Research Article

Spatial and Temporal Modeling of Land use/Land cover Change at the Ca River Basin (North Central Viet Nam) Using Markov Chain and Cellular Automata Approach

Bang Nguyen Thanh [1,2,3*], Phong Doan Ha [1]

[1] Hydro–Meteorology and Climate change Technology Research Division, Vietnam Institute of Meteorology, Hydrology and Climate change, No. 23, Nguyen Chi Thanh street, Dong Da district, Hanoi 100000, Vietnam
[2] Geographic Information System Group, Department of Business and IT, University College of Southeast Norway, Gullbringvegen 36, N–3800, Bø i Telemark, Norway
[3] Department of Health and Environmental Studies (INHM), University College of Southeast Norway, Gullbringvegen 36, N–3800, Bø i Telemark, Norway

*Corresponding author: bangnt.imhen@gmail.com; Tel.: +84–838734488

Abstract: Simulation of Land use/Land cover (LULC) change has been conducted extensively in the past with varying techniques and methodologies with Markov Chain incorporating Cellular Automata approach among those. The Markov–Cellular Automata (Markov_CA) model has been applied worldwide, however, model parameter calibration is site–specific. In Viet Nam, research on LULC change a pressing issue given the rapid socio–economic development. Research on LULC change is a necessary starting point for impacts assessment on water resources, land resources, ecosystems, environment, etc. However, what we lack is a method for modeling our insights to simulate LULC fluctuations and to project future LULC. Therefore, this article offers a way to combine known problems to produce a new result. The change of LULC for the period 2005–2015 will be simulated and will result in a prediction of the LULC of 2030. In addition, the calibrated Markov_CA model adapted to the study area will also be a valuable reference for employment in similar areas. Finally, the expected results and the calibrated model are validated by the Kappa coefficient and provide a good level of agreement.

Keywords: Land use/Land cover Change; Markov Chain; Cellular Automata; Ca River Basin; Viet Nam.

1. Introduction

Land change science has emerged as a fundamental component of global environmental change and sustainability research [1]. Land use/land cover (LULC) has interaction with soil, water resources, biodiversity, ecosystem, climate [2]. Changes in land use and land cover will consequently result in the changes of the latter. LULC changes are often caused by two influencing factors: anthropogenic and natural [3]. Human–induced land cover change such as deforestation has been a major contributor to increasing CO_2 concentration [4], the rapid expansion of agriculture reduce available freshwater given the intensive water use nature of agriculture (70% of total freshwater used by humankind), land exploitation disrupts the biotic function of soils [2]…. However, these changes can also be caused by natural factors, in particular vegetation cover [3]. Therefore, detecting and projecting LULC dynamics is necessary.

Recent advances in remote sensing data and growing advances in their temporal, spatial, and spectral resolutions provide useful data and tools for the detection of changes on LULC at different scales [5]. Remote sensing and geographical information system (GIS) together can provide an accurate depiction of changes in LULC [6], while remaining cost–effective [7].

A typical approach to simulate and predict LULC changes is to investigate the factors contributing to the land transitions and to provide a probabilistic prediction of where the changes may occur through modeling [5, 8]. There are various modelling approaches for the simulation and exploration of LULC changes [9–10]. According to [11] a set of 19 land–use models were reviewed in detail as representative of the broader set of models identified from the more comprehensive review of the literature. They included Markov models, Cellular Automata models, logistic regression models, econometric models, weights models, etc.

Markov model is stochastic modeling that uses evolution from "*t–1*" to "*t*" to project probabilities of changes for a future date "*t+1*" [12]. However, a stochastic Markov model is not appropriate because it does not consider spatial knowledge distribution within each category and transition probabilities are not constant among landscape states; so it may forecast the right magnitude of change but not the right direction [13]. This deficiency of the Markov model can be offset through the integration with other spatial component models [14]. Hence, the Cellular Automata Markov model combines the concepts of Markov Chain, Cellular Automata, Multi–Criteria Evaluation and Multi–Objective Land Allocation [8] is an interesting approach to modeling both spatial and temporal changes. [8] also determined that Cellular Automata Markov gave the approximate results to Multi–Layer Perceptron Markov [15] and outperformed Stochastic Markov in various validation techniques including: per category method, kappa statistics, components of agreement and disagreement, three map comparison, and fuzzy method.

The article aims to combine the scientific basis of the Markov–Cellular Automata method and the practice of the Ca River basin to find the influencing factors and model parameters to simulate the change of land cover in the Ca River basin. At the same time, the results of the article are the future land cover in 2030 of the Ca River basin will also supplement the data source – which is still lacking, to serve the planning and management of water resources in this area.

2. Study area and data used

2.1. Description of the study area

The Ca River system is a transnational basin originating from the upper 2000 m high mountain range in Xieng Khuang province of Laos, flowing northwest–southeast before entering the North Central of Viet Nam and pouring into the sea at the Hoi estuary. The Ca River system is located between 18º15'50" to 20º10'30" north latitude; and 103º45'20" to 105º15'20" east longitude. The outlet of the basin is at 18º45'27" north latitude; and 105º46'40" east longitude. The starting point of the Ca River system within Viet Nam is at 19º24'59" north latitude; and 104º04'12" east longitude [16].

The mainstream of the Ca River system is approximately 513 km in length, of which the length of the reach within Viet Nam's territory is 361 km long [17]. The mainstream flows through most parts of Nghe An Province, known as the Ca river. In Anh Son District, the river receives tributary is the Hieu River. Downstream of the Ca River is its confluent with the La river flowing from Ha Tinh Province. From this reach to the sea, the river is called the Lam river (Figure 1).

The basin area within Viet Nam is 17,730 km^2 in a total basin area of 27,200 km^2 [18]. Every year, the basin receives an average precipitation of 1100 ÷ 2500 mm. In the large rainfall centers such as upstream of Hieu, La Rivers, average annual rainfall could be as high

as 2000 ÷ 2400 mm. Land cover in the river basin composes of 44% forest, 16% paddy rice, 2% vegetable and crops, 38% others [19].
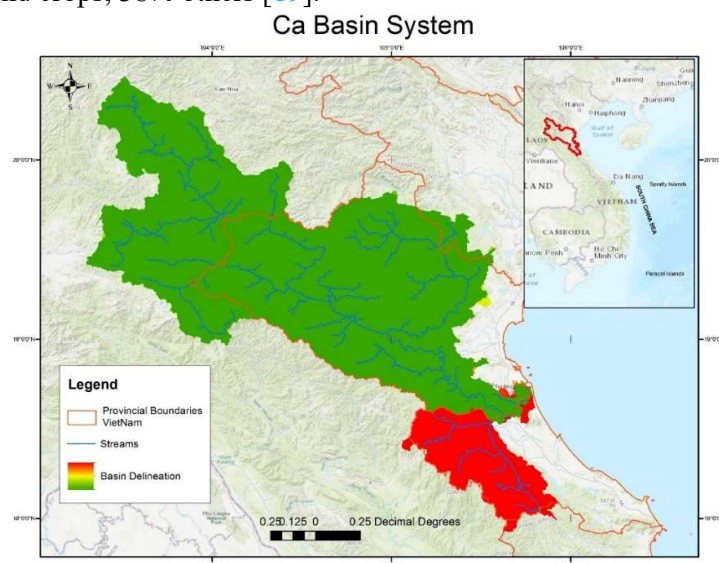

Figure 1. Ca River Basin.

2.2. Data used

2.2.1. Remote sensing data

In order to study how the landscape has changed over the 2005–2015 period, land cover maps of 2005–2010–2015 should be developed. The maps are based on LANDSAT image data from the United States Geological Survey (USGS). The collected images were LANDSAT 5 TM and LANDSAT 8 OLI/ TIRS with the same resolution of 30 m (Table 1), where the image of path 127 rows 47 was the largest covering, approximately 80% of the whole Ca river basin. Selected images based on the criteria: low cloud cover, no "scan line" error, the time of the image is not too far apart, especially the images 127–47.

The Landsat 5 TM images consist of six spectral bands with a spatial resolution of 30 meters for Bands 1–5 and 7. Landsat 8 OLI and TIRS images consist of nine spectral bands with a spatial resolution of 30 meters for Bands 1 to 7 and 9. The ultra–blue Band 1 is useful for coastal and aerosol studies. Band 9 is useful for cirrus cloud detection. The resolution for Band 8 (panchromatic) is 15 meters. Thermal bands 10 and 11 are useful in providing more accurate surface temperatures and are collected at 100 meters [20]. For LULC classification, images are collected as shown in (Table 1). High spatial resolution images from Google Earth and current land use status map published by the Department of Survey, Mapping and Geographic Information Viet Nam are used to validate the results.

