

A machine learning-based classification of LANDSAT images to map land use and land cover of India

Ram Kumar Singh^a, Prafull Singh^b, Martin Drews^c, Pavan Kumar^d, Hukum Singh^e, Ajay Kumar Gupta^e, Himanshu Govil^f, Amarjeet Kaur^g, Manoj Kumar^{h,*}

^a Department of Natural Resources, TERI School of Advanced Studies, 10, Institutional Area, Vasant Kunj, New Delhi – 110070, India

^b Department of Geology, Central University of South Bihar, Gaya, 824236, India

^c Department of Technology, Management and Economics, Technical University of Denmark, 2800 Kgs Lyngby, Denmark

^d Department of Forest Biology and Tree Improvement, College of Horticulture and Forestry, Rani Lakshmi Bai Central Agricultural University, Jhansi - 284003, India

^e Forest Ecology & Climate Change Division, Forest Research Institute, Dehradun, India

^f Department of Applied Geology, National Institute of Technology, Raipur, Chhattisgarh – 492010, India

^g Centre for Disaster Management Studies, Guru Gobind Singh Indraprastha University, Dwarka, New Delhi – 110078, India

^h GIS Centre, IT & GIS Discipline, Forest Research Institute (FRI), PO: New Forest, Dehradun – 248006, India

ARTICLE INFO

Keywords:

Agriculture
Forest
Remote sensing
Logistic model
Machine learning

ABSTRACT

The land use and land cover (LULC) maps are often required by planners and policymakers for effective planning and management interventions at the local, national, regional and global levels. Various attempts have been made to develop LULC maps using field-based surveys and by processing remotely sensed images. These maps can be developed using different tools and methodologies at different scales to achieve different levels of accuracy. With the advent of remote sensing technologies and its application in making LULC maps, attempts have been made to develop such maps with improved accuracy and consistency. The machine learning-based approaches have been attempted to develop LULC maps with varying levels of accuracy using different satellite images. Making LULC maps for a large region such as India covering a total area of ca. 3,287,469 km² can be a cumbersome process using conventional approaches. Thus the map was developed using remotely sensed images using machine learning algorithm (Mnlogit) on LANDSAT images (2005, 2006, 2007 and 2016) for entire India region. We developed LULC maps of years 2005, 2006, 2007 and 2016 to test the consistency of classification using the trained Mnlogit model using field survey based signatures for corresponding years in respective images. We could achieve reasonably good accuracy varying in the range of 80–86% during all four years. A Kappa statistic - K (hat), was obtained in the range of 0.71–0.81 which indicates reasonably good accuracy. The study can be replicated for other regions using other available satellite remote sensing images to obtain LULC maps. In general, the suggested approach in this study will help planners to obtain LULC maps at different time intervals to study land-use change dynamics in a shorter time and cost-effective way.

1. Introduction

Developing new methodologies to map land use and land cover (LULC) using remote sensing images has emerged as one of the prime concerns among the regional and global communities (Singh et al., 2020a, 2020b; Turner et al., 2007). Remote sensing-based images are now available at higher spatial, spectral and temporal resolution providing opportunities to map LULC dynamics in a better way for

multiple time steps (Dingle Robertson and King, 2011; El Jazouli et al., 2019; Grekousis et al., 2015; Kumar et al., 2020). The LULC maps are important for acquiring information about available natural resources such as agriculture, forest, water bodies and soil to estimate their spatial extent and drivers influencing their change over a period (Boles et al., 2004; Wulder et al., 2018). Monitoring LULC also provides the opportunity to evaluate various ecosystem services (Costanza, 1998; Kumar et al., 2019), the productivity of forests (Kumar et al., 2019b), food

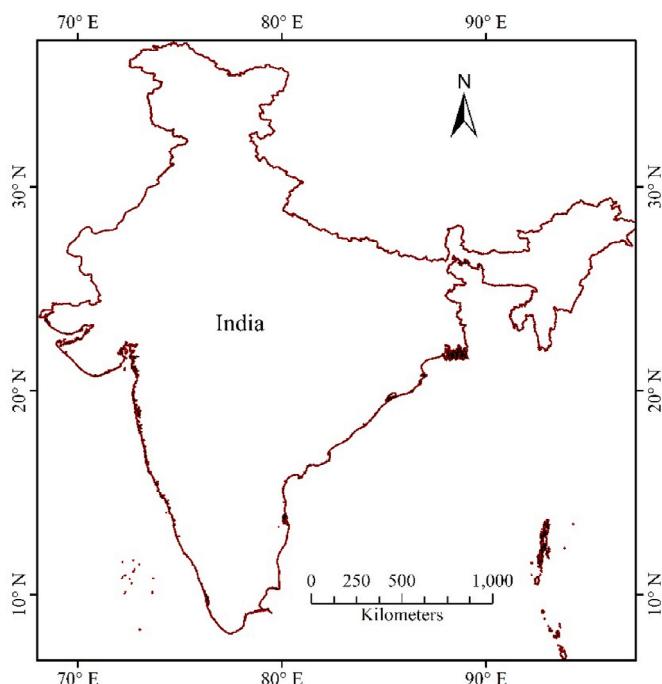
* Corresponding author. GIS Centre, Forest Research Institute (FRI), PO: New Forest, Dehradun – 248006, India.

E-mail addresses: singhramkumar@gmail.com (R.K. Singh), pks.jiwaji@gmail.com (P. Singh), mard@dtu.dk (M. Drews), pawan2607@gmail.com (P. Kumar), hukumsingh97@yahoo.com (H. Singh), ewmajay@gmail.com (A.K. Gupta), himgeo@gmail.com (H. Govil), amarjeet_ip@yahoo.com (A. Kaur), manojfri@gmail.com (M. Kumar).

Table 1

Global land use and land cover (LULC) products using different remote sensing images, classification approach and reported accuracy.

LULC product	Remote Sensing Image/other data used	Classification procedure	Reported accuracy (%)	Spatial resolution	Period of availability	References
Global Land Cover Characterization by (IGBP- DISCover)	AVHRR	K-Mean	81	10 km, 8 km and 1 km	1984	Bai et al. (2014); Jung et al. (2006); Latifovic et al. (2004)
Global Land Cover Classification (UMD Marry land)	AVHRR	Decision-Tree	65–82	1 km	1992–1993	
Global Land Cover (GLC2000)	SPOT 4	ISO-Clustering	66–69	1 km	2000	Bartholome and Belward (2005)
MODIS Land Cover (MCD 12Q1)	MODIS Imagery	Decision-Tree	75	500 m, 5' and 0.5°	2001–2019	Brown et al., 2013; Ran et al., 2010
Synergetic land cover product (SYNMAP)	Merging of GLCC, GLC2000 and MODIS 2001	Threshold and Fuzzy Method	Not available	1 km	2000	Jung et al. (2006)
GlobCover	MERIS satellite Imagery	SVM and Gaussian Likelihood	73 68	300 m	2005 2009	Bontemps et al., 2011
Climate Change Initiative Land Cover (CCI-LC)	MERIS and SPOT	Supervised and Unsupervised	74	300 m	1998–2012	Bontemps et al., 2011
Global Land Survey	Landsat	Hybrid approach	Not available	30 m	1975, 1990 and 2000	Gong et al., 2013
Finer Resolution Observation and Monitoring of Global Land Cover (FROM-GLC)	Landsat ETM	SVM, RF, MLC and J4.8	64–66	30 m	2006	Gong et al., 2013; Price, 2003
GlobalLand30	Landsat ETM	Pixel and Object based	79	30 m	2000 and 2010	Gong et al., 2013; Price, 2003
Global Land Cover-Share	Country-level statistics	Harmonization of Global and Regional datasets	Not available	30 arcSecond	2010	FAO, 2014

**Fig. 1.** The extent of the study region (India) for which land use land cover maps were developed.

security (IPCC, 2006), urban sprawl (Cammarano and Tian, 2018; Sajjad and Iqbal, 2012), climate change-related impacts (IPCC, 2006; Singh et al., 2020b; Sun et al., 2015), forest class change by wildfire and encroachment (Boles et al., 2004; Jin et al., 2017), the degradation of lands (Kumar et al., 2019a; Olokeogun and Kumar, 2020; Pokhriyal et al., 2020; Singh et al., 2020), formulate developmental planning for managing available resources (Long et al., 2007), etc. The “land use” and “land cover” in literature have often been used interchangeably although they are two different terms (Dimyati et al., 1996; Singh et al., 2020a). Land use refers to the kind of use by humans for various activity

on a piece of land while the land cover refers to the physical or the biological cover of the land.

