



National Authority for Remote Sensing and Space Sciences
The Egyptian Journal of Remote Sensing and Space Sciences

www.elsevier.com/locate/ejrs
www.sciencedirect.com



RESEARCH PAPER

Statistical independence test and validation of CA Markov land use land cover (LULC) prediction results



Md. Surabuddin Mondal^a, Nayan Sharma^b, P.K. Garg^c, Martin Kappas^a

^a Dept. of Cartography, GIS & Remote Sensing, Institute of Geography, Georg-August University of Göttingen, Göttingen, 37077, Germany

^b Dept. of W R D & M, Indian Institute of Technology, Roorkee, 247667, India

^c Dept. of Civil (Geomatic) Engineering, Indian Institute of Technology, Roorkee 247667, India

Received 19 August 2015; revised 20 July 2016; accepted 1 August 2016

Available online 7 September 2016

KEYWORDS

LULC;
CA Markov model;
Predictions;
Statistical independence test;
Validation;
Kappa index

Abstract Statistical independence test and validity of the CA (Cellular Automata) Markov process for projecting future land use and land cover (LULC) changes were carried out in this study. Predicting quantity and location changes have been analyzed, and statistically evaluated. Validity of the CA Markov process has been examined using various Kappa Index of Agreement (KIA or Kstandard) and related statistical variations on the KIA. Statistical test of independence (K^2) was performed and markovian suitability has been checked using hypothesis of goodness of fit (χ^2). Hypothesis of statistical independence was rejected, which proved that land use land cover change trends are similar like previous development of land. With acceptance of the hypothesis of goodness of fit (χ^2) proved that actual transition probability of matrix is fitted with expected transition probability prepared using Markov chain method. Statistics indicates Kno, Klocation, Klocation Strata and Kstandard are 0.8347, 0.859, 0.8591 and 0.7928, respectively.

© 2016 National Authority for Remote Sensing and Space Sciences. Production and hosting by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

1. Introduction

Land use/land cover changes (LULCC) are continuous process and have to be understood from more dynamics information.

Corresponding author.

E-mail addresses: msk.iit@gmail.com (M.S. Mondal), nayanfwt@iitr.ac.in (N. Sharma), gargpfce@iitr.ac.in (P.K. Garg), mkappas@uni-goettingen.de (M. Kappas).

Peer review under responsibility of National Authority for Remote Sensing and Space Sciences.

<http://dx.doi.org/10.1016/j.ejrs.2016.08.001>

1110-9823 © 2016 National Authority for Remote Sensing and Space Sciences. Production and hosting by Elsevier B.V.

This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

Traditionally change detection methods can only provide a static diagnosis of the land use/land cover change for the fixed beginning and end dates. Land use/land cover change process model aims at predicting the spatial distribution of the specific land cover and land use classes in a later year utilizing the knowledge gained from previous year. Modeling of land use/land cover change (LULCC) has been a topic of research since over a decade and there are several methods and models existing for the same. Baker (1989), followed by Lambin (1994) reviewed some initial LULCC models. Agarwal et al. (2002) provide details of LULCC models and recently, Mondal

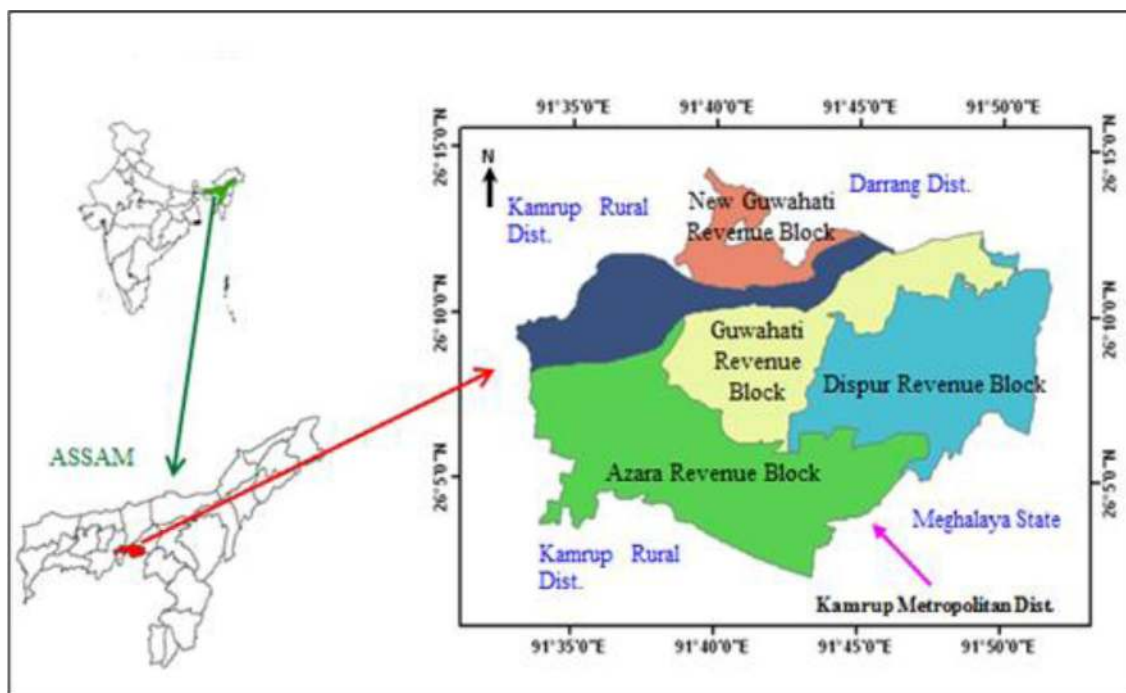


Figure 1a Location map.

et al. (2012) provides an updated detail of LULCC models. The Markov model is very good and useful to understand the stochastic nature and the stability of the land use/land cover (LULC). The Markov model has become more popular due to advancement of Remote Sensing and GIS technology. The Markov model is frequently used to simulate landscape change (Baker, 1989; Muller and Middleton, 1994), analyze land use types, trends and dimension of changes (Weng, 2002; Huang et al., 2008). Two representative models are the Markov chain model (Muller and Middleton, 1994) and the CA (Cellular Automata) Markov model (Clarke, 1997). The Markov chain model treats as a stochastic process; the later state (land cover type) of a pixel is only related to its immediate preceding state, but not to any other previous states. A transition probability is the direct outcome from the Markov chain model. The CA Markov model, on the other hand, achieved a significant improvement in incorporating the spatial contingency information when making predictions. As a step forward, research has been made with the Markov chain model to achieve better accuracy. Pontius and Malanson (2005) reported their success in applying spatial contiguity in a combined CA Markov model when predicting land cover changes in Central Massachusetts.

The CA Markov model combines both the concept of a CA filter and Markov chain procedure. Markov chain and CA both is the discrete dynamic model in time and state. The transition probabilities may be accurate on per category basis, but there is no knowledge of the spatial distribution of occurrences within each LULC category. CA will add spatial character to the model. CA is discrete dynamic systems in which the state of each cell at time $t + 1$ is determined by the stated of its neighboring cells at time according the pre-defined transition rules. CA as a method with temporal–spatial dynamics can simulate the evolution of things in two dimensions. Using the outputs

from the Markov chain analysis, the transition matrix, CA Markov will apply a contiguity filter to ‘grow out’ LULC from the time two to later time periods. CA Markov will use the transition areas tables and the conditional probability images to predict land use and land cover changes over the periods specified in Markov chain analysis. CA Markov will produce much better results geographically using the contiguity filter; those areas likely to change will do so closer to the existing LULC classes.

