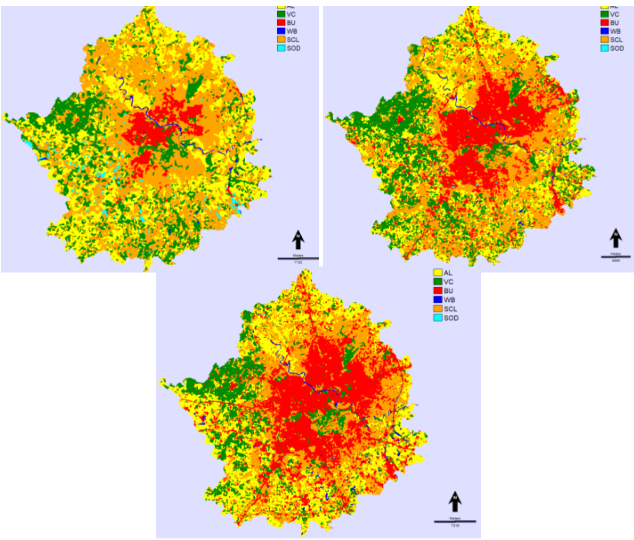

Figure 3. LULC 2000 (top left), 2010 (top right) and 2020 (bottom).

Using this predicted map of 2020 and actual classified map of LULC of 2020 validation process was conducted. In the FOM validation process a map showing hits, misses and false alarms has been generated. Counting each and applying in the formula of calculating FOM the accuracy of prediction is obtained. Here 4 hits, 10 misses and 40 false alarms are found. Thus, FOM accuracy is 7.4%. The ROC/AUC values recorded for agricultural land, vegetative cover, built-up, waterbody, scrub land, and sodic land are 0.62, 0.65, 0.91, 0.71, 0.79, and 0.81, respectively. The ROC index of 0.91 for built-up indicates that the model has performed very well in predicting this built-up class. Similarly, the ROC index values of 0.81 for sodic land and 0.79 for scrub land indicate good prediction accuracy for these classes. The ROC index values of 0.62 for agricultural land, 0.65 for vegetative cover, and 0.71 for waterbody suggest fair prediction accuracy for these categories. Accuracy of Predicted LULC of 2020 shows that this model can be used for further predictions. So for 2030 transition probability has been generated and a predicted layer of LULC of 2030 (Figure 4) was obtained from model.

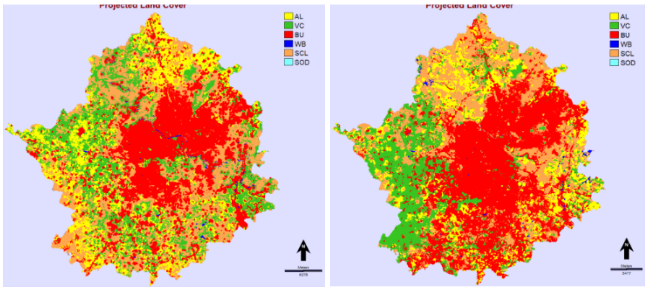

Figure 4. Predicted LULC map of 2020 (left) and 2030 (right).

4 Discussion

Area (hectare) for each LULC class in different years is given in the Table 2. In 2000 the expansion of the built up class was very less as compared to other LULC classes. It was just 5% of the total area. Scrub land was the dominant LULC class in 2000 followed by agricultural land. In 2010 Built-up class experienced a big change in areal extent and jumped to 20.53 % of the total study area. Scrub land remained the dominating class. In 2020 Area of built up again increased to 29.15 % of the total study area. The Agricultural area decreased significantly from 34,931.70 ha in 2000 to 27,128.07 ha in 2010 but then increased to 31,374.36 ha by 2020. This fluctuation might be due to urban expansion and conversion of agricultural lands to other uses, followed by efforts to reclaim or expand agricultural activities. Vegetative Cover category showed an increase from 24,357.60 ha in 2000 to 26,013.96 ha in 2010, likely due to reforestation or conservation initiatives. However, by 2020, it decreased sharply to 17,776.89 ha, possibly due to urbanization. The area under waterbody category slightly decreased from 804.24 ha in 2000 to 652.86 ha in 2010 but increased to 1,224.90 ha by 2020. In Scrub Lad class there was a decrease from 52,566.75 ha in 2000 to 42,848.73 ha in 2010, and further to 35,818.83 ha in 2020. This reduction could be attributed to land conversion for agriculture, urban development, or afforestation efforts replacing scrublands with more productive land uses. The predicted land use/land cover (LULC) map for 2020 within the study area indicates that built-up areas are the predominant category, followed by scrub land, vegetative cover, agricultural land, water bodies, and sodic land. Table 2

Table 2. Table showing area in hectare for LULC classes.