The LULC change dynamics also influence the feedback mechanism of lands and forests upon the climate and thus is an important discipline of research (Meyer and Turner, 1992). Spatially-explicit LULC is important for distinguishing diversity and anthropogenic activity (Turner et al., 2007) which is linked to the physical and human environment (Foody, 2002). The prime classes of LULC could be agriculture, forestry, water, built-up and others (Wulder et al., 2018) based on the specific requirements. In the developing countries such as India and other alike, limited information is readily available on LULC although such maps can readily be made available using satellite-based remote sensing (Otukei and Blaschke, 2010; Singh et al., 2020a). Several advanced algorithms have evolved in recent times that can be used for developing LULC maps to classify remote sensing images, such as the artificial neural networks, decision trees, support vector machines, object-based image analysis, etc. (Grekousis et al., 2015; Kumar et al., 2020; Otukei and Blaschke, 2010; Singh et al., 2020a). The most common classifiers used in processing remote sensing images are K-Means, ISO-Data, Minimum-Distance (Lillesand et al., 2004).

Over the recent decades, various LULC program explored the use of Moderate Resolution Imaging Spectroradiometer (MODIS) (Kastens et al., 2005; Wardlow et al., 2007), Advanced Very High-Resolution Radiometer (AVHRR) (Hansen et al., 2010), and Satellite Pour l'Observation de la Terre (SPOT-4) (Xiao et al., 2002) and other remote sensing data for the mapping of regional and global LULC (Table 1). Global land cover (GLC2000) and the MODIS Collection 4 based LULC products are available for past years at a coarser resolution while with the advent of new sensors providing information at finer resolution can be used for developing maps at a finer scale. The MODIS derived Global Land Cover is available at a spatial resolution of 500-m (DeFries et al., 1998; Shao and Wu, 2008). Congalton et al. (2014) and Grekousis et al. (2015) reviewed the various global LULC products (GLC 2000, IGBP DISCover, UMD Land Cover and GlobCover 2009), where remote sensing data were used to map LULC. As all such products were developed using different images and algorithms with different kinds of limitations thus we can observe the inconsistency in the land use classes of these products from different sources. They have relatively poor mapping accuracy which can further be improved using improved

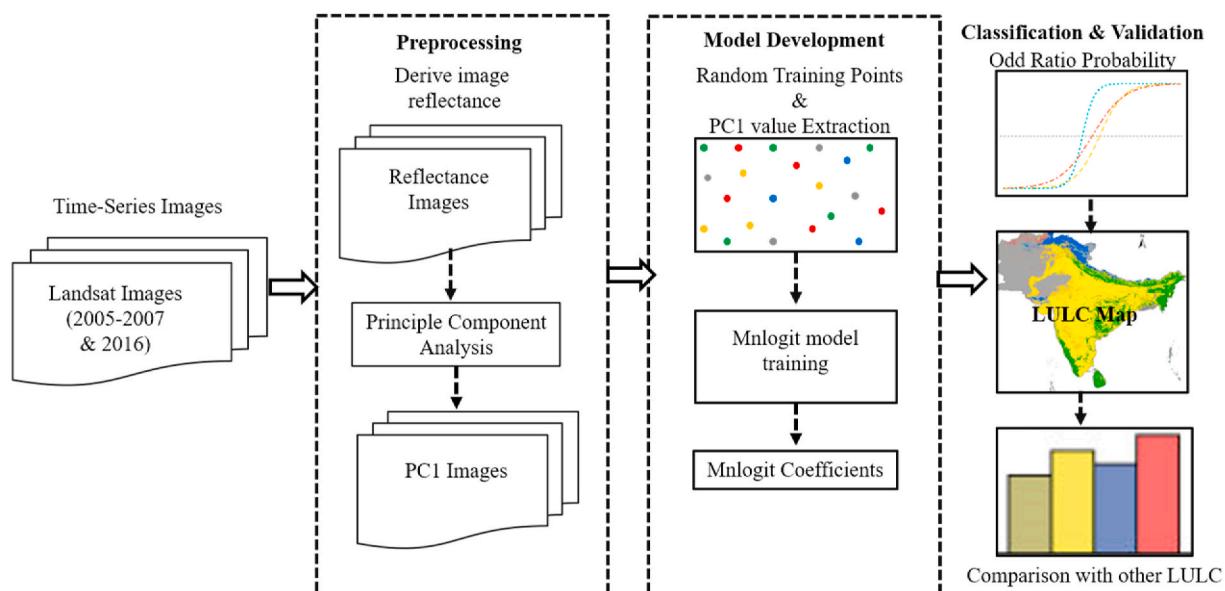


Fig. 2. Schematic representation of adopted methodological steps.

Table 2
Eigen values and Principal Component Analysis (PCA).

PCs	2005		2006		2007		2016	
	Eigen		Eigen		Eigen		Eigen	
	Value	(%)	Value	(%)	Value	(%)	Value	(%)
PC1	39325.6	98.1	41697.3	95.9	41450.7	95.2	52655.0	97.8
PC2	1150.4	1.0	1016.9	2.3	1255.3	2.9	531.6	1.0
PC3	631.7	0.6	624.6	1.4	660.3	1.5	443.9	0.8
PC4	76.0	0.2	63.0	0.1	86.3	0.2	123.7	0.2
PC5	29.3	0.1	23.5	0.1	22.9	0.1	69.1	0.1
PC6	15.8	0.0	15.8	0.0	15.2	0.0	5.6	0.0
PC7	12.5	0.0	12.4	0.0	12.5	0.0	4.5	0.0
PC8	8.5	0.0	8.0	0.0	6.1	0.0	3.5	0.0

algorithms and by using higher resolution images.

In this study, we present Multinomial Logistic Model (Mnlogit) ([Agresti, 2018](#); [Long et al., 2006](#)) based LULC classification of India that can also be replicated for other regions. The Mnlogit model statistically maintains the relationship between the independent predictor that has a logistic function with dependent variables ([Aldrich et al., 1985](#); [Bayaga, 2010](#); [Tabachnick and Fidell, 2007](#)). The LANDSAT time-series multiple images of the year 2005, 2006, 2007 and 2016 were used for training the Mnlogit classifier to map LULC of corresponding years. Our main intention was to compare our results with available LULC products hence we restricted the mapping till 2016 however maps of recent years can also be developed using the suggested approach. In this paper we demonstrate (1) the application of Mnlogit based LULC classification; (2) the testing of consistency while making LULC maps for different years; and (3) to enhance the mapping accuracy using moderate resolution images.