It is also important to validate the model output in an intelligent manner because a negative interpretation of the accuracy can give extremely misleading results. There are various methods of estimating the accuracy of prediction. Pontius et al. (2003) suggested use of kappa for location statistics in order to estimate the pixel level accuracy of a model as it extrapolates backwards in time for several land categories. Several studies estimated the accuracy using kappa for location statistics (et al.). Statistical test of independence (K^2) can be also used to understand whether the changes in LULC are dependent or not. The Markovian suitability can be checked using the hypothesis of goodness of fit (Xc^2) which is availed to test that the land use/land cover change trends are dependent or not dependent on previous development of land. Using the hypothesis of goodness of fit (Xc^2) it will check that actual transition probability of matrix of land use/land cover is fitted or not fitted with expected transition probability prepared using Markov chain method.

The prediction results in this study are tested and validated using traditional kappa for location statistics. Statistical test of independence (K^2) was also performed, the Markovian suitability has been checked using hypothesis of goodness of fit (Xc^2) and tested that the land use/land cover change trends are dependent or not depended on previous development of land. Using the hypothesis of goodness of fit (Xc^2) it has been

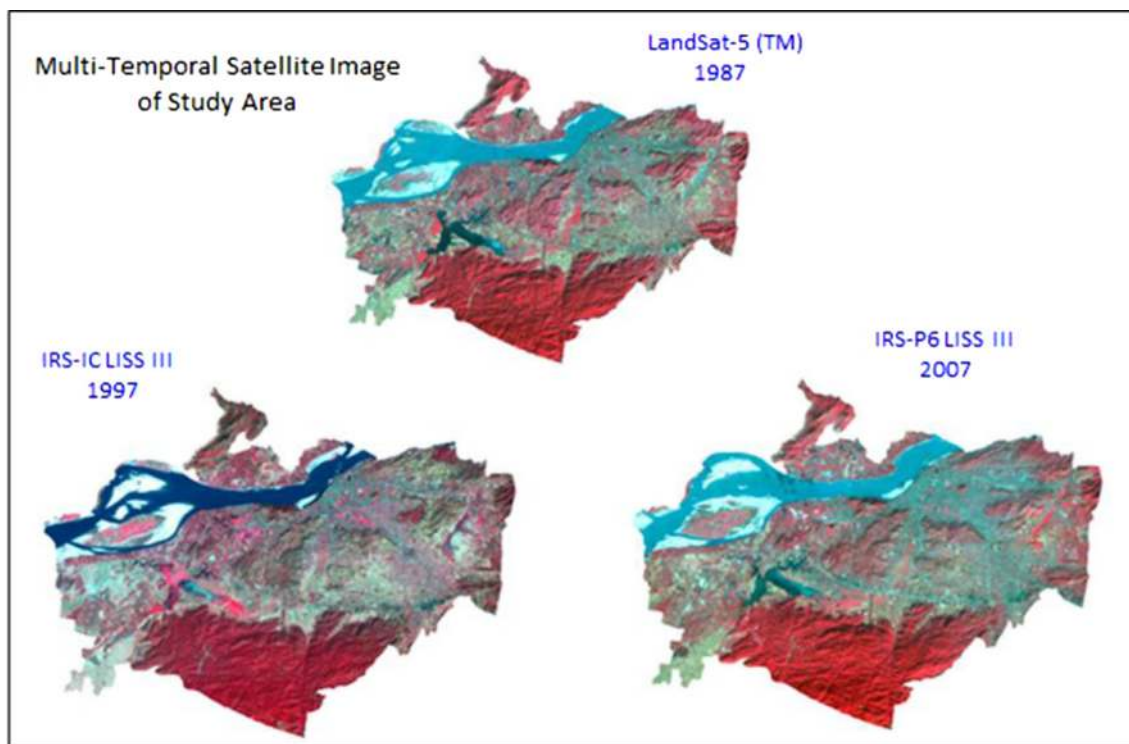


Figure 1b Satellite images of study area.

Table 1 Details of satellite data used in the study.

Satellite	Sensor	Path/row	Data acquired	Spatial resolution (m)	Spectral band	Data sources
LANDSAT-5	TM	136/042 (WRS-2 footprints)	26-12-1987	30 (120 m – thermal (B 6))	B 1 (blue): 0.45–0.52 m B 2 (green): 0.52–0.60 m B 3 (red): 0.63–0.69 m B 4 (NIR): 0.76–0.90 m B 5 (SWIR): 1.55–1.75 m B 6 (thermal IR): 10.4–12.5 m B 7 (Mid-Infrared): 2.08–2.35 m	GLCF-Earth Science Data Interface
IRS-1C	LISS-III	110/53	05-03-1997	23.5 (70 m – B5 (SWIR))	B 2 (green): 0.52–0.59 m B 3 (red): 0.62–0.68 m B 4 (NIR): 0.77–0.86 m B 5 (SWIR): 1.55–1.70 m	NRSC
IRS-P6 (Resourcesat-1)	LISS-III	110/53	14-12-2007	23.5	B 2 (green): 0.52–0.59 m B 3 (red): 0.62–0.68 m B 4 (NIR): 0.77–0.86 m B 5 (SWIR): 1.55–1.70 m	NRSC

The *Global Land Cover Facility (GLCF)* is a NASA-funded member of the Earth Science Information Partnership at the University of Maryland, providing free satellite images to users all over world.

also checked that actual transition probability of matrix of land use/land cover is fitted or not fitted with expected transition probability prepared using Markov chain method.

2. Data used & methods for LULC prediction

In this study, the spatio-temporal CA (Cellular Automata) Markov model of landscape change using multi-temporal

satellite imagery has been used which enabled us to predict spatial pattern of future land use/land cover for the study area – Kamrup Metropolitan district of Assam state in India (Fig. 1a). For this purpose, land use/land cover maps of the study area have been extracted from multi temporal satellite images. LANDSAT-5 TM image acquired on December 26, 1987, IRS-1C LISS III image acquired on March 5, 1997, IRS-P6 LISS III image acquired on 14th December of 2007 digitally classified for land use/land cover mapping (Fig. 1b

Table 2 Levels and LULC (land use land cover) classes considered for classification.

Level I	Level II
1. Built up land	1.1. Built up land
2. Agricultural land	2.1. Agricultural crop land
	2.2. Agricultural fallow land
	2.3. Plantations
3. Forest	3.1. Dense forest
	3.2. Degraded forest
4. Waste land	4.1. Land with or without scrub
	4.2. Marshy/swampy
	4.3. Waterlogged area
	4.4. Sandy area (river bed)
5. Water bodies	5.1. River/stream
	5.2. Lake/reservoir/pond/tank
6. Others	6.1. Open land
	6.2. Aquatic vegetation

and Table 1). Land use/land cover (LULC) maps derived from satellite images of 1987 and 1997 were used to predict future land use/land cover of 2007. The CA Markov model is simu-

lated for a especial study area which covered a large proportion by urban landscape with or surrounding by other 14 classes of LULC. The CA model, coupled with the Markov transition probability, has indicated the capability of trend projection for landscape change. This spatio-temporal model provided not only the quantitative description of change in the past but also the direction and magnitude of change in the future.

