

Validation of CA-Markov for Simulation of Land Use and Cover Change in the Langat Basin, Malaysia

Hadi Memarian¹, Siva Kumar Balasundram^{2*}, Jamal Bin Talib¹, Christopher Teh Boon Sung¹,
Alias Mohd Sood³, Karim Abbaspour⁴

¹Department of Land Management, Faculty of Agriculture, Universiti Putra Malaysia (UPM), Serdang, Malaysia

²Department of Agriculture Technology, Faculty of Agriculture, Universiti Putra Malaysia (UPM), Serdang, Malaysia

³Department of Forest Production, Faculty of Forestry, Universiti Putra Malaysia (UPM), Serdang, Malaysia

⁴Swiss Federal Institute of Aquatic Science and Technology (EAWAG), Ueberlandstrasse, Switzerland

Email: *siva@putra.upm.edu.my

Received August 21, 2012; revised September 23, 2012; accepted October 25, 2012

ABSTRACT

Validity of CA-Markov in land use and cover change simulation was investigated at the Langat Basin, Selangor, Malaysia. CA-Markov validation was performed using validation metrics, allocation disagreement, quantity disagreement, and figure of merit in a three-dimensional space. The figure of merit, quantity error, and allocation error for total landscape simulation using the 1990-1997 calibration data were 5.62%, 3.53%, and 6.13%, respectively. CA-Markov showed a poor performance for land use and cover change simulation due to uncertainties in the source data, the model, and future land use and cover change processes in the study area.

Keywords: Land Use and Cover Change; CA-Markov; Calibration; Validation

1. Introduction

Globally, land use change impacts biodiversity, water and radiation budgets, emission of green house gases, carbon cycling, and livelihoods. The study of Land Use and Cover Change (LUCC) and its dynamics is crucial for environmental management, especially with regard to sustainable agriculture and forestry. Different models, in terms of structure and application, have been used to understand LUCC dynamics [1,2].

As emphasized by Verburg *et al.* (2004) [3], Chen and Pontius (2006) [4], and Luo *et al.* (2010) [5], LUCC models are fundamentally different, and have their own capabilities and limitations. Therefore, performance comparison of LUCC models is very difficult.

In this study, CA-Markov was chosen to simulate land use change based on the following requirements: dynamic simulation capability, high efficiency with data scarcity, simple calibration, ability to simulate multiple land covers and complex patterns.

CA-Markov has the ability to simulate land use changes among multiple categories and combines the CA¹ and Markov chain procedures [6]. Markov analysis does not account for the causes of land use change and it is insensitive to space. However, CA-Markov using the CA ap-

proach relaxes strict assumptions associated with the Markov approach and explicitly considers both spatial and temporal changes [7]. CA-Markov also enables a more comprehensive simulation as compared to other LUCC models such as GEOMOD and CLUE [8]. However, CA-Markov calibration is operationally based on a single period of time, which renders difficulty in simulating land cover dynamics on a temporal scale [9]. In addition, the assumptions underlying CA-Markov method tend to be somewhat simplistic when looking at the micro-level changes in land use [10]. Pontius and Malanson (2005) [11] have demonstrated that the predictive power of CA-Markov is higher for cases where it concentrates on the major signal of land changes and ignores noises.

Pontius and Malanson (2005) [11] compared CA-Markov and GEOMOD in terms of simulation power and suitability for different applications in Central Massachusetts, USA. They applied a three stage method to measure simulation power. At first, calibration process was separated from validation process. Secondly, the accuracy was assessed at multiple resolutions. Finally, the calibrated model was compared to a null model that simulates pure persistence. Their investigation showed that the added complexity of spatial contiguity rule in CA-Markov was of no benefit. Paegelow and Camacho (2005) [9] studied the potential and limitation of prospective GIS²-based LUCC modeling. Their approach comprised the Markov chain for temporal simulation,

*Corresponding author.

¹Cellular Automata.

²Geographic Information System.