2. Study area

The LULC maps were developed for entire India covering a total geographical extent of 3,287,469 km². The physical extent of India considered for mapping LULC in this study lies between 6.5° S and 38° N latitude and 67.5° E – 97.4° E ([Fig. 1](#)). India represents 2.0% of the world's total area and ranks seventh land cover area. According to the World Bank ([2020](#)), total agriculture and forestry land use classes of India represent 26.0 and 20.8%, respectively. The coverage of various land use classes is quite dynamic under the influence of the ever-

increasing human population and ongoing developmental activities. Usually, the agriculture classes in India are under transition into the built-up area while the forested lands have been converted into settlement or agriculture classes.

3. Material and methods

The prime methodological framework adopted in this study is covered under three components, viz., (i) data pre-processing, (ii) training of Mnlogit classifier, and (iii) accuracy assessment. The pre-processing of Landsat images were done for atmospheric correction and obtaining the principal components (PC) used for training of the model. Field based signatures were not possible for the whole country therefore signature of respective land use classes were obtained from the Google imageries (<https://www.google.com/earth/>). Also, it was not possible to obtain landuse class signature of past years thus Google Earth imageries were an obvious choice. The Mnlogit classifier model was trained using Google Earth referenced random points (total 1800 points) for different land use classes. Google Earth provides high resolution imageries where various land use classes can easily be distinguished. The random points were generated in ERDAS Imagine 2020 using the tool “create random point”. The generated random points were initially more than 1800 and each point was labelled manually using on-screen visual interpretation to show its corresponding land use of the classification scheme of our work. The points that were not falling in the designated land use class were eliminated to have a final 1800 points for modelling. Overall methodological steps adopted in this study is

Table 3

The coefficients of the Mnlogit model for the year 2005, 2006, 2007 and 2016.

		2005								2006												
Coef.	Ref.	Agriculture		Forest		Built-up		Others		Water		Agriculture		Forest		Built-up		Others		Water		
		Value	P Value	Value	P Value	Value	P Value	Value	P Value	Value	P Value	Value	P Value	Value	P Value	Value	P Value	Value	P Value	Value	P Value	
Intercept	Agriculture	18.13	0.00	-13.44	0.00	-20.81	0.00	-27.94	0.00	19.18	0.00	-11.38	0.00	-20.68	0.00	-27.83	0.00					
PC1		-0.03	0.00	0.03	0.00	0.06	0.00	0.09	0.00	-0.03	0.00	0.03	0.00	0.06	0.00	0.09	0.00					
PC2		0.01	0.00	-0.02	0.00	-0.02	0.00	-0.04	0.00	0.00	0.03	0.03	0.01	0.00	0.01	0.18	0.04	0.00				
Intercept	Forest	-19.76	0.00	-33.51	0.00	-40.81	0.00	-46.70	0.00	-18.76	0.00	0.00	0.03	0.01	0.00	-30.44	0.00	-40.44	0.00	-49.30	0.00	
PC1		0.03	0.00	0.07	0.00	0.09	0.00	0.12	0.00	0.03	0.00	0.06	0.00	0.09	0.00	0.12	0.00					
PC2		-0.02	0.00	-0.03	0.00	-0.04	0.00	-0.05	0.00	0.00	0.06	0.02	0.00	0.01	0.05	0.04	0.00					
Intercept	Built-up	12.32	0.00	30.00	0.00	-7.74	0.00	-16.90	0.00	11.39	0.00	31.34	0.00	-9.26	0.00	-16.36	0.00					
PC1		-0.03	0.00	-0.06	0.00	0.02	0.00	0.06	0.00	-0.03	0.00	-0.06	0.00	0.03	0.00	0.06	0.00					
PC2		0.02	0.00	0.02	0.00	-0.01	0.02	-0.03	0.00	-0.01	0.00	-0.02	0.00	0.00	0.46	0.02	0.01					
Intercept	Others	21.77	0.00	38.84	0.00	8.59	0.00	-8.17	0.00	20.01	0.00	41.66	0.00	8.84	0.00	-7.86	0.00					
PC1		-0.06	0.00	-0.08	0.00	-0.02	0.00	0.04	0.00	-0.05	0.00	-0.09	0.00	-0.03	0.00	0.04	0.00					
PC2		0.02	0.00	0.03	0.00	0.01	0.11	-0.02	0.00	-0.01	0.01	-0.01	0.18	0.00	0.70	0.02	0.00					
Intercept	Water	29.39	0.00	46.45	0.00	15.75	0.00	8.19	0.00	27.91	0.00	51.18	0.00	18.53	0.00	7.23	27.91					
PC1		-0.10	0.00	-0.12	0.00	-0.06	0.00	-0.04	0.00	-0.09	0.00	-0.14	0.00	-0.07	0.00	-0.03	-0.09					
PC2		0.05	0.00	0.07	0.00	0.04	0.00	0.03	0.00	-0.03	0.00	-0.04	0.00	-0.03	0.00	-0.02	-0.03					
4		2007								2016												
	Coef.	Ref.	Agriculture Value	Agriculture P Value	Forest Value	Forest P Value	Built-up Value	Built-up P Value	Others Value	Others P Value	Water Value	Water P value	Agriculture Value	Agriculture P Value	Forest Value	Forest P Value	Built-up Value	Built-up P value	Others Value	Others P Value	Water Value	Water P Value
	Intercept	Agriculture	19.45	0.00	-12.17	0.00	-19.94	0.00	-29.09	0.00	16.81	0.00	-8.22	0.00	-10.82	0.00	-16.89	0.00				
	PC1		-0.03	0.00	0.03	0.00	0.05	0.00	0.09	0.00	-0.03	0.00	0.02	0.00	0.03	0.00	0.05	0.00				
	PC2		0.00	0.16	0.01	0.01	0.01	0.00	0.04	0.00	0.01	0.00	-0.02	0.00	-0.02	0.00	-0.02	0.00				
	Intercept	Forest	-19.43	0.00	-29.72	0.00	-39.33	0.00	-47.45	0.00	-16.88	0.00	-25.08	0.00	-27.54	0.00	-33.75	0.00				
	PC1		0.03	0.00	0.06	0.00	0.08	0.00	0.12	0.00	0.028	0.00	0.05	0.00	0.06	0.00	0.08	0.00				
	PC2		0.00	0.03	0.01	0.10	0.01	0.08	0.03	0.00	-0.01	0.00	-0.03	0.00	-0.03	0.00	-0.03	0.00				
	Intercept	Built-up	12.42	0.00	31.12	0.00	-7.91	0.00	-16.81	0.00	8.604	0.00	25.63	0.00	-2.77	0.00	-7.62	0.00				
	PC1		-0.03	0.00	-0.06	0.00	0.02	0.00	0.06	0.00	-0.023	0.00	-0.05	0.00	0.01	0.00	0.03	0.00				
	PC2		-0.01	0.00	0.00	0.39	0.00	0.74	0.03	0.00	0.014	0.00	0.03	0.00	0.00	0.41	-0.01	0.08				
	Intercept	Others	20.08	0.00	39.23	0.00	9.22	0.00	-8.29	0.00	11.32	0.00	29.07	0.00	2.87	0.00	-4.17	0.00				
	PC1		-0.05	0.00	-0.08	0.00	-0.03	0.00	0.04	0.00	-0.031	0.00	-0.06	0.00	-0.01	0.00	0.02	0.00				
	PC2		-0.01	0.00	-0.01	0.25	0.00	0.40	0.03	0.00	0.018	0.00	0.03	0.00	0.00	0.26	0.00	0.21				
	Intercept	Water	27.39	0.00	47.27	0.00	17.98	0.00	7.88	27.39	15.06	0.00	33.60	0.00	6.63	0.00	3.66	15.06				
	PC1		-0.09	0.00	-0.12	0.00	-0.06	0.00	-0.03	-0.09	-0.048	0.00	-0.08	0.00	-0.02	0.00	-0.02	-0.048				
	PC2		-0.03	0.00	-0.03	0.00	-0.03	0.00	-0.02	-0.03	0.027	0.00	0.04	0.00	0.01	0.12	0.01	0.027				