2.1. Preparation of LULC maps

The image dataset used in this study consists of LANDSAT-5 TM images of December 1987, IRS-1C images of March 1997 and IRS-P6 images of December 2007. Only images acquired in December and March months (winter season) were considered. The available images were selected based on the absence of cloud cover. When multi-data images from different sources are used, different atmospheric and terrain conditions may cause variations in data. Therefore, radiometric corrections including atmospheric correction - Top-of-Atmosphere (TOA) reflectance calibration were applied in this study. After radiometric correction, geometric correction was applied to the

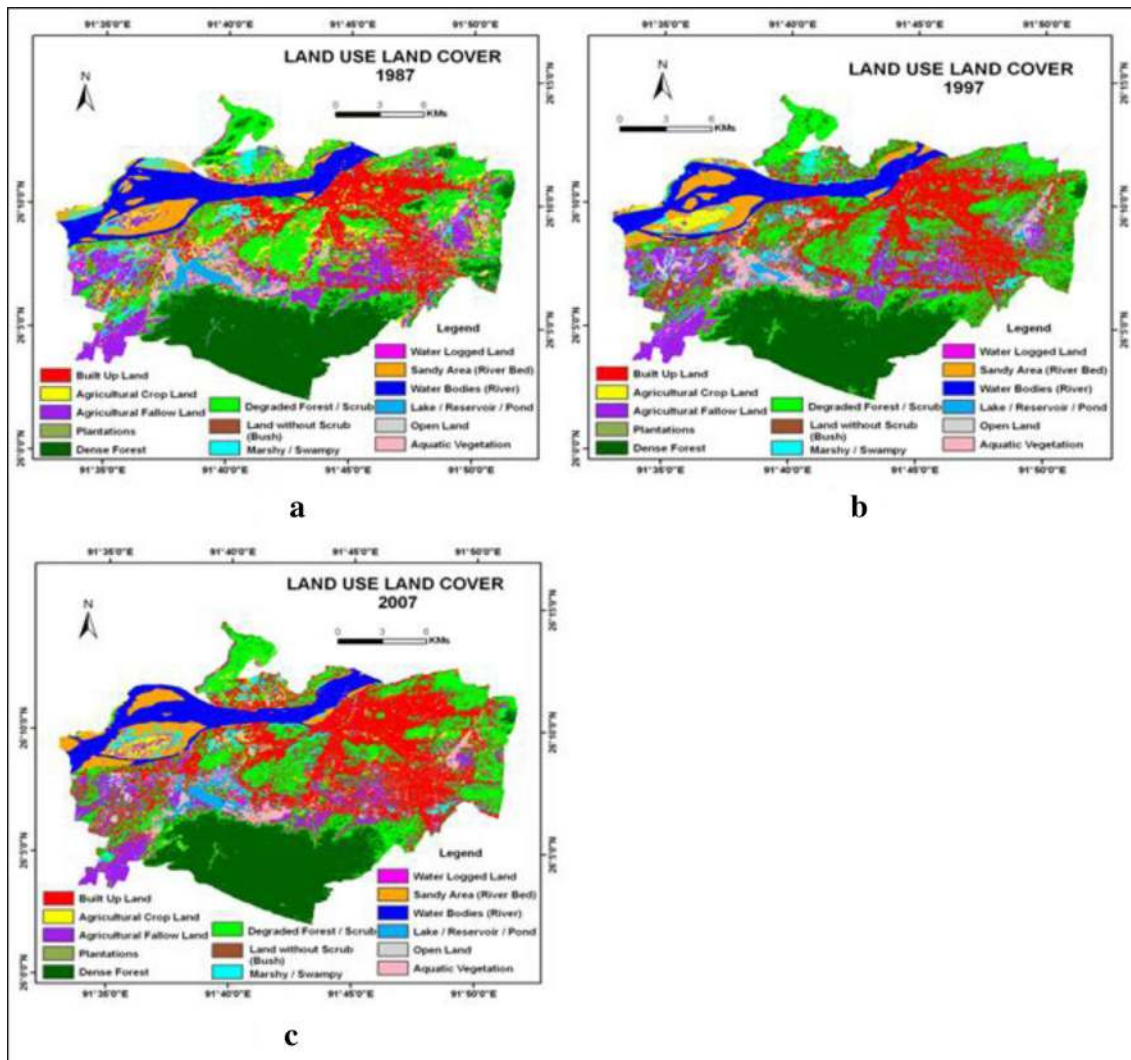


Figure 2 Classified land use land cover map of 1987, 1997 and 2007.

images. For accurate change detection, an accurate geometric registration is needed. The 1987 Landsat image from Global Land Cover Facility (GLCF) was chosen which has been orthorectified by the United States Geological Survey (USGS). Then, IRS-1C images of 1997 and IRS-P6 images of 2007 were rectified (geometrically corrected) with reference to the orthorectified Landsat satellite image of 1987 with two-order polynomial transformation and more than 14 ground control points (GCPs—mainly road junctions) to further improve the georeferencing accuracy. All images were resampled using Nearest Neighbor resampling method with a root mean square error of less than ±0.5 pixels per image to a 23.5 m resolution with the UTM coordinate system (zone 46, WGS 84 datum system).

For this study, supervised maximum likelihood classifier is used to classify all satellite images. Modified (modified from NRSA classification system for India and classification scheme adopted for European Commission sponsored Brahmatwin projects) classification scheme (level II) is adopted for different categories of LULC (Table 2). 14 LULC classes i.e., built up land, agricultural crop land, agricultural fallow land, plantation, dense forest land, degraded forest land, land with or without scrub, marshy/swampy land, waterlogged area, sandy area, river, lakes/reservoirs/ponds, open land, aquatic vegetation area derived from satellite images. As supervised classification technique has been used for this study, it requires a priori knowledge of the number of classes, as well as knowledge concerning statistical aspects of the classes. Areas of visually homogeneous spectral response were chosen (10–12 training set for per class) well distributed all over images as AOI (area of interest) and added to the spectral signature editor. Limited pre-classification ground truth (using GPS) helped to select the training samples. The pre-classification ground truth was conducted on 14 December 2007, the same date when satellite collected the images for the study area. In the classification, the signature separability functions were used to examine the quality of training sites and class signature, before performing the classification. The land use and land cover types derived from digital image classification validate with data obtained from limited post-classification ground verification and using high-

resolution Google earth images. Land use/land cover (LULC) maps derived from satellite images of 1987, 1997 and 2007 are shown in Fig. 2 & area statistics are shown in Table 3.

2.2. CA Markov model

CA Markov model is a combination of the concept of a CA filter and Markov chain procedure. The CA model can be expressed as follows:

$$S(t, t + 1) = f(S(t), N)$$

where, S is the set of limited and discrete cellular states, N is the Cellular field, t and $t + 1$ indicate the different times, and f is the transformation rule of cellular states in local space.