Table 1. Description of the data sources and types used in this study.

Year	Data type and resolution	Path–row	Date	Source
2005	Landsat 5 TM	126–47	14th July 2005	https://earthexplorer.usgs.gov/
	30 m	126–48	09th April 2005	
		127–46	18th May 2005	
		127–47	18th May 2005	
		128–46	23rd April 2005	
		128–47	07th April 2005	

Year	Data type and resolution	Path–row	Date	Source
2010	Landsat 5 TM	126–47	12th July 2010	https://earthexplorer.usgs.gov/
	30 m	126–48	12th July 2010	
		127–46	08th Nov 2010	
		127–47	08th Nov 2010	
		128–46	21st April 2010	
		128–47	30th Oct 2010	
2015	Landsat 8 OLI/TIRS	126–47	11th Aug 2015	https://earthexplorer.usgs.gov/
	30 m	126–48	28th Sept 2015	
		127–46	30th May 2015	
		127–47	30th May 2015	
		128–46	28th Oct 2015	
		128–47	28th Oct 2015	

2.2.2. Ancillary data and field data (GPS)

a) Digital Elevation Model–DEM

ASTER Global Digital Elevation Model 2.0 data (GDEM 2.0) is a product of the Ministry of Economy, Trade, and Industry (METI) and National Aeronautics and Space Administration (NASA) collected from the US Geological Survey (USGS). GDEM 2.0 was announced by METI and NASA in mid–October 2011, inheriting almost all the features of GDEM 1.0 with a resolution of 30 m, covering from latitude 83º North to 83º South. But GDEM 2.0 has a higher horizontal resolution by using a 5×5 correlation kernel instead of 9×9 as used for GDEM 1.0. GDEM 2.0 has a total accuracy of 17 m compared to 20 m of GDEM 1.0 along with a 95% certainty [21] (ASTER–GDEM, October 2011).

The DEM data for the study area were collected from latitude 18º to 19º North, longitude 103º to 105º East. The ArcSWAT tool is then used to calculate flow direction, accumulate flow, create sub–basin area, create flow net, discharge outlet, etc [22].

b) Current land use status map

Beginning in 1999, under Directive 24/1999/CT–TTg of the Prime Minister of Viet Nam on land inventory is issued in 2000. Since then, the inventory and mapping of current land use status have been performed in 2005, 2010, 2015 (Figure 2). They are valuable ancillary references for LULC classification.

The current land use status map is a map showing the distribution of land categories according to the regulations on an inventory of land use purposes at the time of land inventory. The current land use status map is drawn up on the basis of the cadastral map, in comparison with the field data and land inventory data; In case no cadastral map is available, use aerial photographs or high–resolution satellite images converted into orthogonal photographs combined with field data and land inventories to make the current land use status map; In case there are no such maps, current land use status map of the previous period is used and also will be checked with field data and land inventory data.

The land cover based on use purpose includes agriculture production, forestry, aquaculture, salt production, other agricultural, built–up, specialized (eg.: State's office, defence, and security, transport, medical, education, etc.), rivers and water surfaces, bare and unused land. They can be regrouped into 5 classes: Forest, Agriculture, Built–up, Waterbody, and Bare area (Figure 2).
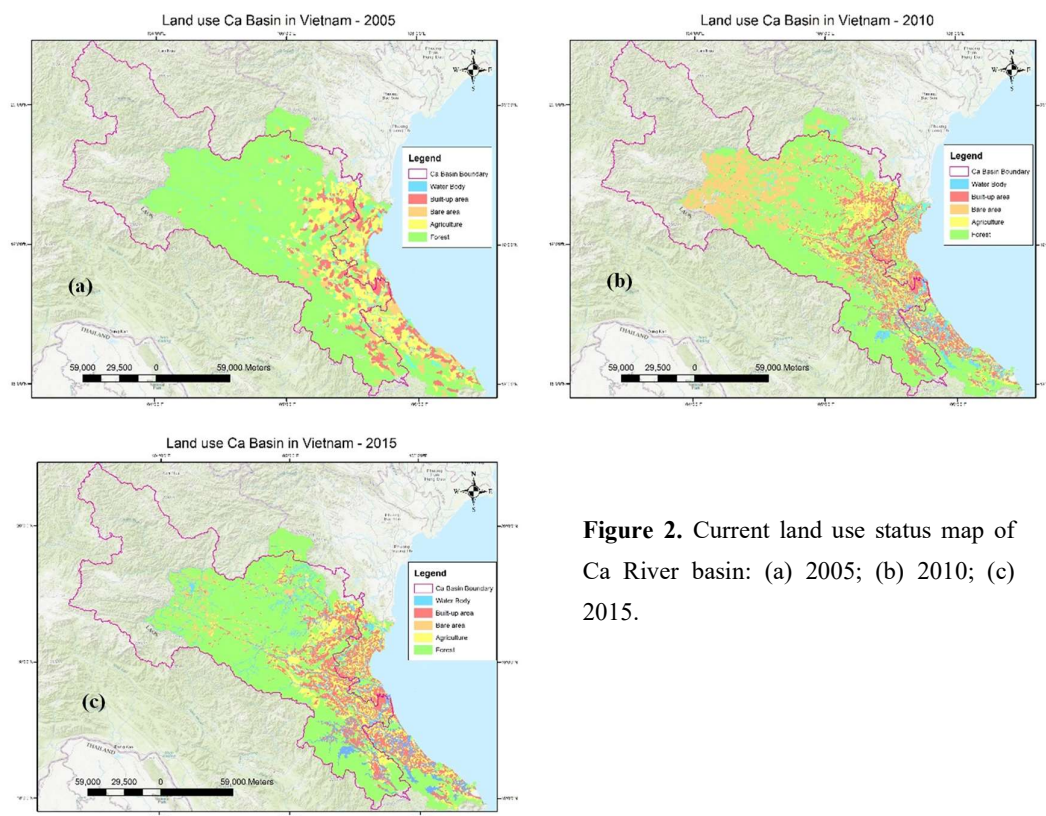
Figure 2. Current land use status map of Ca River basin: (a) 2005; (b) 2010; (c) 2015.

c) Field data

Ground data is collected for classification and verification of classification results. The total number of samples acquired is 120 samples. Because the latter works related to three periods, all the ground data were consolidated; care was taken to ensure that areas that had undergone a change (e.g., burn regeneration) were excluded from the investigation. Through consultation with local people, five classes were sampled – Forest, Agriculture, Built–up, Waterbody, Bare area – with about 20 ground data for each class. Some test data is additionally collected using Google Maps by random points algorithm.
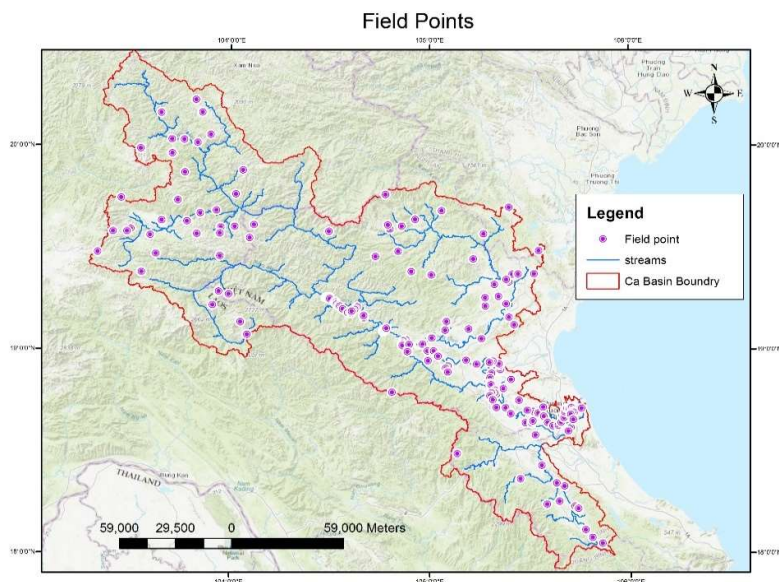

Figure 3. Location of ground data.

3. Theoretical background of the method used

3.1. Maximum Likelihood classification

The classification method used is the Maximum Likelihood Classification (MLC), which is one of the methods of the Supervised Classification [7, 23]. This method is based on a given set of sample pixels and hence identifies pixels with the same spectral characteristics. Next, the estimated (Gaussian) probability density function is used to identify other pixels of the same land use/land cover [24]. The MLC principle also can be found in Foody and Strahler's researches [24–25]. This is a commonly used method of image classification and provides relatively high classification accuracy.