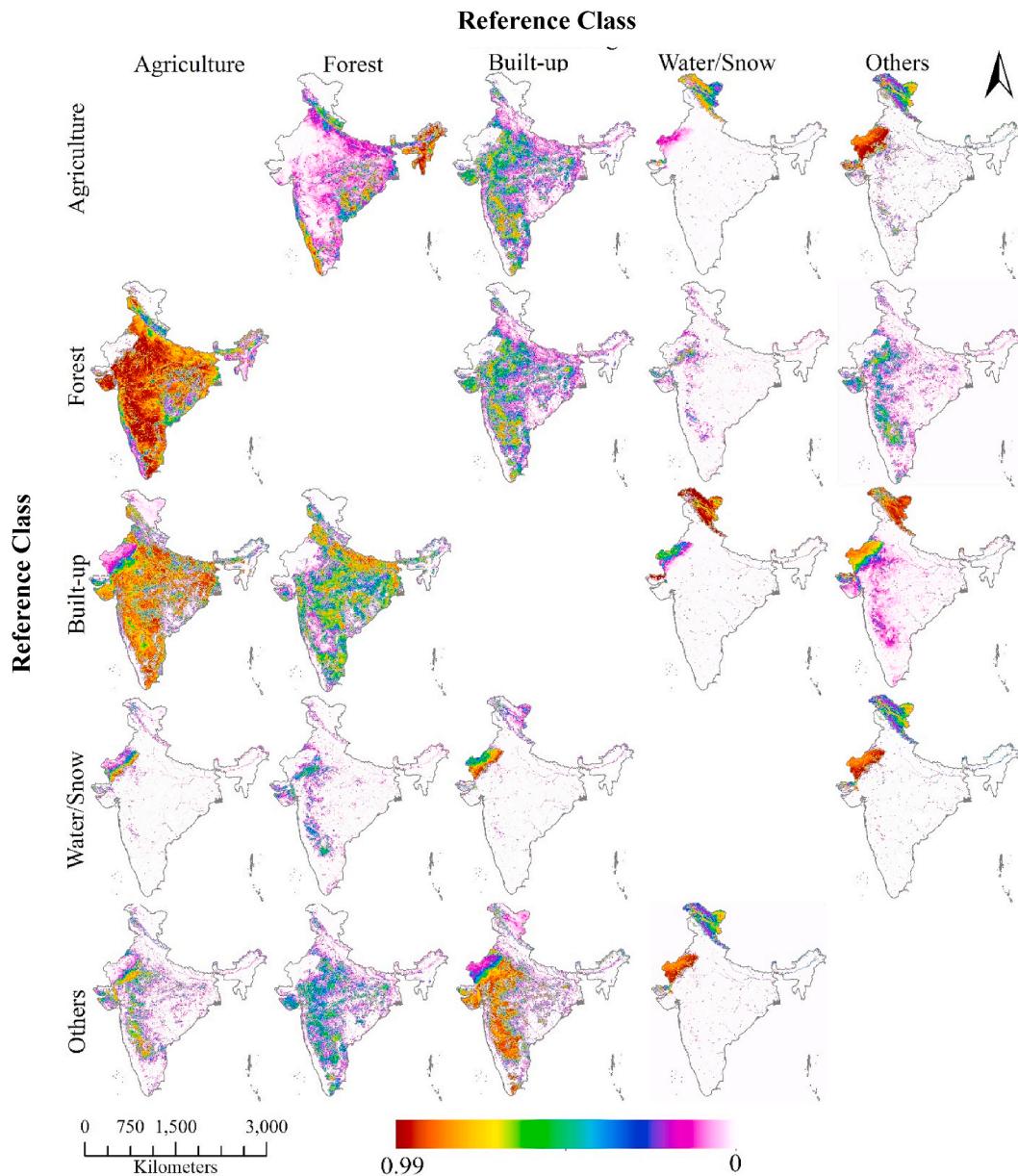


Fig. 3. The Odd Ratio probability land use land cover class maps of the year 2005.

presented in Fig. 2.

3.1. Data procurement and pre-processing

The time series images of Landsat were acquired from the archives of USGS Earth Explorer (<https://earthexplorer.usgs.gov/>) for the years 2005, 2006, 2007 (Landsat 5) and 2016 (Landsat 8). The Landsat images are available at a spatial resolution of 30 m. Initial image processing was done before deriving the reflectance values at pixel levels that included absolute radiometric correction, conversion of digital numbers into radiance unit by applying solar correction using parameters provided in the Landsat metadata files. The problem of misalignment among Landsat scenes was resolved using ERDAS IMAGINE tool “AutoSync”. We obtained principal components (PCs) of all the images in a year (multiple images available for a year) to use all bands that had 30 m resolution while we also excluded aerosol and thermal bands. The PCA of multiple year images were obtained using ERDAS IMAGINE 2020 “Spectral analysis-PCA tool”. The principal component analysis was done to remove data redundancy and the principal components were obtained

for each of the years separately. PCs that explained ~98% variability were used to train Mnlogit classifier using random training points. The steps used for Mnlogit development using PCs are explained in the subsequent section.

3.2. Mnlogit model development

The PC1 and PC2 were found to explain ~ >98% variability for all four year assessments. The extracted pixel values of PC1 and PC2 images and its corresponding signature land use class were used to train the Mnlogit model using R software package “Mlogit” (Croissant, 2012). The coefficients obtained were used to develop Mnlogit Model for each corresponding year. We used 75% (i.e 1350 points) of the equalized random points to train the model and other remaining 25% points (450 points) were used for the validation purpose.

The PC1 and PC2 were used as the explanatory (independent) variables for different land use class (agriculture, forestry, water, built up and other) that acted as the dependent variables. The logit P_b is a logistic probability of class occurrence P_b associated with each pixel for its