The Markov model is a theory based on the process of the formation of Markov random process systems for the prediction and optimal control theory method. Based on the conditional probability formula—Bayes, the prediction of land use changes is calculated by the following equation:

$$S(t + 1) = P_{ij}eS(t) \tag{1}$$

where, $S(t)$, $S(t + 1)$ are the system status at the time of t or $t + 1$; P_{ij} is the transition probability matrix in a state which is calculated as follows:

$$P = (p_{ij}) = \begin{pmatrix} P_{11} & P_{12} & P_{1n} \\ P_{21} & P_{22} & P_{2n} \\ P_{n1} & P_{n2} & P_{nn} \end{pmatrix}, \quad \sum_{j=i}^n p_{ij} = 1 \tag{2}$$

where, P is the Markov transition matrix P ,
 i, j is the land use land cover type of the first and second time period,
 and P_{ij} is the probability from land use and land cover type i to land type j .

In this expression, n is the number of land use and land cover types in the target area, and “ P_{ij} ” is the probability of transition of type i into that of type j from the initiation to the end. In the transition matrix, it requests that each rate is

Table 3 Area statistics of LULC (land use land cover).

Sl. No.	Class name	1987		1997		2007	
		Area (km ²)	% of area	Area (km ²)	% of area	Area (km ²)	% of area
1.	Built up land	60.54	14.63	102.4	24.73	141.35	34.14
2.	Agricultural crop land	25.91	6.26	5.99	1.45	7.17	1.73
3.	Agricultural fallow land	48.27	11.66	34.08	8.23	25.12	6.07
4.	Plantations	1.38	0.33	3.68	0.89	3.35	0.81
5.	Dense forest	86.26	20.84	80.56	19.46	74.84	18.08
6.	Degraded forest	83.48	20.17	76.95	18.59	60.31	14.57
7.	Land with or without scrub	9.48	2.29	24.82	6	23.78	5.74
8.	Marshy/swampy	13.42	3.24	10.26	2.48	6.82	1.65
9.	Water logged area	3.57	0.86	1.86	0.45	1.52	0.37
10.	Sandy area (river bed)	14.83	3.58	16.08	3.88	15.92	3.85
11.	River/stream	37.27	9	32.51	7.85	33.42	8.07
12.	Lake/reservoir/pond/tank	7.99	1.93	6.05	1.46	6.59	1.59
13.	Open land	13.8	3.33	7.28	1.76	6.97	1.68
14.	Aquatic vegetation	7.78	1.88	11.46	2.77	6.82	1.65
Total		413.98	100.00	413.98	100.00	413.98	100.00

Table 4 Transition matrix of 1987–1997.

LULC classes	Built up land	Agricultural crop land	Agricultural fallow land	Plantations	Dense forest	Degraded forest	Land with or without scrub	Marshy/swampy	Water logged area	Sandy area (river bed)	River/stream	Lake/reservoir / pond/tank	Open land	Aquatic vegetation
Built up land	0.0001	0.083	0.0101	0.0133	0.0002	0.0003	0.0212	0.0038	0.0028	0.0006	0.0007	0.0005	0.0005	0.0057
Agricultural crop land	0	0	0.0027	0.0023	0	0	0.0004	0	0.0009	0	0.0019	0	0	0.0001
Agricultural fallow land	0.0001	0	0.0041	0.0302	0	0	0.004	0.0011	0.0021	0.0008	0.0008	0	0.0001	0.0035
Plantations	0	0.0007	0.0005	0.0002	0.0009	0.0001	0.002	0.0001	0.0003	0.0001	0	0	0	0.0001
Dense forest	0.0001	0	0.0004	0.0003	0.0002	0.1025	0.0088	0.0002	0.0002	0	0	0	0.0001	0.0001
Degraded forest	0.0002	0.0004	0.0096	0.0075	0.0004	0.0163	0.061	0.0032	0.0028	0.0005	0.0003	0.0003	0.0007	0.0038
Land with or without scrub	0	0	0.0057	0.0054	0.0001	0.0011	0.0134	0.0035	0.0006	0.0002	0.0001	0	0.0003	0.0028
Marshy/swampy	0	0.0002	0.0012	0.0023	0	0.0001	0.0026	0.0003	0.003	0.0001	0.0026	0.0007	0.0001	0.0007
Water logged area	0	0	0.0001	0.0006	0	0	0.0004	0	0.0007	0.0006	0	0	0	0.0001
Sandy area (river bed)	0.0001	0	0.0002	0.0001	0	0	0.0001	0	0.0006	0.0001	0.01	0.0113	0	0
River/stream	0.0001	0	0.0001	0	0	0	0	0	0.0017	0.0001	0.0041	0.0394	0	0
Lake/reservoir/pond/tank	0	0.0001	0.0002	0.0004	0.0001	0	0.0006	0.0001	0.0012	0.0005	0.0001	0	0.0037	0.0002
Open land	0	0.0003	0.0007	0.0041	0	0	0.0009	0.0003	0.0009	0.0007	0.0001	0	0.0002	0.0017
Aquatic vegetation	0	0.0001	0.0006	0.0007	0.0001	0	0.0014	0.0005	0.001	0.0007	0	0	0.0054	0.0006

Table 5 Transition matrix of 1997–2007.

LULC classes	Built up land	Agricultural crop land	Agricultural fallow land	Plantations	Dense forest	Degraded forest	Land with or without scrub	Marshy/swampy	Water logged area	Sandy area (river bed)	River/stream	Lake/reservoir/pond/tank	Open land	Aquatic vegetation
Built up land	0.0001	0.1393	0.0008	0.0114	0.0022	0.0014	0.0249	0.0082	0.0029	0.0004	0.0005	0.0003	0.0009	0.0031
Agricultural crop land	0	0.0001	0.0031	0.001	0.0002	0.0003	0.0023	0.001	0.0008	0	0.0006	0	0.0001	0.0001
Agricultural fallow land	0.0001	0.0002	0.002	0.0202	0.0001	0.0001	0.0033	0.003	0.0021	0.0002	0.0005	0	0.0003	0.0025
Plantations	0	0.0001	0.0001	0.0003	0.0012	0.0006	0.0013	0.0006	0	0	0	0	0.0001	0
Dense forest	0.0001	0	0	0	0	0.0984	0.006	0.0003	0	0	0	0	0	0
Degraded forest	0.0001	0.0008	0.0002	0.0039	0.0007	0.0104	0.0526	0.0117	0.0016	0.0005	0	0	0.0003	0.0006
Land with or without scrub	0.0001	0.0004	0.0003	0.0062	0.0004	0.0009	0.0112	0.0072	0.0015	0.0007	0.0001	0	0.0009	0.0016
Marshy/swampy	0	0.0002	0.0003	0.0005	0.0001	0.0002	0.0023	0.0004	0.0019	0.0003	0.0016	0.0008	0.0004	0.0003
Water logged area	0	0	0	0.0002	0	0	0.0005	0.0003	0.0001	0.0001	0.0002	0.0002	0.0001	0.0002
Sandy area (river bed)	0.0001	0	0.0007	0.0003	0	0	0.0001	0	0.0016	0	0.0109	0.0084	0.0001	0.0001
River/stream	0	0.0002	0.0008	0.0003	0	0	0.0003	0.0001	0.0011	0	0.0082	0.0358	0	0
Lake/reservoir/pond/tank	0	0.0001	0	0.0002	0	0	0.0005	0.0002	0.0001	0.0001	0	0	0.0037	0.0002
Open land	0	0.0018	0.0001	0.0026	0.0001	0.0001	0.0014	0.0009	0.0005	0.0001	0	0	0.0002	0.0015
Aquatic vegetation	0	0.0001	0	0.0005	0	0.0001	0.0009	0.0009	0.0002	0.0002	0	0	0.0014	0.0002

Table 6 Transition probability of prepared LULC data for 1987–2007.