In this study, the Landsat data were classified with the maximum likelihood decision rule and some ancillary data (e.g., DEM, land use data, vegetation index, and textural analysis of the Landsat images) were combined through an expert (or hypothesis testing) system to improve the classification accuracy [26]. Considering the spectral characteristics of the satellite images and existing knowledge of land use/land cover of the study area, five LULC categories were identified and classified for 2005, 2010, and 2015 (Table 2).

Table 2. LULC categories distributed for the classification (Circular 08/2007/TT–BTNMT).

LULC category	Description
Forest	Land with natural forest or planted forests meeting the forest standards (e.g., production forests, protection forests, and special–use forests).
Agriculture	Land for agriculture production including Land for planting annual crops (e.g., paddy land, grassland used for breeding, other annual crops); Land for perennial crops (e.g., orchards, perennial crops).
Built–up	Land for construction of dwelling houses, construction of works, land for offices of agencies and non–business works; land protection, security.
Waterbody	Land for rivers and streams and specialized water surfaces, coastal water surface.
Bare area	Land with no purpose of use including unused plain land, unused hill or mountain land, Rocky Mountains without forests.

3.2. Markov Chain

Markov process is a special random moving from one state to another state at each time step via the use of transition probability matrices [14, 27]. The transition probability matrix is calculated by assuming that probability distribution over the next state only depends on the current state, but not on previous ones [10]. In this study, a probability matrix based on the likelihood of the LULC variations between 2005, 2010, and 2015 was used to predict the LULC map in 2030. The transition matrix can be presented as follows [14]:

$$P = (P_{ij}) = \begin{vmatrix} P_{11} P_{12} \dots P_{1n} \\ P_{21} P_{22} \dots P_{2n} \\ \dots \\ P_{n1} P_{n2} \dots P_{nn} \end{vmatrix} 0 \le P_{ij} \le 1 \quad \sum_{i=1}^{n} P_{ij} = 1 \qquad (1)$$

where P is the transition probability matrix, P_{ij} is the probability of the ith LULC changing to j[th] LULC from initial year to illation year and n is the number of LULC classes.

The Markov chain model is very powerful to determine the possibility of land–use change between two time periods. However, the Markov chain model cannot provide the spatial distribution of occurrences of land–use change [28].

3.3. Cellular automata

Generally, Cellular automata (CA) models aim to simulate the real nature regulations. Land–use change modeling using the CA technique is one of the preferred methods because it gives explicit spatial modeling results based on defined transition rules [23]. A CA consists of discrete cell space, in which states characterize every cell. [29] define a simple CA to include the following components: (1) a grid space L on which the model operates, (2) cell states Q in the grid space, (3) transition rules f, which determine the spatial dynamic process, (4) status of the neighborhood Δ that influences the central cell. Hence, the spatiotemporal changes of state in a system can be described as [30]:

$$A = [L, Q, Δ, f] \tag{2}$$

Each Q cell of L grid space will change their state in discrete time steps. The state of a cell Q depends on its neighborhood Δ (the surrounding cells) and the corresponding f transition rules. However, the most important concern in the CA model is defining appropriate transition rules f based on training data that controls the model [31].

3.4. Accuracy assessment

There are many accuracy assessment methods that have been discussed in the remote sensing context e.g., [32–35], but the most widely proposed and used method is confusion matrix or error matrix. A measurement termed "percentage of cases correctly allocated" derived from a confusion matrix has been used to measure classification accuracy [36]. The accuracy of the individual class may be derived from the matrix by relating the number of cases correctly allocated to the class to the total number of cases of that class (Figure 3). This leads to two concepts: user's accuracy and producer's accuracy. The user's accuracy provides the user information on the accuracy of the LCLU data against actual data. Producer's accuracy indicates the percentage of samples of a certain (reference) class that were correctly classified [37]. These accuracies are calculated based upon the confusion matrix's row or column [38].

Actual Class

		A	B	C	D	Σ
	A	n_{AA}	n_{AB}	n_{AC}	n_{AD}	n_{A+}
	B	n_{BA}	n_{BB}	n_{BC}	n_{BD}	n_{B+}
Predicted Class	C	n_{CA}	n_{CB}	n_{CC}	n_{CD}	n_{C+}
	D	n_{DA}	n_{DB}	n_{DC}	n_{DD}	n_{D+}
	Σ	n_{+A}	n_{+B}	n_{+C}	n_{+D}	n

$$\text{User's accuracy} = \frac{n_{ii}}{n_{i+}} \tag{3}$$

$$\text{Producer's accuracy} = \frac{n_{ii}}{n_{+i}} \tag{4}$$

Figure 4. Error matrix with n_{ij} representing the proportion of predicted class i and the actual class j.

On the issue of the chance of agreement, Cohen's kappa coefficient has been used and be adopted as a standard measure of classification accuracy [39]. Kappa takes the chance

agreement into account and Kappa adjusts the percentage correct measure by subtracting the estimated contribution of the chance agreement [40]. The definition of Kappa (k) is:

$$k = \frac{p_0 - p_e}{1 - p_e} \qquad (5)$$

where p_0 is the observed proportion correct, p_e is the expected proportion correct due to change.

4. Proposed methodology for Spatial and Temporal Modeling of Land Cover Change at the Ca River Basin (North Central Viet Nam)

This study employed an integrated Markov – Cellular Automata (Markov–CA) model to predict LULC changes for the Ca River basin in the target years 2030. Data preprocessing and format unification were achieved using GIS, which provides numerous functions for visualizing and analyzing the data [41]. Markov–CA model is applied by TerrSet, developed by Clark Labs at Clark University, is an integrated geospatial software with the ability to incorporate the IDRISI GIS analysis for monitoring and modeling purposes [42]. In general, the flow chart of the methodology is summarized in (Figure 5).
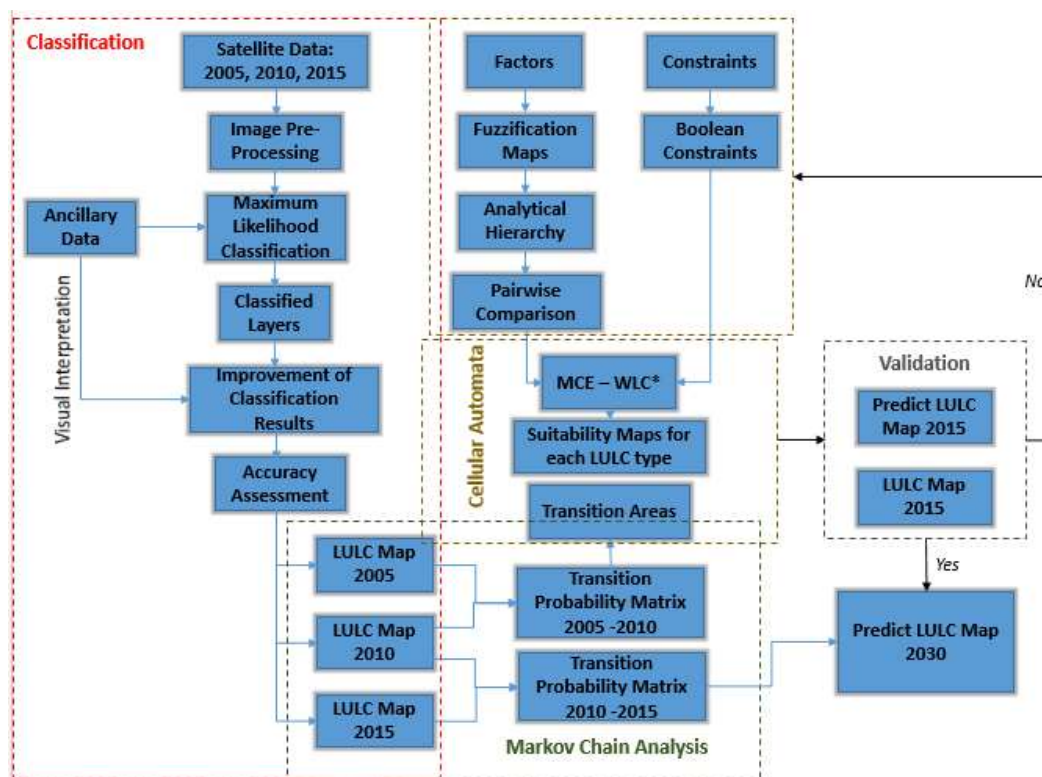

Figure 5. Workflow showing the methodology in the study; *MCE–WLC: Multi–criteria evaluation–Weighted linear combination.