LULC class	2000	2010	2020	Predicted 2020	Predicted 2030
AL	34931.7	27128.07	31374.36	21832.92	18181.8
VC	24357.6	26013.96	17776.89	24578.19	22190.13
BU	7281	24984.81	35496.99	38877.3	50321.07
WB	804.24	652.86	1224.9	550.98	491.04
SCL	52566.75	42848.73	35818.83	35534.88	30174.66
SOD	1582.38	79.02	72.54	30.96	46.53

provides the area of each LULC class in the 2020 predicted LULC map. However, the actual classified LULC map for 2020 presents a different pattern, where scrub land is the dominant category, followed by built-up areas, agricultural land, vegetative cover, water bodies, and sodic land. The specific areas for each LULC class in the actual classified map are also provided in Table 2. Notably, the area under the built-up category differs between the predicted and actual classified maps: the actual classified map shows 35,496.99 ha, while the predicted map shows 38,877.3 ha, a difference of 3,380.31 ha. Despite this discrepancy, the error margin of 2.7% is considered negligible given the study area's total extent of 122,072 ha. Thus, the difference between the predicted and actual classified LULC maps does not significantly affect the overall conclusions of the study.

The LULC classes have been changed significantly in the predicted maps for 2030. In the 2030 predicted map the built-up category has become the dominant LULC class with an area of 50321.07 ha. These tremendous changes in built-up class indicate a significant transformation of the landscape in the study area over the coming years.

It is apparent from the predicted built-up maps of 2020 and 2030 that there is a gradual and continuous increase in the built-up area. This increase in built-up area will expend the areal extent of Lucknow city towards outer sides. The map of 2030 clearly explained the sprawling character built up class.

5 Conclusions

The accuracy of simulation model results depends on many factors. One key factor is the quality of the input data, especially multi-scale data from various satellite images at different resolutions. Using diverse data can introduce errors and uncertainty. Properly preparing this data, including correct classification of satellite images, is challenging but crucial. Mixed pixels can lower classification accuracy, and the choice of a classification algorithm is important. Every simulation model has internal limitations that affect accuracy. For simulating urban growth, all relevant factors (economic, social, ecological, etc.) must be considered.

Acknowledgments: The authors thank the USGS for free Landsat data and Google Earth for high-resolution data. No funding was granted by any authority or institution for the study. No funding was received. The authors declare that they have no conflict of interest.

6 References

Eastman JR. 2012. IDRISI Selva manual. Worcester, Mass: Clark labs-Clark University.

Eastman JR. 2015. TerrSet manual. Accessed in TerrSet version 18, 1-390.

Eastman JR, Van Fossen ME, Solarzano LA., 2005. Transition potential modeling for land cover change. GIS, Spatial Analysis and Modeling 17, 357-386.

Gidey, E., Dikinya, O., Sebego, R., Segosebe, E., Zenebe, A., 2017. Cellular automata and markov chain (CA_Markov) model-based predictions of future land use and land cover scenarios (2015–2033) in Raya, northern Ethiopia. Modeling Earth System Environment 3, 1245-1262.

Onate-Valdivieso F., Sendra J.B., 2010. Application of GIS and remote sensing techniques in generation of land use scenarios for hydrological modeling. Journal of Hydrology 395 (3-4), 256-263.

Sangermano, F., Eastman, J. R., Zhu, H., 2010. Similarity weighted instance-based learning for the generation of transition potentials in land use change modeling. Transactions in GIS 14 (5), 569-580.

Singh, V. G., Singh, S. K., Kumar, N., Singh, R. P., 2022. Simulation of land use/land cover change at a basin scale using satellite data and markov chain model. Geocarto International 37 (26), 11339-11364.

Sohl, T.L., Wimberly, M.C., Radeloff, V.C., Theobald, D.M., Sleeter, B.M., 2016. Divergent projections of future land use in the United States arising from different models and scenarios. Ecological Modelling 337, 281-297.