LULC classes	Built up land	Agricultural crop land	Agricultural fallow land	Plantations	Dense forest	Degraded forest	Land with or without scrub	Marshy/swampy	Water logged area	Sandy area (river bed)	River/stream	Lake/reservoir/pond/tank	Open land	Aquatic vegetation
Built up land	0.4190	0.0001	0.0001	0.0002	0	0.0004	0.0003	0	0	0	0	0	0	0
Agricultural crop land	0.0002	0.0821	0.0178	0.0264	0.0006	0.0011	0.0415	0.0066	0.0052	0.0013	0.0009	0.0009	0.0011	0.0105
Agricultural fallow land	0	0	0.0024	0.0016	0	0.0004	0.0021	0.0001	0.0009	0.0002	0.0017	0	0.0001	0.0002
Plantations	0.0001	0.0001	0.0026	0.0223	0	0.0001	0.0033	0.0009	0.0011	0.0005	0.001	0	0.0001	0.002
Dense forest	0	0.0001	0.0005	0.0003	0.0009	0.0005	0.0016	0.0002	0.0002	0.0001	0	0	0.0001	0.0001
Degraded forest	0.0001	0	0	0.0001	0.0001	0.1005	0.0039	0	0.0001	0	0	0	0	0
Land with or without scrub	0.0002	0.0005	0.0076	0.0048	0.0002	0.0157	0.0481	0.0019	0.0018	0.0003	0.0001	0	0.0004	0.0023
Marshy/swampy	0.0001	0.0003	0.0031	0.007	0	0.0012	0.0114	0.0022	0.0024	0.0016	0.0002	0	0.0005	0.0021
Water logged area	0	0.0002	0.0004	0.001	0	0.0006	0.0018	0.0002	0.0019	0.0001	0.0017	0.001	0.0001	0.0001
Sandy area (river bed)	0	0	0.0001	0.0002	0	0.0002	0.0003	0.0003	0.0002	0.0001	0.0002	0.0002	0.0001	0.0001
River/stream	0.0001	0	0.0003	0.0001	0	0	0	0	0.0016	0.0001	0.0105	0.0094	0	0
Lake/reservoir/pond/tank	0.0001	0	0.0001	0	0	0	0	0	0.0016	0	0.0043	0.0406	0	0
Open land	0	0	0.0001	0.0003	0.0001	0	0.0004	0.0001	0.0002	0.0001	0	0	0.006	0.0001
Aquatic vegetation	0	0.0014	0.001	0.0026	0	0.0001	0.0015	0.0003	0.0009	0.0001	0.0001	0	0.0001	0.0013

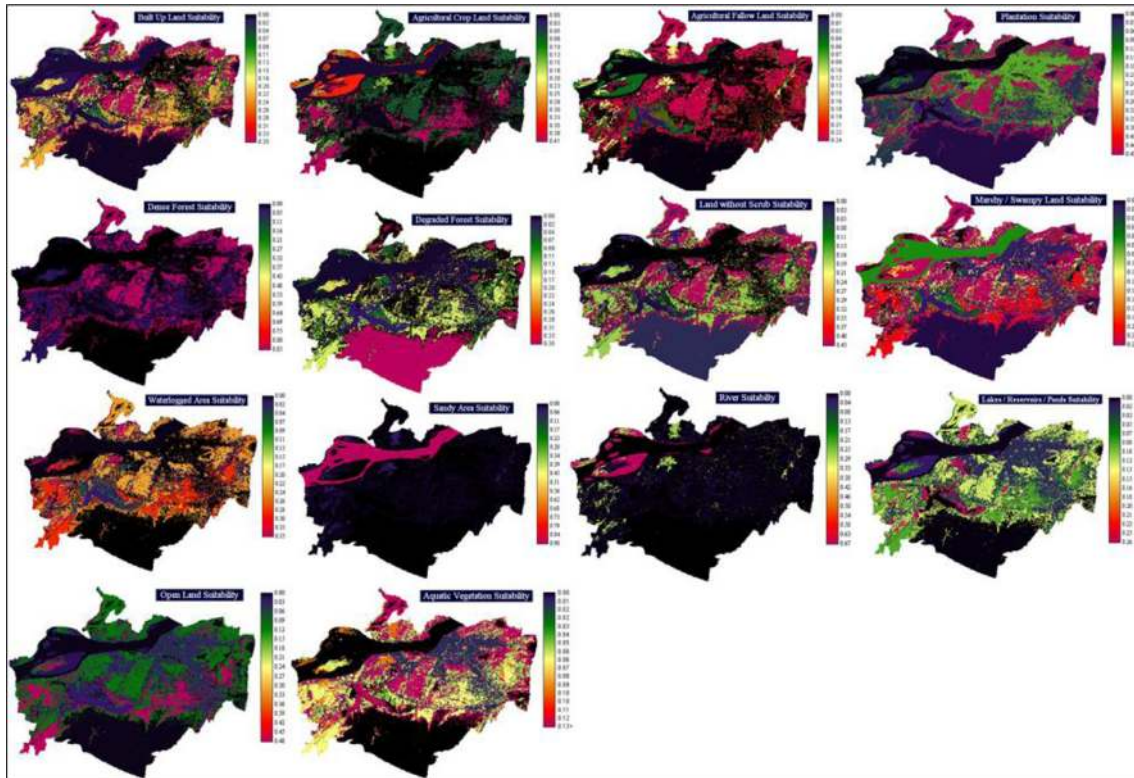


Figure 3 Suitability (evidence likelihood) map used to predict future LULC.

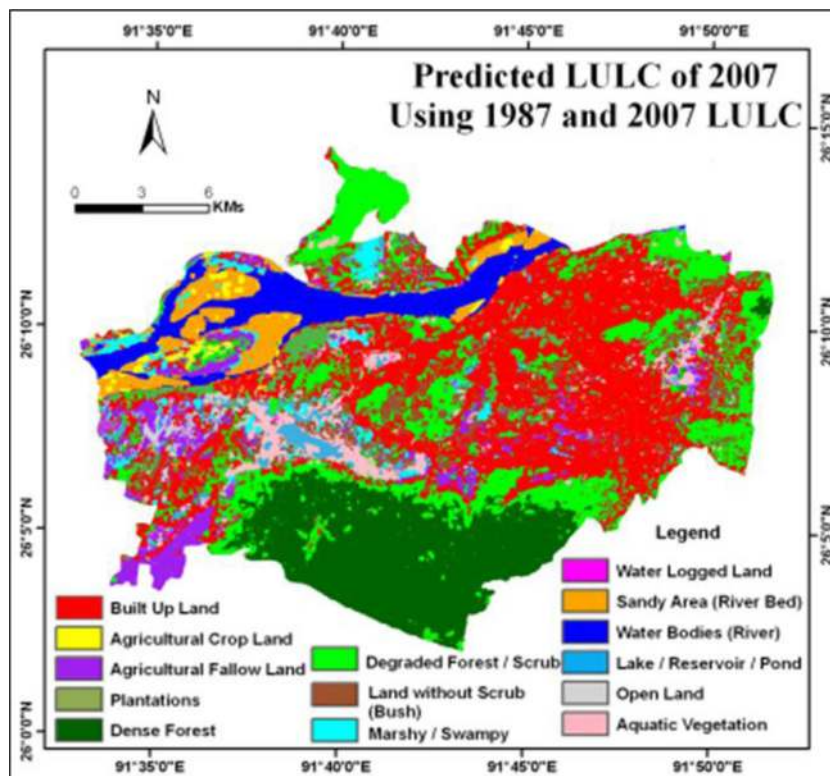


Figure 4 Predicted LULC of 2007 using 1987 & 1997 LULC image.