In the first phase, the Landsat images are classified and LULC layers are prepared. In the second phase, the Transition Probability Matrix and Transition Areas are calculated with Markov Chain Analysis. At the same time, factors and constraints are set up for each land–use class and fuzzy functions are applied for each factor and assigned Boolean values (0 or 1) for constraints. Then Analytical Hierarchy Process and Pairwise Comparison are used to assign the weight of each factor. Weighted factors and Boolean constraints are used in the MCE–WLC function to generate suitability maps for each LULC type. In the third phase, all previous components are thrown into the Cellular Automata module and output to the projected LULC map of the next period (2015). In the validation phase, the projected LULC

map is compared to the LULC map on the agreement level by the Kappa coefficient. If the validation results indicate that a bad agreement, the Weighted factors, and Boolean constraints will be reconsidered. Otherwise, the model is ready to predict LULC maps in the future.

5. Result

5.1. Image processing and classification

For an accurate assessment of LULC between 2005 and 2015, atmospherically–corrected surface reflectance Landsat 5 TM and Landsat 8 Operational Land Imager (OLI) and Thermal Infrared Sensor (TIRS) images were collected from the United States Geological Survey (USGS) website. All scenes were verified for geometric accuracy and all data were projected on WGS 1984, UTM zone 48N.

Images were stacked, subset, and analyzed in ENVI, ArcGIS software, and classified using the maximum likelihood algorithm. Supervised approaches using a maximum likelihood classifier algorithm were applied for the extraction of LULC. A modified land–cover classification system was used for remote sensing data as recommended by [43] and 5 classes were identified: built–up, forest, agriculture, waterbody, bare area (Table 2).

Segmentation provides an approach to extracting features from imagery based on objects. These objects are created via an image segmentation process where pixels in close proximity and having similar spectral characteristics are grouped into a segment. Segments exhibiting certain shapes, spectral, and spatial characteristics can be further grouped into objects – meaningful object–oriented feature class. The result is a grouping of image pixels into a segment characterized by an average color.

A supervised classification requires collecting training samples as the basis for the maximum likelihood algorithm speculating other pixels of the same class. The more accurate the data collected, the more accurate the classification. Bands composite method is widely used in remote sensing to support that. Each composite has its advantages in classifying LULC. A composite image using bands 4, 3, 2 in Landsat 5 TM images or bands 7, 6, 5 in Landsat 8 for the red, green, blue channels, respectively will be easier to detect roads, water–body, and agriculture class. Other composite images are also used to expose other land classes e.g. bands 5, 4, 3 or 4, 5, 3 in Landsat 5 TM or 6, 5, 4 in Landsat 8 for forest classification. The results are presented in (Figure 6).
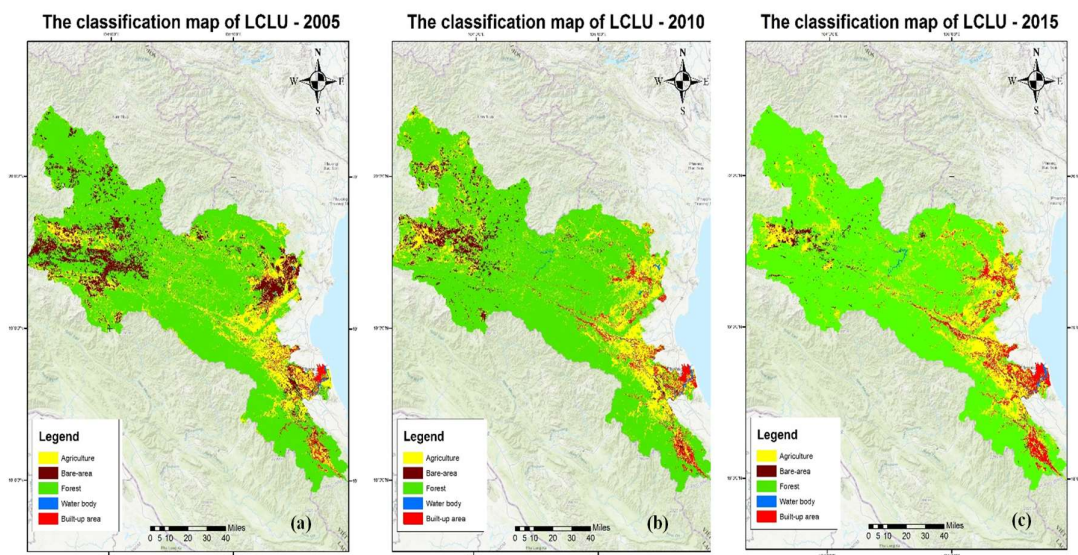

Figure 6. The classification map of LULC (a) 2005, (b) 2010, (c) 2015.

5.2. Landuse/landcover classification and analysis

Accuracy assessment has been used to evaluate the accuracy of classified data. 175 test samples were selected, of which 90 samples were GPS points collected in the field, the rest were randomly selected points from Google Maps. The study calculated and evaluated PA, UA, CA, and kappa index for classification data of 2005, 2010, and 2015. The results are shown in Table 3.

Table 3. Accuracy assessment of the Land use/Land cover classification using the validation dataset
PA: Producers Accuracy; UA: Users Accuracy; CA: Classification Accuracy.

Land use/Land cover class	2005		2010		2015	
	PA (%)	UA (%)	PA (%)	UA (%)	PA (%)	UA (%)
Agriculture	68.40	74.28	77.77	80.00	83.33	85.71
Bare area	74.19	65.71	84.85	80.00	86.11	88.57
Forest	78.38	82.86	80.00	91.43	91.67	94.29
Waterbody	87.88	82.86	94.44	97.14	100	100
Built–up	77.78	80.00	93.33	80.00	93.75	85.71
Overall CA (%)	77.14		85.71		90.86	
Kappa index	0.7143		0.8214		0.8857	

Accordingly, the PA and UA of each class are greater than 75% for the classification data of 2010 and 2015. Particularly for 2005, the PA of agriculture class is only 68.40%, the UA of bare area class is only 65.71%, however, PA and UA of other classes still reach over 75% for all 3 years. Overall Classification Accuracy of classification data for 2005, 2010 and 2015 is 77.14%, 85.71%, 90.86%, respectively. The Kappa index for 2005 data is 0.71, 2010 is 0.82, 2015 is 0.88. The CA and Kappa index of 2005 was relatively lower than other years can be explained by the fact that test data was collected in 2018, so there are certain differences compared to 2005.

5.3. Generation of Transition Probability Matrix (TPM), Transition Area Matrix (TAM)

a) TPM

A transition probabilities matrix determines the likelihood that a cell or pixel will move from a land–use category or class to every other category from date 1 to date 2. This matrix is the result of cross–tabulation of the two images adjusted by the proportional error and is translated in a set of probability images, one for each land–use class [12]. As mentioned above, TerrSet software is one of the best platforms to conduct CA–Markov model, which is developed by Clark Labs in the U.S. Hence, transition probabilities matrix are built from the land–use/land–cover images of 2005–2010 and 2010–2015 by ArcGIS and TerrSet software (Table 4).

Table 4. Transition probability matrix of 2005–2010 and 2010–2015 periods (%).

		Agriculture	Bare Area	Forest	Water Body	Built–up
		(%)	(%)	(%)	(%)	(%)
2005–2010	Agriculture	42.98	7.14	38.58	1.76	9.54
	Bare Area	35.74	17.35	43.15	0.39	3.37
	Forest	22.53	5.97	70.52	0.37	0.61
	Water Body	24.07	0.63	7.92	63.22	4.15
	Built–up	18.75	3.15	5.71	1.96	70.43
2010–2015	Agriculture	63.89	4.26	21.69	1.20	8.96

	Agriculture (%)	Bare Area (%)	Forest (%)	Water Body (%)	Built–up (%)
Bare Area	25.76	66.40	4.70	0.38	2.75
Forest	24.81	3.34	70.76	0.47	0.63
Water Body	15.00	0	0	85.00	0
Built–up	0	0	0.01	15.97	84.02

Table 4 shows that the transition probability of agriculture and bare–area is higher than forest, built–up, and water–body in the 2005–2010 period. Built–up and water body has only little probability to change to another type of land cover, about 15% in the 2010–2015 period.

b) TAM

A transition area matrix that records the number of cells or pixels that are expected to change from each land–use class to each other land–use class over the next period. This matrix is produced by the multiplication of each column in the transition probability matrix by the number of cells of corresponding land use in the later image [12]. Overlapping of land cover maps in 2005–2010 and 2010–2015 (Figure 6). Set the time interval between two maps to be five years, the proportional error to 0.15 in case of Maximum Likelihood Classification. The transition area matrix is presented in (Table 5).