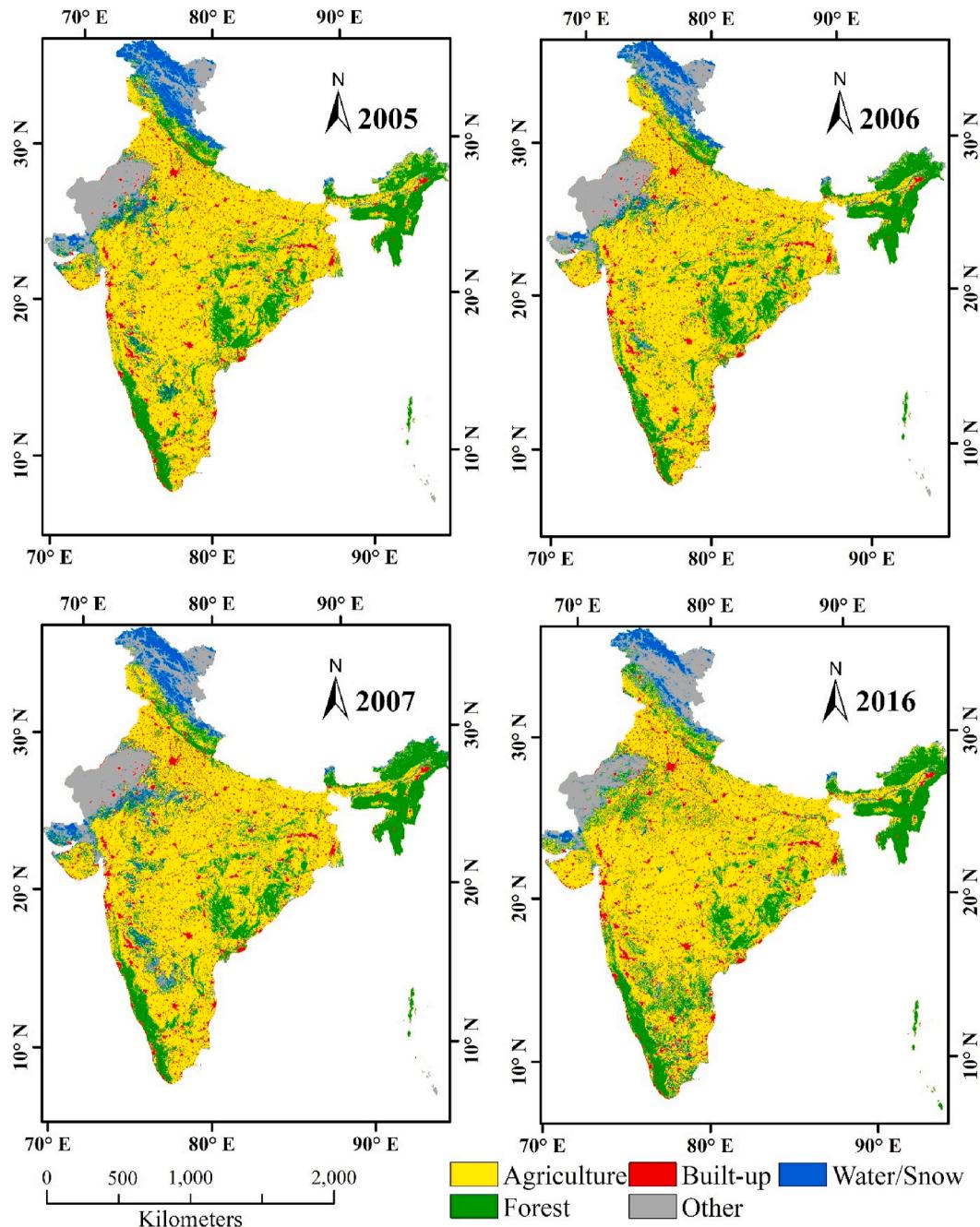


Fig. 4. The land use land cover maps of India obtained using Mnlogit model for the year 2005, 2006, 2007 and 2016.

different land use class. The coefficients β_0 , β_1 , and β_2 were estimated for the explanatory variables PC1 and PC2 as indicated in Equation (1) (Aldrich et al., 1985; Goeman and Le Cessie, 2006). The Odd ratio of different land use classes was obtained by keeping one selected class as reference value by using Equation-2. The Odd ratio of reference class is defined as the ratio of probability of obtaining reference class ($Pb_{\text{reference class}}$) and the probability of not obtaining reference class (i.e. $1 - Pb_{\text{reference class}}$). The Odd ratio images obtained for all reference land use classes were used to finally mark the land use having highest Odd ratio value. We predicted five different land use/land cover classes of agriculture, built up, forest, others and water shown as AGR, BUT, FOR, OTH and WAT, respectively.

$$Pb_{\text{reference class}} = \beta_0 + \beta_1 \cdot PC1 + \beta_2 \cdot PC2 \quad \text{Eq. (1)}$$

$$\text{Odd Ratio}_{\text{reference class}} = \frac{Pb_{\text{reference class}}}{1 - Pb_{\text{reference class}}} \quad \text{Eq. (2)}$$

$$\text{Mnlogit } (Pb_{\text{reference class}}) = \text{Log} \left(\frac{Pb_{\text{reference class}}}{1 - Pb_{\text{reference class}}} \right) \text{ where } 0 < Pb < 1 \quad \text{Eq. (3)}$$

3.2. Consistency and accuracy assessment

The consistency of LULC classification was defined by matching the LULC class area mapped by Mnlogit approach with the available statistics provided by the International Geosphere Biosphere Programme (IGBP) and Food and Agriculture Organisation (FAO). We also compared Mnlogit individual land cover class with the aggregate area of land cover

Table 4

Pseudo R² values (McFadden, Cox & Snell and Nagelkerke) obtained to indicate the performance of the Mnlogit model.

	Reference	R ² McFadden	R ² Cox & Snell	R ² Nagelkerke
2005	Agriculture	0.65	0.88	0.92
	Forest	0.64	0.88	0.92
	Built-up	0.63	0.88	0.92
	Others	0.65	0.88	0.92
	Water	0.64	0.88	0.92
2006	Agriculture	0.65	0.87	0.92
	Forest	0.64	0.88	0.93
	Built-up	0.63	0.85	0.91
	Others	0.65	0.87	0.90
	Water	0.64	0.87	0.89
2007	Agriculture	0.66	0.88	0.93
	Forest	0.63	0.88	0.92
	Built-up	0.61	0.85	0.90
	Others	0.66	0.87	0.93
	Water	0.63	0.85	0.91
2016	Agriculture	0.62	0.87	0.90
	Forest	0.63	0.88	0.91
	Built-up	0.61	0.87	0.88
	Others	0.63	0.87	0.91
	Water	0.62	0.88	0.90

Table 5

The assessment of K (hat) and overall accuracy for the years 2005, 2006, 2007 and 2016.

	Agriculture	Forest	Water	Built-up	Other
2005	K (hat)	0.75	0.81	0.79	0.81
	Overall	82.01	84.25	78.35	81.32
2006	K (hat)	0.88	0.84	0.86	0.85
	Overall	85.11	86	80.48	81.83
2007	K (hat)	0.79	0.79	0.78	0.76
	Overall	83.75	84.26	80.25	78.25
2016	K (hat)	0.74	0.78	0.74	0.73
	Overall	83.29	85.48	78.28	79.07
					79.73

classes mapped by the Atlas of India created by “National Remote Sensing Center (NRSC)” available for the year 2005. The 16 different land use land cover classes shown in the IGBP and NRSC maps were converted into five classes for comparing with Mnlogit classified map. The accuracy assessment of land cover classes was carried by generating equalized random points over Mnlogit classified land cover class and were matched with the actual land cover observed through Google Earth high-resolution images. Accuracy assessment of classified images for each year was done by comparing with 450 random land use points obtained from the Google Earth images. Google Earth images however are not available for all years and hence the years for which images were not available we used immediate year images. The K (hat) and overall accuracy value was used to define the accuracy of mapping.

4. Results and discussion

The results obtained for each of the significant steps of the study is presented in separate sections ahead.

4.1. Data processing and compression

The reflectance values represented in multiple images of a year were processed to obtain eight PCA images (Table 2). It was found that PC1 and PC2 explain ~98% variability and the Eigen value of each PCs for different year images is given in Table 2.