a non-negative quantity, and each line factor 0 to 1. The estimate of Markov chain is the relative frequency of transitions observed over the entire time period. The result of the estimation can be used for prediction.

2.2.1. Markov chain – transition probability matrix

The transition probability matrix has been calculated for the time period of 1987–1997 & 1997–2007 for the prediction of LULC of 2007. The transition probability matrix for the time period of 1987–1997 & 1997–2007 displayed in Tables 4 and 5, respectively. The expected probability of transition of LULC category is displayed in Table 6. The transition probability matrix is the cross tabulation of the two images (1987 and 1997 & 1997 and 2007), that each LULC category will change to every other category. The transition probability areas matrix records the number of pixels that are expected to change over the specified time (1987–2007).

2.2.2. Preparation of suitability map (evidence likelihood map) and calibration of the CA Markov model

According to the underlying land use and land cover change dynamics between years 1987 and 1997, a series of suitability maps (evidence likelihood map) consisting of built up land suitability, agricultural crop land suitability, agricultural fallow land suitability, plantation suitability, dense forest land suitability, degraded forest land suitability, land with or without scrub suitability, marshy/swampy land suitability, waterlogged area suitability, sandy area suitability, river suitability, lakes/reservoirs/ponds suitability, open land suitability, aquatic vegetation land suitability were prepared (Fig. 3). The number thus expresses the likelihood of finding the LULC at the pixel in question, if this lies in transition area. These images (evidence likelihood maps) are calculated as projections from the later date image (1997) of two input LULC images (before image 1987 and later image 1997). The output images are the conditional probability images. This conditional probability images report the probability that each LULC type would be found at each pixel in future after the

specified time. The procedure looks at the relative frequency of pixels belonging to the different categories of that variable within areas of change. In effect, it asks the question of each category of the variable, “How likely is it that you would have a value like this if you was an area that would experience change?” (Eastman et al., 2009). To project land use and land cover change for next 10 years using known LULC of 1987 and 1997, probability statistics for land use and land cover change for 2007 has been generated through cross tabulation of two LULC maps. Thus, the CA Markov model combines both the concepts of Markov chain procedure and CA filters, after getting Markov transition probability, CA Markov used the transition probability matrix and probability images (here, suitability/evidence likelihood map) to predict the LULC over a 10 years period i.e., 2007. The total numbers of iterations are based on the number of time steps, for 10 years model will choose to complete run in 10 iterations. The predicted locations of LULC are shows in Fig. 4. The quantitative results of predicted LULC are shows in Table 7.

3. Statistical independence test for Markov chain transition probability

The Markov model considers that LULC as stochastic process, and different categories of LULC as the states of chain. A chain is defined as stochastic process having the conditional probability distribution of the process at time $n + 1$, X_{n+1} depends upon only value of X_n , and is not dependent on all other previous value $X_{n-1}, X_{n-2}, \dots, X_0$. It can be explained as:

$$P[X_{n+1} = X_{n+1} | X_n = X_n] \quad X_0 = x_0$$

$$P[X_{n+1} = x_{n+1} | X_n = x_n] \tag{1.1}$$

This can also be expressed as

$$P_{ij} = P[X_{n+1} = j | X_n = i] \tag{1.2}$$

$$ij = 0, 1, 2,$$

Table 7 Area statistics of predicted land use land cover (LULC) of 2007 using 1987 & 1997 LULC (land use land cover) image and LULC (land use land cover) derived from LISS III image of 2007.

LULC class	Area (in Km ²)		
	Predicted LULC 2007 (Using 1987 & 1997 LULC Image)	LULC 2007 (Derived from LISS III Image of 2007)	Differences
Built up land	125.09	141.35	16.26
Agricultural crop land	4.32	7.17	2.85
Agricultural fallow land	23.62	25.12	1.50
Plantation	10.57	3.35	+ 7.22
Dense forest	66.26	74.84	8.58
Degraded forest	76.19	60.31	+ 15.88
Land with or without scrub	24.95	23.78	+ 1.17
Marshy/swampy	10.91	6.82	+ 4.09
Waterlogged	1.46	1.52	0.06
Sandy area	17.39	15.92	+ 1.47
River	25.72	33.42	7.70
Lakes/reservoirs/ponds	6.31	6.59	0.28
Open land	8.67	6.97	+ 1.70
Aquatic vegetation	12.52	6.82	+ 5.70
Total	413.98	413.98	

Table 8 Transition probability of LULC from 1987 to 2007 under Markov Hypothesis.

LULC classes	Built up land	Agricultural crop land	Agricultural fallow land	Plantations	Dense forest	Degraded forest	Land with or without scrub	Marshy/swampy	Water logged area	Sandy area (river bed)	River/stream	Lake/reservoir/pond/tank	Open land	Aquatic vegetation
Built up land	0.9799	0	0	0.0078	0	0.0047	0	0.0019	0	0	0.0003	0.0012	0.0037	0.0014
Agricultural crop land	0.2812	0.0736	0.1141	0.0149	0.0098	0.2637	0.1569	0.0342	0.0016	0.006	0.0033	0.005	0.019	0.0172
Agricultural fallow land	0.1998	0.0343	0.4459	0.0031	0.0045	0.1109	0.0801	0.0343	0.0084	0.0016	0.0007	0.0061	0.0602	0.0103
Plantations	0.0821	0.0056	0.0111	0.4736	0.0984	0.1829	0.044	0.002	0.0004	0	0	0.0595	0.0052	0.0353
Dense forest	0.0055	0	0.0002	0.0004	0.8486	0.1349	0.0093	0.0005	0	0	0.0001	0.0001	0.0003	0
Degraded forest	0.1827	0.0032	0.0343	0.017	0.0751	0.5218	0.1143	0.0221	0.0038	0.0005	0.0001	0.0049	0.0078	0.0121
Land with or without scrub	0.2883	0.0015	0.0837	0.0112	0.0137	0.243	0.2666	0.0215	0.0019	0.0001	0	0.0054	0.0227	0.0406
Marshy/swampy	0.1491	0.0472	0.1102	0.0135	0.0083	0.1487	0.0336	0.1599	0.0369	0.0316	0.0904	0.066	0.0499	0.0546
Water logged area	0.1301	0.0008	0.1624	0.0159	0.0017	0.0946	0.036	0.0259	0.1196	0.0259	0.0102	0.1072	0.1379	0.132
Sandy area (river bed)	0.0334	0.0928	0.039	0.002	0.0001	0.0132	0.0063	0.1235	0	0.4823	0.1981	0.0026	0.0067	0
River/stream	0.0092	0.0009	0.0003	0.0003	0	0.005	0	0.0137	0.0001	0.2156	0.7537	0.0002	0.0009	0
Lake/reservoir/pond/tank	0.0442	0.0018	0.0133	0.0026	0.0067	0.0585	0.0287	0.0097	0.0028	0.0008	0.0014	0.3293	0.0158	0.4845
Open land	0.2967	0.0032	0.1786	0.0056	0.0067	0.1966	0.1448	0.0369	0.0032	0.0023	0.0007	0.0102	0.0861	0.0293
Aquatic vegetation	0.0639	0	0.0774	0.0077	0.005	0.0872	0.1318	0.0400	0.0091	0.0001	0	0.1151	0.0155	0.4473