Table 5. Transition area matrix of 2005–2010 and 2010–2015 periods (pixel).

		Agriculture (pixel)	Bare Area (pixel)	Forest (pixel)	Water Body (pixel)	Built–up (pixel)
2005–2010	Agriculture	3,511,522	583,006	3,151,858	144,081	779,401
	Bare Area	744,148	361,252	898,504	8,171	70,201
	Forest	4,778,388	1,265,217	14,955,883	78,376	129,558
	Water Body	95,896	2,527	31,554	251,882	16,539
	Built–up	309,583	51,943	94,306	32,381	1,162,909
2010–2015	Agriculture	5,974,382	398,421	2,028,508	112,114	837,973
	Bare Area	579,574	1,494,112	105,851	8,601	61,880
	Forest	4,479,663	602,590	12,778,159	84,471	114,425
	Water Body	81,596	0	0	462,369	0
	Built–up	19	56	169	358,662	1,886,814

Table 5 shows a clearer view of how many pixels have changed from a class to another class. Specifically, the number of pixels in the forest class has the greatest change in both phases, followed immediately by the agriculture class. The class that has the least change is the water body.

5.4. Suitability Map

Suitability maps present the probability of suitability of a pixel belonging to the corresponding LULC class. They range from 0 to 255 with 255 being the most likely and 0 is unlikely. Each suitability map is created by transition rules that are formed by the linkage between socioeconomic, ecological, and spatial variations (e.g. built–up tends to develop near the road). Besides, there are also restrictions on each type of LULC class (e.g. forest areas are planned for conservation). Therefore, factors and constraints are two driving forces of change that determine which lands to be considered for further development.

In this study, slope, digital elevation model (DEM), distance to water bodies, distance to main roads were set as driving factors, but also there were constraints considered (e.g. water and built–up represented constraints for transition to bare–area). They were chosen because of the similarity in their use in many previous studies such as [5, 12, 15, 28, 44–47] and the author's knowledge of the study area. The constraints and factors were standardized into a Boolean (0 and 1) character and a continuous scale of suitability from 0 (least suitable) to 255 (most suitable), respectively. To do that, three types of fuzzy membership function (linear, sigmoidal, and J–shaped) and control points were determined as a necessity to measure the scale of potential suitability for each class (Table 6). Selection of the type of fuzzy membership function and control points is prone to subjectivity and can change according to the knowledge of decision–makers [15]. A fuzzy set theory can be found at [48–50].

Table 6. Standardization of factors by Fuzzy module.

Class	Factors	Functions	Control Points
			0 degree highest suitability
	Slope	J–shaped	0–20 degree decreasing suitability
			>20 degree no suitability
			0 m highest suitability
			0–350 m decreasing suitability (Agriculture)
	DEM	J–shaped	0–150 m decreasing suitability (Built–up)
Agriculture			> 350 m no suitability (Agriculture)
Built–up			> 150 m no suitability (Built–up)
			< 1.5 km highest suitability
	Distance to rivers	Sigmoidal	1.5–5.5 km decreasing suitability
			> 5.5 km no suitability
			< 0.2 km highest suitability
	Distance to main roads	J–shaped	0.2–5 km decreasing suitability
			> 5 km no suitability
			0 degree highest suitability
	Slope	J–shaped	0–15 degree decreasing suitability
			> 15 degree no suitability
			0 m highest suitability
Waterbody	DEM	J–shaped	0–300 m decreasing suitability
			> 300 m no suitability
			< 1 km highest suitability
	Distance to rivers	Sigmoidal	1–5 km decreasing suitability
			> 5 km no suitability
			< 5 degree no suitability (Forest)
			< 20 degree no suitability (Bare)
Forest			5–18 degree increasing suitability (Forest)
Bare area	Slope	Sigmoidal	20–40 degree increasing suitability (Bare)
			> 18 degree highest suitability (Forest)
			> 40 degree highest suitability (Bare)

Class	Factors	Functions	Control Points
	DEM	Sigmoidal	< 150 m no suitability (Forest)
			< 1300 m no suitability (Bare)
			150–700 m increasing suitability (Forest)
			1300–1700 m increasing suitability (Bare)
			> 700 m highest suitability (Forest)
			> 1700 m no suitability (Bare)
	Distance to main roads	Sigmoidal	< 1 km no suitability
			1–10 km increasing suitability
			> 10 km highest suitability

Analytical hierarchy process and pairwise comparison were then applied to develop a set of relative weights for a group of factors in a multi–criteria evaluation. The weights were developed by providing a series of pairwise comparisons of the relative importance of factors to the suitability of pixels for the activity being evaluated. These pairwise comparisons were then analyzed to produce a set of weights that sum to 1 [42]. The procedure by which the weights are produced follows the logic developed by [51–52]. The larger weight denoted a more important criterion in terms of overall factors (Table 7).

Table 7. Factors and their weights used in the construction of suitability maps.

Factors	Forest	Agriculture	Built–up	Water body	Bare area
Slope	0.5917	0.1740	0.5232	0.3874	0.1571
DEM	0.3332	0.2696	0.2976	0.1692	0.2493
Distance to main roads	0.0751	0.0795	0.1222		0.5936
Distance to rivers		0.4768	0.0570	0.4434	
Consistency ratio	0.01	0.02	0.03	0.02	0.05

Then, the Multi–Criteria Evaluation (MCE) module was used to make decisions which is a choice between alternatives. In an MCE, an attempt is made to combine a set of criteria to achieve a single composite basis for a decision according to a specific objective. Through a Multi–Criteria Evaluation, these criteria images representing suitability may be combined to form a single suitability map from which the final choice will be made [42]. Weighted linear combination (WLC) methods are used to include both weighted factors and constraints by the logical AND operation. The intersection of all the criteria leads to obtaining suitable areas for a specific LULC class (Figure 7).