4.2. Mnlogit model

The coefficients of Mnlogit model obtained for each of the years is presented in Table 3. The Odd ratio images for the year 2005, 2006,

2007 and 2016 were developed using these coefficients considering PC1 and PC2 as predictor variables. The pixel values of Odd ratio images show the probability value of the land use class obtained with respect to the reference land use class. The Odd ratio probability LULC maps obtained for the year 2005 is shown in Fig. 3. Every time, one land use class was considered as the reference and probability values of other land use classes were obtained. Thus total 20 such maps were obtained and the final land use class to pixels were marked that qualified as having maximum probability value for a given class. The final LULC maps of years 2005, 2006, 2007 and 2016 prepared by adopting the mentioned approach is provided in Fig. 4. The performance of the Mnlogit model is given by Pseudo R² value as indicated in Table 4.

4.3. Accuracy and consistency assessment

The accuracy of mapping different land use classes obtained as K (hat) statistics and overall accuracy is presented in Table 5. We could obtain the highest overall accuracy for the agriculture and forestry classes that varied between 82 and 86% in all four years. The overall accuracy of rest of the classes was in the range of 78–80%. The consistency of LULC map was judged by visualizing the total land area representing corresponding classes in subsequent years and by matching their values with the value reported by FAO (2005–2007) and IGBP (2005–2007 and 2016) (Fig. 5). It was observed that FAO reports a constant value of agriculture and forest area for all three years (i.e. 2005–2007) which may not be a real case. Moreover, the land use area presented by FAO is not derived by processing remotely sensed images while it is a value as reported by individual countries using available statistics. In contrary to FAO approach, IGBP reports the area of land use classes by processing remote sensing images (MODIS NDVI product) with 67–80% accuracy (Foody 2002; Herold et al., 2008; Scepan and Hansen 1999). The comparison of Mnlogit based LULC map with the National Remote Sensing Center (NRSC) natural resource atlas 2005 (NRSC, 2006), is presented in Table 6. We observed forestry class match accurately, while the other classes differed in the absolute area.

We attempted the Mnlogit based classification to obtain the land use map of India by processing LANDSAT images for the year 2005, 2006, 2007 and 2016. The Mnlogit based classification achieved reasonable accuracy in mapping LULC for a large area representing India. This approach can also be applied and tested for other regions to compare the results and accuracy for suggesting a more reliable approach for obtaining maps with improved accuracy. Although various attempts have been made to map the LULC at the national, regional and global scale, still people are attempting to achieve more reliable maps (Sun et al., 2020). We could observe an inconsistency in reporting different land use area of India by different agencies which indicate that different approaches may provide different area for the same situation. The maps provided by different agencies report various levels of accuracy such as the IGBP land cover map is reported to have 67–80% accuracy. We demonstrated here the mapping of LULC with high accuracy for forestry and agriculture classes that ranged between 82 and 86% which are the two most important and dominant land use class while the other classes were mapped with an accuracy of 78–82%. The method may also be tested to map the global LULC mapping using other remote sensing data of higher spatial resolution. Sun et al. (2020) used deep neural network to classify agricultural land and reported that such set ups mapped more accurately compared to the wetlands and other land features. They achieved an overall accuracy of 82% for all land types considered together to match with the US Department of agriculture cropland. They emphasized using multiple time layers to train the model for achieving a better accuracy. Our accuracy of mapping forestry and agriculture is at par with the US department of agriculture cropland data.

The land cover of IGBP programme is mapped using MODIS derived product NDVI at 500 m spatial resolution using decision tree supervised classification method (Lambin et al., 2006). The other land cover products “Global Land cover 2000” (GLC2000) uses NDVI of

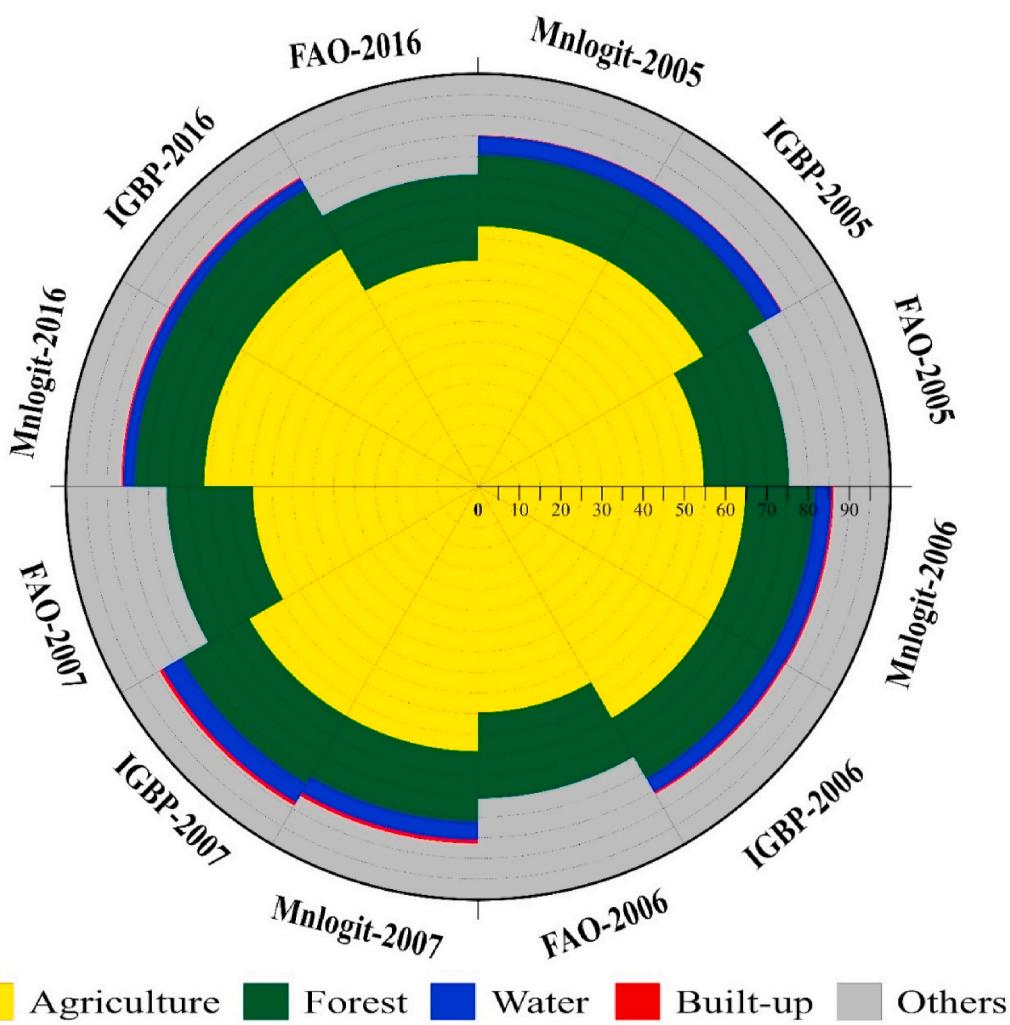


Fig. 5. The polar bar chart showing the comparison of land use classes mapped by IGBP and FAO and its comparison with Mnlogit to observe the consistency.

Table 6

Land use land cover area of India mapped using Mnlogit and its comparison with the area reported in the natural resource atlas of NRSC.