Here, P_{ij} is transition probability of one step, which can be analyzed as the conditional probability at time n when the process is in state i and at time $n + 1$ the process is in state j . Two step transition probabilities are defined with generalization of Chapman–Kolmogorov equation.

$$P_{ij}^2 = P[X_{n+2} = j | X_n = i] = P[X_{n+2} = j | X_{n+1} = k]P[X_{n+1} = k | X_n = i] \quad (1.3)$$

$$\text{This is equivalent to } (P)m + n = (P)n(P)m \quad (1.4)$$

3.1. Hypothesis test for statistical independence

To follow the hypothesis of statistical independence involves a process of comparing the actual data with expected data of land use adopting following formula:

$$K^2 = (A_{ik} - E_{ik})^2 / E_{ik} \quad (1.5)$$

where,

E_{ik} = expected value under Markov hypothesis

A_{ik} = actual value of data from category in i to category in k .

If the value of K^2 is greater than the tabulated value on the critical region 0.05 with degree of freedom (D.F. 1)² the hypothesis will be rejected. Expected value calculated with the use of Chapman–Kolmogorov equation following the Markov method. For calculation of transition probability matrix for the period 1987–2007 (Table 8) can be obtained by multiplying the 1987–1997 matrices (Table 4) and 1997–2007 matrices (Table 5). The expected value is calculated by following formula:

$$E_{ik} = E(E_{ij})(E_{jk})/E_j \quad (1.6)$$

where,

E_{ij} = the number of transition from category i to j during the period 1987–1997,

E_{jk} = the number of transition from category j to k during the period 1997–2007,

E_j = the number of cells in category j in 1987.

3.2. Test of goodness of fit

Chi square test of goodness of fit is used to test order Markovian suitability with the data. This test analyzes whether the particular distribution is adequately described or not. By making comparison between actual observed probability and expected probability.

$$Xc^2 = \sum \sum (O_{ik} - E_{ik})^2 / E_{ik} \quad (1.7)$$

where,

O_{ik} = observed transition probability data from 1987 to 1997,

E_{ik} = expected data of transition probability from 1987 to 2007.

If the Xc^2 is less than the value of $X_{1-\alpha}$ on the 0.05 critical regions then the hypothesis is accepted.

3.3. Output of statistical independence test

The transition probability matrix has been calculated between 1987–1997 & 1997–2007 for prediction of LULC for 2007. The expected probability of transition of LULC category is displayed in Table 8. The transition probability matrix is the cross tabulation of the two images (images of 1987 and 1997).

The statistical test (Table 9) of independence is used to understand whether the changes in LULC are dependent or not. For this statistical test of independence, (K^2) is performed on LULC data. The results of K^2 is 497.12 which is more than the significance 201.1 on critical region 0.05 with degree of freedom (14 1)². So the hypothesis of statistical independence is rejected. Therefore, the changes in LULC are dependent. One can say that the land use and land cover change trends are dependent on previous development of land use/land cover or in another language one can say the land use/land cover change trends are likely to similar kinds of previous trends of land use/land cover change.

The Markovian suitability has been checked using hypothesis of goodness of fit. In this test, actual LULC from 1987 to 2007 has been compared with expected data (LULC) which were calculated using the Markov model. This hypothesis is accepted for these data. The calculated value of Xc^2 is 0.52 and it is very much less than the significance of 22.4 on the critical region of 0.05 with 13 degrees of freedom (Table 9). With acceptance of the hypothesis one can say that actual transition probability of matrix from 1987 to 2007 is fitted with expected transition probability prepared using Markov method. Actual transition probability of matrix from 1987 to 2007 is similar to expected transition probability prepared using Markov method.

4. Validation of CA Markov prediction – kappa indices of agreement and disagreement

The international scientific community has called for research into land cover change, specifically models that predict spatial patterns of future change (Turner et al., 1995; Lambin et al., 2003). Modelers are satisfying this need with a variety of approaches (Baker, 1989; Pontius et al., 2004; Hall et al., 1995; Veldkamp and Fresco, 1996; Geoghegan et al., 1997; Mertens and Lambin, 1997; Liverman et al., 1998; Wu and Webster, 1998). In most cases, the models are connected to a raster-based GIS. Scientists are required to necessarily develop statistical methods to validate such models, because it is essential to know its prediction accuracy (Pontius and Schneider, 2001). Pontius (2002) have suggested the use of Kappa statistics for testing accuracy in terms of location (Kappa for location) and quantity of correct cells (Kappa for quantity). Therefore, land use and land cover change data derived from

Table 9 Statistical results of data.

Test perform	Calculated value	Chi sq. table value on .05 critical region
Statistical independence test (K^2)	497.12	201.1
Goodness of fit test (Xc^2)	0.52	22.4

Table 10 Agreement/disagreement according to ability to specify accurately quantity and location to predict 2007 LULC.

Sl. No.	Information of location	Information of quality		
		No [n]	Medium [m]	Perfect [p]
1.	Perfect [P(x)]	P(n) = 0.4592	P(m) = 0.9478	P(p) = 1.0000
2.	Perfect Stratum [K(x)]	K(n) = 0.4592	K(m) = 0.9478	K(p) = 1.0000
3.	Medium Grid [M(x)]	M(n) = 0.4398	M(m) = 0.8550	M(p) = 0.8856
4.	Medium Stratum [H(x)]	H(n) = 0.1522	H(m) = 0.3235	H(p) = 0.3261
5.	No [N(x)]	N(n) = 0.1522	N(m) = 0.3235	N(p) = 0.3261
Agreement chance				0.1522
Agreement quantity				0.1713
Agreement strata				0.0000
Agreement grid cell				0.5315
Disagree grid cell				0.0928
Disagree strata				0.0000
Disagree quantity				0.0522

Table 11 Kappa Index of Agreement to ability to specify accurately quantity and location to predict 2007 LULC.