5.5. Simulation of the land use/land cover

The integrated model of Cellular Automata and Markov models can predict LULC changes based on two–time intervals. Therefore, the transition probabilities for the period 2005–2010 along with the basis LULC 2010 were used to simulate LULC in 2015. Each pixel of each LULC type was attributed future suitability by the suitability map for each LULC class. In addition, a standard 5×5 boolean mask filter was used to analyze the neighborhood definition that the suitability weight of the pixels will decrease far from the existing areas and allocate preference to the neighboring suitable areas [41]. The number of iterations i is also the number of time steps that will be used in the model. Choosing this number is also one of the elements that will influence the model's expected results. Indeed, to reach the optimal parameters, the number of iterations was examined in [53]. In the case of this study, the number of iterations was tested as 5, 10, and 20 (Figure 8).
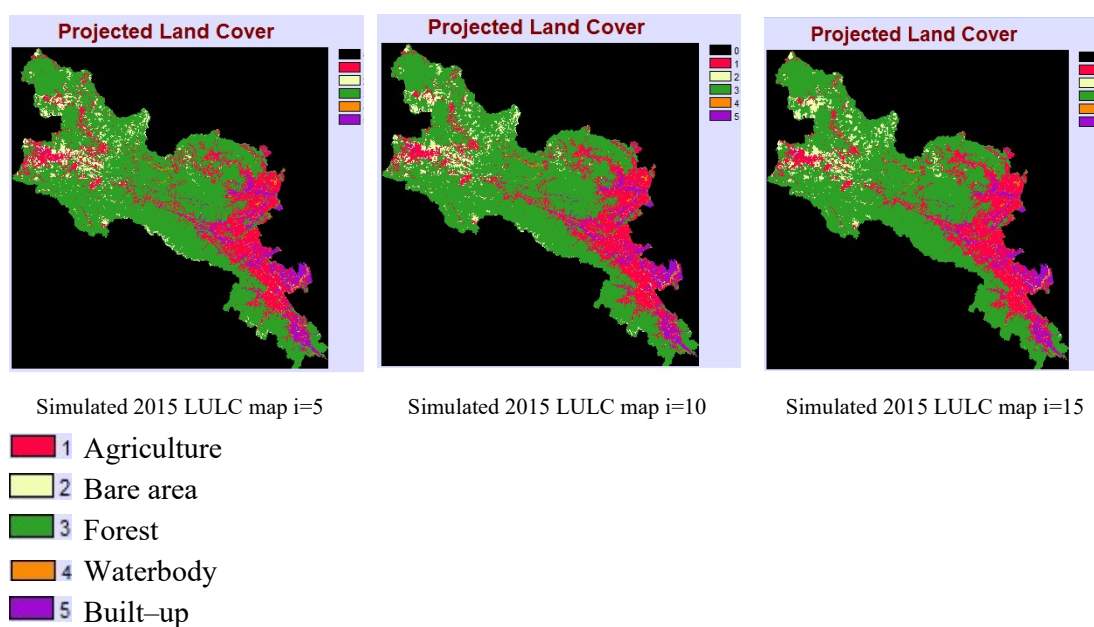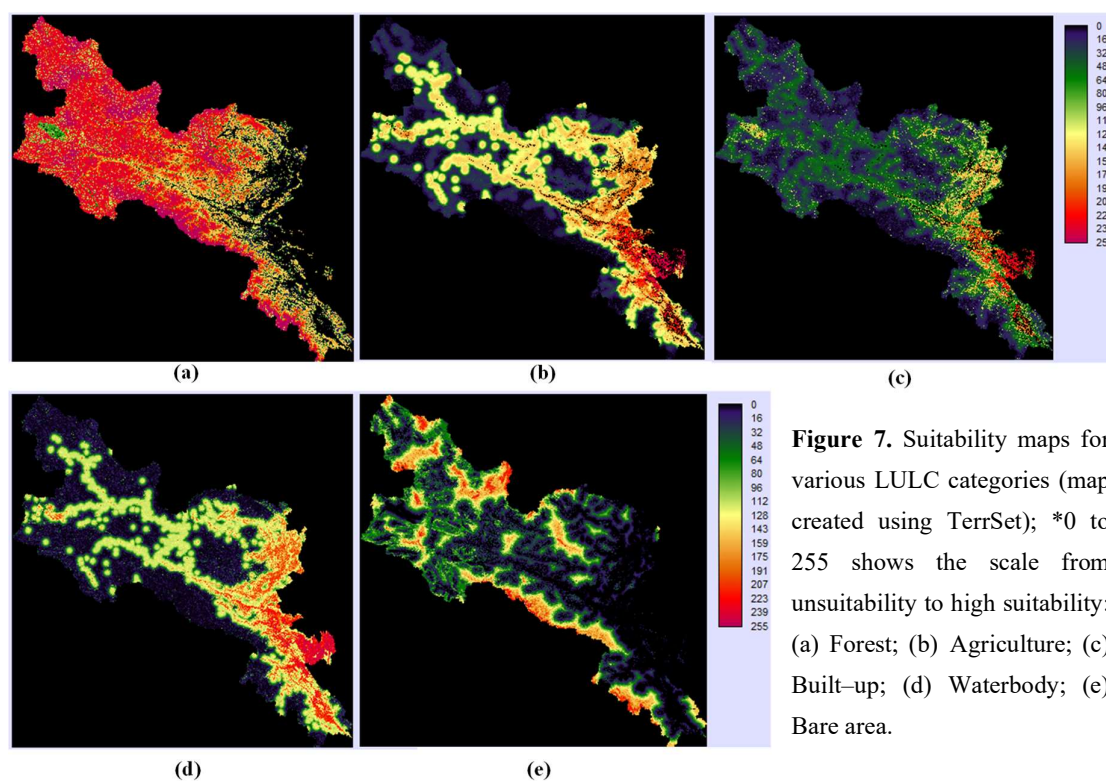
Figure 7. Suitability maps for various LULC categories (map created using TerrSet); *0 to 255 shows the scale from unsuitability to high suitability: (a) Forest; (b) Agriculture; (c) Built–up; (d) Waterbody; (e) Bare area.

Simulated 2015 LULC map i=5   Simulated 2015 LULC map i=10   Simulated 2015 LULC map i=15

1 Agriculture
2 Bare area
3 Forest
4 Waterbody
5 Built–up

Figure 8. Simulated LULC map of 2015 at different number of iterations through CA–Markov.

5.6. Model validation

Model validation is always an important part to verify and evaluate the accuracy of a model. Nevertheless, there are no consolidation criteria for assessing the feasibility of land change models [54]. To quantify the proficiency of the model, we need to compare the predicted result of the model with a similar and reliable map using the Kappa coefficient [55]. But [54, 56] proved that standard Kappa (Cohen's Kappa) offers almost no useful information because it confounds quantification error with location error. Hence, in addition to Kappa standard ($K_{standard}$), different components of the Kappa index including the Kappa

for no information (K_{no}), Kappa for grid–cell level location ($K_{location}$), and Kappa for stratum–level location ($K_{locationStrata}$) were used to supplement the deficiencies [54]. In short, the simulated 2015 LULC map was validated with the classified 2015 map, and the results are shown below (Table 8).

Table 8. Summary of Kappa statistics for the models on validation data (2015).

Kappa Indices	2015 LULC (i = 5)	2015 LULC (i = 10)	2015 LULC (i = 20)
K_{no}	0.9507	0.9349	0.9119
$K_{location}$	0.9178	0.8887	0.8451
$K_{locationStrata}$	0.9178	0.8887	0.8451
$K_{standard}$	0.9156	0.8865	0.8420

[57] claimed that associations between two variables that both rely on coding schemes with K < 0.7 is often impossible and said that content analysis researchers generally think of K > 0.8 as good reliability, with 0.67 < K < 0.8 allowing tentative conclusions to be drawn. Therefore, the simulation provided valid results, then the calibrated model could be applied for the prediction of future patterns – 2030 LULC map.

5.7. Future land use/land cover modeling

After calibration, the CA–Markov model has proven its viability in performing future LULC simulations, 2030. Therefore, this model has continued to be used with parameters that have been demonstrated to be accurate in study area conditions – the Ca river basin – as factors along with its weights and constraints, number of iterations, etc. However, the difference was the input data: 1) satellite–derived LULC maps for 2010–2015 were used to project the LULC for 2030; 2) constituents that generated factors such as main roads, rivers were updated until 2015. The predicted LULC map of 2030 is shown in Figure 9.
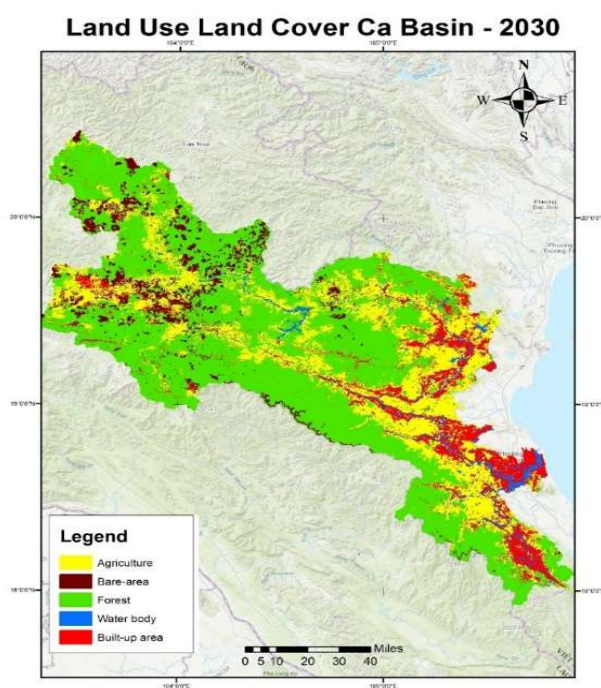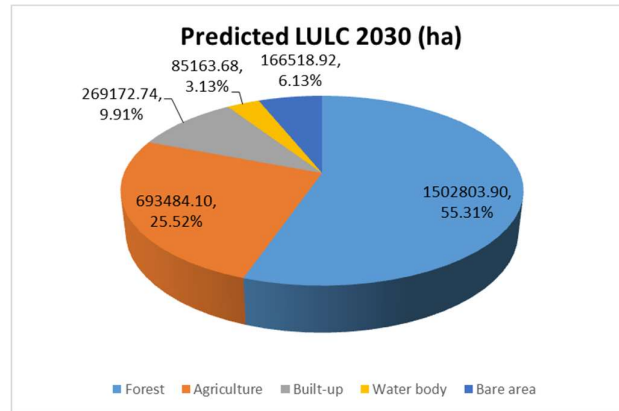

Figure 9. Predicted LULC map of 2030.