Mapping approach	Land use land cover class (2005)				
	Agriculture	Forest	Water	Built-Up	Other
Mnlogit	60.14	20.20	2.11	2.34	15.21
Natural resource atlas of NRSC	58.16	17.74	1.67	2.54	19.89

SPOT-VEGETATION-1 images and ISO-Data clustering method to classify land cover with an overall accuracy of ~67% (Bartholome and Belward, 2005; Xiao et al., 2002). The global land use map GLC2000 is available for the year 2000 at a spatial resolution of 1 km. The GlobCover-2005/2009 is created using MEdium Resolution Imaging Spectrometer (MERIS) images of 13 spectral bands at 300 m spatial resolution that has been classified using Support Vector Machine and Gaussian likelihood method with overall accuracy ranging between 68 and 73% (Bontemps et al., 2015). Global Land Cover by UMD Marry land used MODIS derived NDVI of 500 m spatial resolution to obtain LULC map by applying decision-tree based classification method to obtain 65–81% accuracy (Bai et al., 2014; Jung et al., 2006; Latifovic et al., 2004). In most of the cases, the accuracy of classification is tested by using random test sample points. The maps prepared by various

agencies use a different approach and different images at varying spatial resolution thus the inconsistency in reporting is quite obvious. Nevertheless, we present here another approach for mapping a relatively larger area with improved accuracy.

5. Conclusion

The mapping of different land use classes is an essential component for the management and planning purpose. The two dominant mainland cover representing forestry and agriculture supports the economy of the developing countries including India while they are most often under extreme anthropogenic pressure. The climatic variabilities and climate change pose additional stress on these resources. The change dynamics of LULC classes at varying time steps can be used for linking the possible drivers of change. Such studies help to plan management interventions for the protection of valuable natural resources. The remote sensing time-series data from the various platforms can be used to obtain such information more pragmatically. With increasing technological advancement in procuring remote sensing-based Earth observations, it is possible to trace the temporal dynamics of land use change and its associated drivers. However, there is a requirement to develop and test new methodologies for assimilating the huge bunch of data which is made available by multiple satellite based sensors. Machine learning based approach has often been applied to derive valuable information for multiple purposes. Processing of remote sensing images has also been attempted by adopting different algorithms of machine learning to

utilize available information in the most effective way to achieve better accuracy of classification. The approach demonstrated by us can be tested for the local, regional as well as global LULC mapping to test its reliability and accuracy.

Author statement

RK and MK developed the original draft of the MS, implemented analysis, developed maps, revised the MS and compiled relevant references. **PKS, MD and PK** helped in the building of the Mnlogitmodel. **HS, AKG, HG** and **AK** contributed in the development of concept and methodology, data curation, and analysis.

Funding

There was no funding support for the study.

Ethical statement

I along with my all learned co-authors jointly declare that.

- All ethical practices have been followed in relation to the development, writing, and publication of the article.
- The research meets all applicable standards with regards to the ethics of experimentation and research integrity and being certified/declared true.
- There is no duplicate publication, fraud, plagiarism, or other malicious concerns involved while submitting the MS.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

The authors acknowledge USGS Earth Explorer for providing Landsat images free of cost. Authors are thankful to Hexagon Geospatial, USA and Forest Research Institute, India for providing resources including valuable insight by its resource persons at multiple stages for the completion of this study.