Statistics	Index
Kno	0.8347
Klocation	0.8591
Klocation Strata	0.8591
Kstandard	0.7928

satellite images for describing and projecting land use and cover changes establishes the validity of the predicted results of the CA Markov process in this study. For validation, a map of simulated future change is compared to a map of recent real land cover change. For appropriate validation, the map of reality used for validation should not be used in calibration (Pontius and Schneider, 2001). Here, LULC of 2007 is predicted using LULC maps of 1987 and 1997, derived from Landsat and IRS-P6 satellite images, respectively. This provides a method to measure agreement between two categorical images, a “comparison” map (here the predicted LULC of 2007 – Fig. 4) and a “reference” map (LULC map derived from IRS-P6 LISS III image of 2007 – Fig. 2c). The comparison map is the result of CA Markov model simulation results, whose validity is to be assessed against a reference map that depicts reality.

The statistical methods separate error and agreement by components due to specification of quantity and location. The simulated map of 2007 is compared to the reference map of 2007, a Kappa for quantity and location statistic is derived (Table 10). The statistics for location showing Kno is 0.8347, Klocation is 0.859, Klocation Strata is 0.8591 and Kstandard is 0.7928 (Table 11). The results indicate that CA Markov model’s ability to specify grid cell level location of future change is nearly perfect (here Klocation value is 0.859, where Klocation value of 1 is perfect).

5. Conclusions

Currently, land-change modelers are not being held accountable for their prediction of future landscapes. Most land-change modelers fail to validate models and fail to state the

uncertainty in future prediction. Consequently, policy makers and the general public develop opinions based on misleading research that fails to give them the appropriate interpretations required to make informed decisions. Validation efforts to a known point in time are necessary to make an estimate of the uncertainty for the extrapolation to an unknown point in time. CA Markov LULCC Model prediction results were tested and validated in this study using traditional kappa for location statistics. Statistical test of independence (K^2) was performed; the Markovian suitability has been checked using hypothesis of goodness of fit (Xc^2) and proved that the land use/land cover change trends are dependent on previous development of land. The calculated value of Xc^2 is 0.52 and it is very less than significance 22.4 on critical region 0.05 with 13 degree of freedom. With acceptance of the hypothesis established that actual transition probability of matrix from 1987 to 2007 is fitted with expected transition probability prepared using Markov method. Hypothesis of goodness of fit (Xc^2) value established that the actual transition probability of matrix of land use/land cover is similar to expected transition probability prepared using the Markov chain method. The validation for CA Markov model land use/land cover prediction results calculated using various Kappa Index of Agreement (KIA or Kstandard) and related statistical variations on the KIA. The simulated map of 2007 was compared to the reference map of 2007, Kappa for quantity and location statistic was derived and statistics for location showing Kno is 0.8347, Klocation is 0.859, Klocation Strata is 0.8591 and Kstandard is 0.7928 and this results indicated that CA Markov model’s ability to specify grid cell level location of future change is nearly perfect. This study concludes that use of statistical independence test, Kappa indices are potentially useful techniques for purposes of validation of CA Markov model land use/land cover (LULC) prediction results.

References

Agarwal, C., Green, G.M., Grove, J.M., Evans, T.P., Schweik, C.M., 2002. A review and assessment of land-use change models: dynamics of space, time, and human choice, Gen. Tech. Rep. NE-297. U.S. Department of Agriculture, Forest Service, Northeastern Research Station, Newton Square, PA, p. 61.

- Baker, W.L., 1989. A review of models of landscape change. *Landscape Ecol.* 2, 111–133.
- Clarke, K.C., 1997. Land transition modeling with deltatrons. In: *Proceedings of the Land Use Modeling Workshop, June 1997, Santa Barbara, California*, Available online at: <http://www.ncgia.ucsb.edu> (accessed 1 June 2007).
- Eastman, J.R., 2009. IDRISI users guide. Clark University, Massachusetts.
- Geoghegan, J., Wainger, L.A., Bockstael, N.E., 1997. Spatial landscape indices in a hedonic framework. An ecological economics analysis using GIS. *Ecol. Econ.* 23, 251–264.
- Hall, D.K., Foster, J.L., Chien, J.Y.L., Riggs, G.A., 1995. Determination of actual snow covered area using Landsat TM and digital elevation model data in Glacier National Park, Montana. *Polar Rec.* 31 (177), 191–198.
- Huang, W., Liu, H., Luan, Q., Jiang, Q., Liu, J., Liu, H., 2008. Detection and prediction of land use change in Beijing based on remote sensing and GIS. *Int. Arch. Photogram. Rem. Sens. Spatial Inf. Sci.* XXXVII, 75–82.
- Lambin, E.F., 1994. Modelling deforestation processes: A review TREES Series B. Research Report 1. Office of Official publications of the European Community, Luxemburg, p. 133.
- Lambin, E.F., Geist, H.I., Lepers, E., 2003. Dynamics of land use and land cover change in tropical regions. *Ann. Rev. Environ. Resour.* 28, 205–241.
- Liverman, D.E., Moran, R., Rindfuss, R., Stern, P.C. (Eds.), 1998. *People and Pixels: Linking Remote Sensing and Social Science*. Natl. Acad. Press, Washington, D.C.
- Mertens, B., Lambin, E.F., 1997. Spatial modeling of deforestation in Southern Cameroon. *Appl. Geogr.* 17, 143–162.
- Mondal, M.S., Sharma, N., Kappas, M., Garg, P.K., 2012. Modeling of spatio-temporal dynamics of LULC – a review and assessment. *J. Geomatics* 6 (2), 93–103.
- Muller, M.R., Middleton, J., 1994. A Markov model of land-use change dynamics in the Niagara region, Ontario, Canada. *Landscape Ecol.* 9 (2), 151–157.
- Pontius Jr., R.G., 2002. Statistical methods to partition effects of quantity and location during comparison of categorical maps at multiple resolutions. *Photogram. Eng. Rem. Sens.* 68 (10), 1041–1049.
- Pontius Jr., R.G., Malanson, J., 2005. Comparison of the structure and accuracy of two land change models. *Int. J. Geogr. Inf. Sci.* 19 (2), 243–265.
- Pontius Jr., R.G., Schneider, L., 2001. Land-use change model validation by a ROC method for the Ipswich watershed, Massachusetts, USA. *Agric. Ecosyst. Environ.* 85, 239–248.
- Pontius Jr., R.G., Agarwal, A., Huffaker, D., 2003. Estimating the uncertainty of land cover extrapolations while constructing a raster map from tabular data. *J. Geogr. Syst.* 5, 253–273.
- Pontius Jr., R.G., Huffaker, D., Denman, K., 2004. Useful techniques of validation for spatially explicit land-change models. *Ecol. Model.* 179 (4), 445–461.
- Turner II, B.L., Skole, D., Sanderson, S., Fischer, G., Fresco, L., Leemans, R., 1995. *Land-Use and Land-Cover Change, Science/Research Plan, IHDP Report No. 07, IGBP report No. 35*.
- Veldkamp, A., Fresco, I.O., 1996. CLUE: A conceptual model to study the conversion of land use and its effects. *Ecol. Model.* 85 (2–3), 253–270.
- Weng, Q., 2002. Land use change analysis in the Zhujiang Delta of China using satellite remote sensing, GIS and stochastic modelling. *J. Environ. Manage.* 64, 273–284.
- Wu, F., Webster, C.J., 1998. Simulation of land development through the integration of CA and multi-criteria evaluation. *Environ. Plan. B* 25, 103–126.