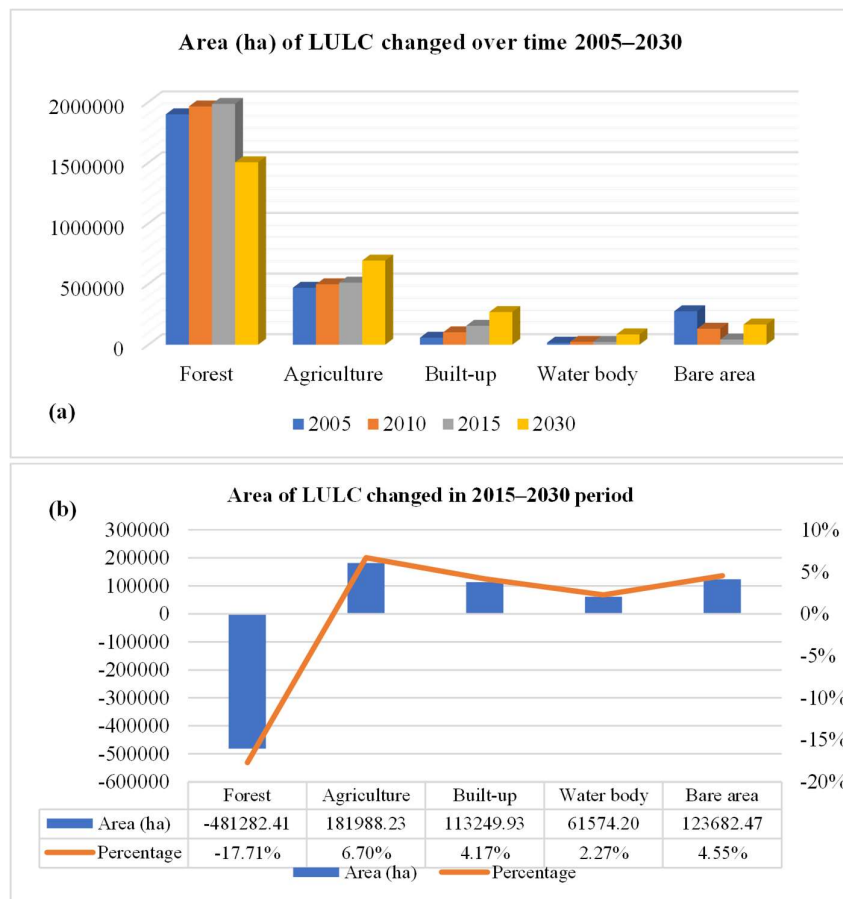**Figure 10.** Area (ha) and percentage distribution (%) for each LULC class in 2030.

Figure 11. The predicted amount of change from a) 2005 to 2030; b) 2015 to 2030.

Figure 9 demonstrates the spatial distribution of each LULC class in 2030 that predicted by the CA_Markov model. Figure 10 is a quantitative figure for the number of hectares of each type of land: forest, agriculture, built–up, waterbody, bare area corresponding to 55%, 26%, 10%, 3%, 6% of the total land in 2030. We can also easily see the increase of bare areas in the Western uplands by 2030 compared to previous years. In addition, forest areas are projected to decline sharply, especially in areas close to agricultural land and built–up land. Agricultural land in 2030 increased compared to 2005 (Figure 11a). Specifically, in Figure

11b, bare land increased by 4.55%, waterbody increased by 2.27%, built–up land increased by 4.17%, agricultural land increased by 6.70% between 2015 and 2030. The area of forest land has increased slightly in the period of 2005–2010, but in the years after that figure kept going down, especially from 2015 to 2030, the area of forest land decreased by 17.71%. In total, hectares were projected to change from 2015 to 2030 were 961777.24 ha, equivalent to 35.39% of the total studied land area.

6. Conclusion

This study demonstrates the feasibility of the Markov Chain and Cellular Automata approach for modeling the LULC in the Ca river basin, Viet Nam. The research process also clarified the model test and evaluation options, resulting in a calibrated model with appropriate parameters for the conditions of the study area. Validation results with Kappa coefficients of $K_{no} = 0.95$, $K_{location} = 0.91$, $K_{locationStrata} = 0.91$, $K_{standard} = 0.91$ showing strong agreement between satellite–derived and simulated LULC maps also denote that the model has good reliability.

The results of this study reveal that LULC in the Ca river basin has been, and will continue to change. Specifically, by 2030 the area of forest land will be reduced 17.71% and transformed into other types of land such as agricultural land 6.70%, construction 4.17%, and vacant land 4.55%. This also accurately reflects current socio–economic development trends: urbanization, agricultural land expansion, deforestation, etc. The change will increase the pressure on other natural factors such as soil status, water resources in this area. This still requires further research, but knowing the specific quantitative numbers of the area and type of LULC that will change will greatly assist in the process of assessing the impacts of LULC on other natural factors.

Author contribution: The authors confirm contribution to the paper as follows: study conception and design: B.N.T.; data collection: B.N.T.; analysis and interpretation of results: B.N.T.; draft manuscript preparation: B.N.T., P.D.H. All authors reviewed the results and approved the final version of the manuscript.

Conflict of interest: The authors declare that there are no conflicts of interest.

References
1. Turner, B.L.; Lambin, E.F.; Reenberg, A. The emergence of land change science for global environmental change and sustainability. *Proceedings of the National Academy of Sciences*, **2007**, *104(52)*, 20666–20671.
2. Sciences, Committee on Grand Challenges in Environmental Sciences. Grand challenges in environmental sciences. *National Academy Press*, **2001**.
3. Lambin, E.F. Modeling and monitoring land–cover change processes in tropical regions. *Prog. Phys. Geogr.* **1997**, *21(3)*, 375–393.
4. Kabat, P. The role of biospheric feedbacks in the hydrological cycle; the IGBP–BAHC special issue. *Global Change Newsletter* **1999**, *39*, 1–3.
5. Halmy, M.W.A.; Gessler, P.E.; Hicke, J.A.; Salem, B.B. Land use/land cover change detection and prediction in the north–western coastal desert of Egypt using Markov–CA. *Appl. Geogr.* **2015**, *63*, 101–112.

6. Corner, R.J.; Dewan, A.M.; Chakma, S. Monitoring and prediction of land–use and land–cover (LULC) change. Change. In: Dewan A., Corner R. (eds) Dhaka Megacity. Springer Geography. Springer, Dordrecht, **2014**, 75–97.

7. Rogan, J.; Chen, D. Remote sensing technology for mapping and monitoring land–cover and land–use change. *Prog. Plann.* **2004**, *61(4)*, 301–325.

8. Ahmed, B.; Ahmed, R.; Zhu, X. Evaluation of model validation techniques in land cover dynamics. *ISPRS Int. J. Geo-Inf.* **2013**, *2(3)*, 577–597.

9. Overmars, K.d.; De Koning, G.; Veldkamp, A. Spatial autocorrelation in multi–scale land–use models. *Ecol. Modell.* **2003**, *164(2–3)*, 257–270.

10. Veldkamp, A.; Lambin, E. F. Predicting land–use change. In: *Elsevier,* **2001**.

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

12. Houet, T.; Hubert–Moy, L. Modeling and projecting land–use and land–cover changes with Cellular Automaton in considering landscape trajectories. *EARSeL eProceedings* **2006**, *5(1)*, 63–76.

13. D BEHERA, M.; Borate, S.N.; Panda, S.N.; Behera, P.R.; Roy, P.S. Modeling and analyzing the watershed dynamics using Cellular Automata (CA)–Markov model–A geo–information–based approach. *J. Earth Syst. Sci.* **2012**, *121(4)*, 1011–1024.

14. Guan, D.; Li, H.; Inohae, T.; Su, W.; Nagaie, T.; Hokao, K. Modeling urban land–use change by the integration of cellular automaton and Markov model. *Ecol. Modell.* **2011**, *222(20–22)*, 3761–3772.

15. Ozturk, D. Urban growth simulation of Atakum (Samsun, Turkey) using cellular automata–Markov chain and multi–layer perceptron–Markov chain models. *Remote Sen.* **2015**, *7(5)*, 5918–5950.

16. Tung, H.T.; Cat, V.M.; Ranzi, R.; Hoa, T.T. Research on medium-term flood forecasting in Ca River basin. *Journal of Water Resources and Environmental Engineering*, **2010**, 28. (In Vietnamese)

17. Ly, N.T.K. Flood Characteristics of Lam River basin. *University of Science - Viet Nam National University*, 2017. In Vietnamese.

18. Giang, P.T. Studying flood characteristics for flood warning in the Lower Lam River basin. (Master Thesis), *Hanoi University of Science – Viet Nam National University*, **2014**. In Vietnamese.