References

- Agresti, A., 2018. *An Introduction to Categorical Data Analysis*. Wiley.
- Aldrich, J.H., Nelson, F.D., Adler, E.S., 1985. *Linear Probability, Logit, and Probit Models*. Sage. SAGE Publications, London, Thousand Oaks, CA.
- Bai, Y., Feng, M., Jiang, H., Wang, J., Zhu, Y., Liu, Y., 2014. Assessing Consistency of Five Global Land Cover Data Sets in China 8739–8759. <https://doi.org/10.3390/rs6098739>.
- Bartholome, E., Belward, A.S., 2005. GLC2000: a new approach to global land cover mapping from Earth observation data. *Int. J. Rem. Sens.* 26, 1959–1977.
- Bayaga, A., 2010. Multinomial logistic regression: usage and application IN risk analysis. *J. Appl. Quant. methods* 5.
- Boles, S., Xiao, X., Liu, J., Zhang, Q., Munkhtuya, S., Chen, S., Ojima, D., 2004. Land cover characterization of Temperate East Asia using multi-temporal VEGETATION sensor data. *Remote Sens. Environ.* 90, 477–489. <https://doi.org/10.1016/j.rse.2004.01.016>.
- Bontemps, S., Boettcher, M., Brockmann, C., Kirches, G., Lamarche, C., Radoux, J., Santoro, M., Van Bogaert, E., Wegmüller, U., Herold, M., Achard, F., Ramoim, F., Arino, O., Defourny, P., 2015. Multi-year global land cover mapping at 300 M and characterization for climate modelling: achievements of the land cover component of the ESA climate change initiative. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci. - ISPRS Arch.* 40, 323–328. <https://doi.org/10.5194/isprarchives-XL-7-W3-323-2015>.
- Cammarano, D., Tian, D., 2018. The effects of projected climate and climate extremes on a winter and summer crop in the southeast USA. *Agric. For. Meteorol.* 248, 109–118. <https://doi.org/10.1016/j.agrformet.2017.09.007>.
- Congalton, R., Gu, J., Yadav, K., Thenkabail, P., Ozdogan, M., 2014. Global land cover mapping: a review and uncertainty analysis. *Rem. Sens.* 6, 12070–12093.
- Costanza, R., 1998. Using dynamic modelling to scope environmental problems. *Environ. Manag.* 22, 183–195.
- Croissant, Y., 2012. Estimation of Multinomial Logit Models in R: the Mlogit Packages. R Packag. Version 0.2-2. URL <http://cran.r-project.org/web/packages/mlogit/vignettes/mlogit.pdf>.
- DeFries, R.S., Hansen, M., Townshend, J.R.G., Sohlberg, R., 1998. Global land cover classifications at 8 km spatial resolution: the use of training data derived from Landsat imagery in decision tree classifiers. *Int. J. Rem. Sens.* 19, 3141–3168. <https://doi.org/10.1080/014311698214235>.
- Dimyati, M.U.H., Mizuno, K., Kobayashi, S., Kitamura, T., 1996. An analysis of land use/cover change in Indonesia. *Int. J. Rem. Sens.* 17, 931–944.
- Dingle Robertson, L., King, D.J., 2011. Comparison of pixel-and object-based classification in land cover change mapping. *Int. J. Rem. Sens.* 32, 1505–1529.
- El Jazouli, A., Barakat, A., Khellouk, R., Rais, J., El Baghdadi, M., 2019. Remote sensing and GIS techniques for prediction of land use/land cover change effects on soil erosion in the high basin of the Oum Er Rbia River (Morocco). *Remote Sens. Appl. Soc. Environ.* 13, 361–374.
- Foody, G.M., 2002. Status of land cover classification accuracy assessment. *Remote Sens. Environ.* [https://doi.org/10.1016/S0034-4257\(01\)00295-4](https://doi.org/10.1016/S0034-4257(01)00295-4).
- Goeman, J.J., Le Cessie, S., 2006. A goodness-of-fit test for multinomial logistic regression. *Biometrics* 62, 980–985. <https://doi.org/10.1111/j.1541-0420.2006.00581.x>.
- Grekousis, G., Mountrakis, G., Kavouras, M., 2015. An overview of 21 global and 43 regional land-cover mapping products. *Int. J. Rem. Sens.* 36, 5309–5335.
- Hansen, M.C., DeFries, R.S., Townshend, J.R.G., Sohlberg, R., 2010. Global land cover classification at 1 km spatial resolution using a classification tree approach. *Int. J. Rem. Sens.* 21, 1331–1364.
- Herold, M., Mayaux, P., Woodcock, C.E., Baccini, A., Schmullius, C., 2008. Some challenges in global land cover mapping: an assessment of agreement and accuracy in existing 1 km datasets. *Remote Sens. Environ.* 112, 2538–2556. <https://doi.org/10.1016/j.rse.2007.11.013>.
- IPCC, 2006. VOLUME 4 AGRICULTURE , FORESTRY and OTHER LAND USE (AFOLU), Forestry.
- Jin, S., Yang, L., Zhu, Z., Homer, C., 2017. A land cover change detection and classification protocol for updating Alaska NLCD 2001 to 2011. *Remote Sens. Environ.* 195, 44–55. <https://doi.org/10.1016/j.rse.2017.04.021>.
- Jung, M., Henkel, K., Herold, M., Churkina, G., 2006. Exploiting synergies of global land cover products for carbon cycle modeling. *Remote Sens. Environ.* 101, 534–553. <https://doi.org/10.1016/j.rse.2006.01.020>.
- Kastens, J.H., Kastens, T.L., Kastens, D.L.A., Price, K.P., Martinko, E.A., Lee, R.Y., 2005. Image masking for crop yield forecasting using AVHRR NDVI time series imagery. *Remote Sens. Environ.* <https://doi.org/10.1016/j.rse.2005.09.010>.
- Kumar, M., Padalia, H., Nandy, S., Singh, H., Khaiter, P., Kalra, N., 2019a. Does spatial heterogeneity of landscape explain the process of plant invasion? A case study of *Hypsis suaveoleens* from Indian Western Himalaya. *Environ. Monit. Assess.* 191, 794. <https://doi.org/10.1007/s10661-019-7682-y>.
- Kumar, M., Savita, Singh, H., Pandey, R., Singh, M.P., Ravindranath, N.H., Kalra, N., 2019b. Assessing vulnerability of forest ecosystem in the Indian Western Himalayan region using trends of net primary productivity. *Biodivers. Conserv.* 28, 2163–2182.
- Kumar, M., Singh, H., Padalia, H., 2020. Remote sensing for mapping invasive alien plants: opportunities and Challenges. In: *Invasive Species: A Handbook*. Indian Council of Forestry Research and Education, pp. 16–31.
- Lambin, E.F., Geist, H., McConnell, W., Moran, E., Alves, D., Rudel, T., 2006. Land-use and land-cover change: Local processes and global impacts. *Land-Use Land-Cover Chang* 222. <https://doi.org/10.1007/3-540-32202-7>.
- Latifovic, R., Zhu, Z.L., Cihlar, J., Giri, C., Olthof, I., 2004. Land cover mapping of north and Central America - global land cover 2000. *Remote Sens. Environ.* 89, 116–127. <https://doi.org/10.1016/j.rse.2003.11.002>.
- Lillesand, T.M., Kiefer, R.W., Chipman, J.W., 2004. Remote sensing and image interpretation. *Geogr. J.* <https://doi.org/10.2307/634969>.
- Long, S.J., Long, J.S., Freese, J., 2006. *Regression Models for Categorical Dependent Variables Using Stata*. Stata press.
- Long, H., Tang, G., Li, X., Heilig, G.K., 2007. Socio-economic driving forces of land-use change in Kunshan, the Yangtze River Delta economic area of China. *J. Environ. Manag.* 83, 351–364. <https://doi.org/10.1016/j.jenvman.2006.04.003>.
- Meyer, W.B., Turner, B.L., 1992. Human population growth and global land-use/cover change. *Annu. Rev. Ecol. Systemat.* 23, 39–61. <https://doi.org/10.1146/annurev.es.23.110192.000351>.
- National Remote Sensing Agency (NRSC), 2006. *The Atlas of Natural Resource Census*. Department of Science and Technology, India.
- Olokeogun, O.S., Kumar, M., 2020. An indicator based approach for assessing the vulnerability of riparian ecosystem under the influence of urbanization in the Indian Himalayan city, Dehradun. *Ecol. Indicat.* <https://doi.org/10.1016/j.ecolind.2020.106796>.
- Otukey, J.R., Blaschke, T., 2010. Land cover change assessment using decision trees, support vector machines and maximum likelihood classification algorithms. *Int. J. Appl. Earth Obs. Geoinf.* <https://doi.org/10.1016/j.jag.2009.11.002>.
- Pokhriyal, P., Rehman, S., Krishna, G.A., Rajiv, P., Manoj, K., 2020. Assessing forest cover vulnerability in Uttarakhand, India using analytical hierarchy process. *Model. Earth Syst. Environ.* <https://doi.org/10.1007/s40808-019-00710-y>.
- Sajjad, H., Iqbal, M., 2012. Impact of urbanization on land use/land cover of dudhganga watershed of Kashmir Valley, India. *Int. J. Urban Sci.* 16, 321–339. <https://doi.org/10.1080/12265934.2012.743749>.
- Scepan, J., Menz, G., Hansen, M.C., 1999. The DISCover validation image interpretation process. *Photogramm. Eng. Rem. Sens.* 65, 1075–1081.

- Shao, G., Wu, J., 2008. On the accuracy of landscape pattern analysis using remote sensing data. *Landsc. Ecol.* 23, 505–511. <https://doi.org/10.1007/s10980-008-9215-x>.
- Singh, H., Kumar, N., Kumar, M., Singh, R., 2020. Modelling habitat suitability of western tragopan (*Tragopan melanocephalus*) a range-restricted vulnerable bird species of the Himalayan region, in response to climate change. *Clim. Risk Manag.* 29, 100241.
- Singh, R.K., Sinha, V.S.P., Joshi, P.K., Kumar, M., 2020a. A multinomial logistic model-based land use and land cover classification for the South Asian Association for Regional Cooperation nations using Moderate Resolution Imaging Spectroradiometer product. *Environ. Dev. Sustain.* 1–22.
- Singh, R.K., Sinha, V.S.P., Joshi, P.K., Kumar, M., 2020b. Modelling agriculture, forestry and other land use (AFOLU) in response to climate change scenarios for the SAARC nations. *Environ. Monit. Assess.* 192, 1–18.
- Sun, Y., Yang, Y., Zhang, Y., Wang, Z., 2015. Assessing Vegetation Dynamics and Their Relationships with Climatic Variability in Northern China 88. <https://doi.org/10.1016/j.pce.2015.09.018>.
- Sun, Z., Di, L., Fang, H., Burgess, A., 2020. Deep learning classification for crop types in north Dakota. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* 13, 2200–2213.
- Tabachnick, B.G., Fidell, L.S., 2007. Using Multivariate Statistics. Allyn & Bacon/Pearson Education.
- Turner, B.L., Lambin, E.F., Reenberg, A., 2007. The emergence of land change science for global environmental change and sustainability. *Proc. Natl. Acad. Sci. Unit. States Am.* <https://doi.org/10.1073/pnas.0704119104>.
- Wardlow, B.D., Egbert, S.L., Kastens, J.H., 2007. Analysis of time-series MODIS 250 m vegetation index data for crop classification in the US Central Great Plains. *Remote Sens. Environ.* 108, 290–310.
- Wulder, M.A., Coops, N.C., Roy, D.P., White, J.C., Hermosilla, T., 2018. Land cover 2.0. *Int. J. Rem. Sens.* <https://doi.org/10.1080/01431161.2018.1452075>.
- Xiao, X., Boles, S., Froliking, S., Salas, W., Moore, I., Li, C., He, L., Zhao, R., 2002. Observation of flooding and rice transplanting of paddy rice fields at the site to landscape scales in China using VEGETATION sensor data. *Int. J. Rem. Sens.* 23, 3009–3022. <https://doi.org/10.1080/01431160110107734>.