19. Jayawardena, A.; Takahasi, Y.; Tachikawa, Y.; Takeuchi, K. Catalogue of Rivers for Southeast Asia and the Pacific – Volume 6: UNESCO–IHP Regional Steering Committee for Southeast Asia and the Pacific, **2012**.

20. Barsi, J.A.; Lee, K.; Kvaran, G.; Markham, B.L.; Pedelty, J.A. The spectral response of the Landsat–8 operational land imager. *Remote Sen.* **2014**, *6(10)*, 10232–10251.

21. ASTER–GDEM. ASTER GDEM 2 README, **2011**.

22. Olivera, F.; Valenzuela, M.; Srinivasan, R.; Choi, J.; Cho, H.; Koka, S.; Agrawal, A. ARCGIS&SWAT: A geodata model and GIS interface for SWAT 1. *JAWRA J. Am. Water Resour. Assoc.* **2006**, *42(2)*, 295–309.

23. Al–sharif, A.A.; Pradhan, B. Monitoring and predicting land–use change in Tripoli Metropolitan City using an integrated Markov chain and cellular automata models in GIS. *Arabian J. Geosci.* **2014**, *7(10)*, 4291–4301.

24. Foody, G.M.; Campbell, N.; Trodd, N.; Wood, T. Derivation and applications of probabilistic measures of class membership from the maximum–likelihood classification. *Photogramm. Eng. Remote Sen.* **1992**, *58(9)*, 1335–1341.

25. Strahler, A.H. The use of prior probabilities in maximum likelihood classification of remotely sensed data. *Remote Sen. Environ.* **1980**, *10(2)*, 135–163.

26. Manandhar, R.; Odeh, I. O.; Ancev, T. Improving the accuracy of land use and land cover classification of Landsat data using post–classification enhancement. *Remote Sen.* **2009**, *1(3)*, 330–344.

27. López, G.G.I., Hermanns, H.; Katoen, J.P. Beyond memoryless distributions: Model checking semi–Markov chains. In Process Algebra and Probabilistic Methods. Performance Modelling and Verification. Springer, **2001**, 57–70.

28. Araya, Y.H.; Cabral, P. Analysis and modeling of urban land cover change in Setúbal and Sesimbra, Portugal. *Remote Sens.* **2010**, *2(6)*, 1549–1563.

29. Moreno, N.; Wang, F.; Marceau, D. J. Implementation of a dynamic neighborhood in a land–use vector–based cellular automata model. *Comput. Environ. Urban Syst.* **2009**, *33(1)*, 44–54.

30. Balzter, H.; Braun, P.W.; Köhler, W. Cellular automata models for vegetation dynamics. Ecol. Modell. **1998**, *107(2–3)*, 113–125.

31. Al–shalabi, M.; Billa, L.; Pradhan, B.; Mansor, S.; Al–Sharif, A.A. Modeling urban growth evolution and land–use changes using GIS–based cellular automata and SLEUTH models: the case of Sana'a metropolitan city, Yemen. *Environ. Earth Sci.* **2013**, *70(1)*, 425–437.

32. Aronoff, S. Classification accuracy: a user approach. *Photogramm. Eng. Remote Sens.* **1982**, *48(8)*, 1299–1307.

33. Aronoff, S. The minimum accuracy value is an index of classification accuracy. *Photogramm. Eng. Remote Sens.* **1985**, *51(1)*, 99–111.

34. Kalkhan, M.A.; Reich, R.M.; Czaplewski, R.L. Statistical properties of five indices in assessing the accuracy of remotely sensed data using simple random sampling. Paper presented at the Proceedings ACSM/ASPRS Annual Convention and Exposition, **1995**.

35. Koukoulas, S.; Blackburn, G.A. Introducing new indices for accuracy evaluation of classified images representing semi–natural woodland environments. *Photogramm. Eng. Remote Sen.* **2001**, *67(4)*, 499–510.

36. Foody, G.M. Status of land cover classification accuracy assessment. *Remote Sen. Environ.* **2002**, *80(1)*, 185–201.

37. Lucas, I.; Janssen, F.; van der Wel, F.J. Accuracy assessment of satellite–derived landcover data: A review. *Photogramm. Eng. Remote Sen.* **1994**, *60(4)*, 479–426.

38. Story, M.; Congalton, R.G. Accuracy assessment: a user's perspective. *Photogramm. Eng. Remote Sens.* **1986**, *52(3)*, 397–399.

39. Smits, P.; Dellepiane, S.; Schowengerdt, R. Quality assessment of image classification algorithms for land–cover mapping: a review and a proposal for a cost–based approach. *Int. J. Remote Sen.* **1999**, *20(8)*, 1461–1486.

40. Campbell, J.B.; Wynne, R.H. Introduction to remote sensing: Guilford Press. **2011**.

41. Lu, Q.; Chang, N.B.; Joyce, J.; Chen, A.S.; Savic, D.A.; Djordjevic, S.; Fu, G. Exploring the potential climate change impact on urban growth in London by a cellular automata–based Markov chain model. *Comput. Environ. Urban Syst.* **2018**, *68*, 121–132. doi:10.1016/j.compenvurbsys.2017.11.006.

42. Eastman, J.R. TerrSet manual. *TerrSet Version* **2015**, *18*, 1–390.

43. Anderson, J.R. Land use and land cover classification system for use with remote sensor data, US Government Printing Office, **1976**, pp. 964.

44. Hadi Memarian, S.K.B.; Talib, J.B.; Sung, C.T.B.; Sood, A.M.; Abbaspour, K. Validation of CA–Markov for Simulation of Land Use and Cover Change in the Langat Basin, Malaysia. *J. Geog. Inf. Sys.* **2012**, *4*, 13. doi:10.4236/jgis.2012.46059.

45. Shafizadeh Moghadam, H.; Helbich, M. Spatiotemporal urbanization processes in the megacity of Mumbai, India: A Markov chains–cellular automata urban growth model. *Appl. Geogr.* **2013**, *40*, 140–149. doi:10.1016/j.apgeog.2013.01.009.

46. Yang, Y.; Zhang, S.; Liu, Y.; Xing, X.; Sherbinin, A. Analyzing historical land–use changes using a Historical Land Use Reconstruction Model: a case study in Zhenlai County, northeastern China. *Sci. Rep.* **2017**, *7*, 41275. doi:10.1038/srep41275.

47. Zabihi, H.; Ahmad, A.; Vogeler, I.; Said, M.N.; Golmohammadi, M.; Golein, B.; Nilashi, M. Land suitability procedure for sustainable citrus planning using the application of the analytical network process approach and GIS. *Comput. Electron. Agric.* **2015,** *117*, 114–126.

48. An, P.; Moon, W.; Rencz, A. Integration of geological, geophysical, and remote sensing data using fuzzy set theory. *Can. J. Explor. Geophys.* **1991**, *27(1)*, 1–11.

49. Luo, X.; Dimitrakopoulos, R. Data–driven fuzzy analysis in quantitative mineral resource assessment. *Comput. Geosci.* **2003**, *29(1)*, 3–13.

50. Ranjbar, H.; Honarmand, M. Integration and analysis of airborne geophysical and ETM+ data for exploration of porphyry type deposits in the Central Iranian Volcanic Belt using fuzzy classification. *Int. J. Remote Sen.* **2004**, *25(21)*, 4729–4741.

51. Saaty, T.L. A scaling method for priorities in hierarchical structures. *J. Math. Psychol.* **1977**, *15(3)*, 234–281.

52. Saaty, T.L. Analytic hierarchy process. In Encyclopedia of operations research and management science. Springer, **2013**, 52–64.

53. Arsanjani, J.J.; Kainz, W.; Mousivand, A.J. Tracking dynamic land–use change using spatially explicit Markov Chain based on cellular automata: the case of Tehran. *Int. J. Image Data Fusion* **2011**, *2(4)*, 329–345.

54. Pontius, R.G. Quantification error versus location error in comparison of categorical maps. *Photogramm. Eng. Remote Sen.* **2000**, *66(8)*, 1011–1016.

55. Cohen, J. A coefficient of agreement for nominal scales. *Educ. Psychol. Meas.* **1960**, *20(1)*, 37–46.

56. Pontius Jr, R.G. Statistical methods to partition effects of quantity and location during comparison of categorical maps at multiple resolutions. *Photogramm. Eng. Remote Sen.* **2002**, *68(10)*, 1041–1050.

57. Kirppendorff, K. Content analysis: An introduction to its methodology. Beverley Hills: Sage. 1989.