MCE³, MOLA⁴, and CA to perform spatial contiguity on the simulated land cover scores. Results showed three distinct limitations; the first was caused by complex variability within the land cover categories, the second was caused from using only two land cover maps for calibration, and the third was caused by the fact that MCE, MOLA, and CA only affected spatial distribution, and not temporal distribution, of simulated scores. In another study, Samat (2009) [12] applied CA-Markov to evaluate urban spatial growth in Seberang Perai, Malaysia. Validation analysis was performed using Kappa Agreement Index (KIA). In this work, the poor performance of CA-Markov in the development modeling of commercial/public facilities and industrial activities mainly resulted from: Model development based on physical factors, model assumption based on the uniform spatio-temporal growth of urbanized area, and inability of the model in recognizing new development. Samat *et al.* (2011) [13] showed that the prediction accuracy of CA-Markov decreased when the model tried to predict for a longer period of time, possibly due to the fact that a uniform transition rule was used by the model throughout the simulation period. According to Jokar Arsanjani *et al.* (2011) [14], the critical issue in employing CA-Markov is to combine the social, human and economic dynamics in the simulation, which can be realized in agent-based modeling systems. However, they tried to solve this problem using the integration of logistic regression and CA-Markov.

Araya and Cabral (2010) [15] modeled and analyzed urban land use change using CA-Markov and landscape metrics in Portugal. They reported that CA-Markov was validated successfully, with Klocation and Kquantity of 87% and 83%, respectively. However, they did not analyze the model capability for change simulation using a three-dimensional approach. Wang *et al.* (2010) [16] studied land use change based on vector data source using CA-Markov in China. Although the vector data improved simulation accuracy, validation analysis was performed only based on Kappa values.

Recent validation techniques proposed by Pontius and Millones (2011) [17] and Pontius *et al.* (2011) [18] allow the understanding of complex relationships caused by multiple simulation parameters. Additionally, these techniques facilitate three-dimensional spatial and temporal comparisons of LUCC, and eliminate the weakness of Kappa values for validation analysis.

This work was aimed at evaluating the capability of CA-Markov in simulating LUCC in a tropical catchment using the validation techniques proposed by Pontius and Millones (2011) [17] and Pontius *et al.* (2011) [18].

³Multi Criteria Evaluation.

⁴Multi Objective Land Allocation.

⁵Tropical Rainfall Measuring Mission.

2. Materials and Methods

2.1. Study Area

In recent decades, the Langat Basin has undergone rapid urbanization, industrialization and agricultural development [19,20]. The Langat Basin is also a main source of drinking water for surrounding areas, a source of hydro-power and plays an important role in flood mitigation. Over the past four decades, the Langat Basin has served approximately 50% of the Selangor State population. However, the Selangor State is currently facing water shortage problems, especially in urban areas [21-23]. Hydrometeorologically, the Langat basin is affected by two types of monsoon, *i.e.* the northeast (November-March) and the southwest (May-September) monsoons. Average annual rainfall is about 2400 mm. The wettest months are April and November with average monthly rainfall exceeding 250 mm, while the driest month is June with average monthly rainfall not exceeding 100 mm. Topographically, the Langat basin can be divided into three distinct areas with reference to the Langat River: the mountainous area in the upstream, undulating land in the centre and flat flood plain in the downstream (**Figure 1**). The basin has a rich diversity of landforms, surface features and land cover [23,24]. Dominant soil types in the basin are steepland and Rengam-Jerangau soil series with sandy clay loam and clay textures, respectively [20].

Due to the national importance of the Langat Basin, three sub basins (upstream of the Langat River), *i.e.* Lui, Hulu Langat, and Semenyih, with a total area of 694.13 km² were selected for investigation of land use change (**Figure 1**).

2.2. Data Set

Land use maps dated 1990, 1997 and 2002 and soil maps, with the scale of 1:50,000 were collected from the Soil Resource Management and Conservation Division, Department of Agriculture, Malaysia. Ancillary data such as road and stream networks, contour lines, and population centers were extracted from the topographic maps, with the scale of 1:50,000. Landslide points, geology map (scale: 1/100,000), and location of the water treatment plants were obtained through the library of Universiti Putra Malaysia. Sawmill and oil refinery location maps were extracted from the agriculture yellow pages of Malaysia (www.yellowpages.com.my) and Google Earth. Daily TRMM⁵ radar data during 1998-2009 were obtained through the TOVAS and utilized to produce rainfall map. TRMM Online Visualization and Analysis System (TOVAS), developed by the NASA, provides a user friendly web-based interface for visualization and analysis of TRMM gridded rainfall products and other precipitation data [25]. In this work, land use classification

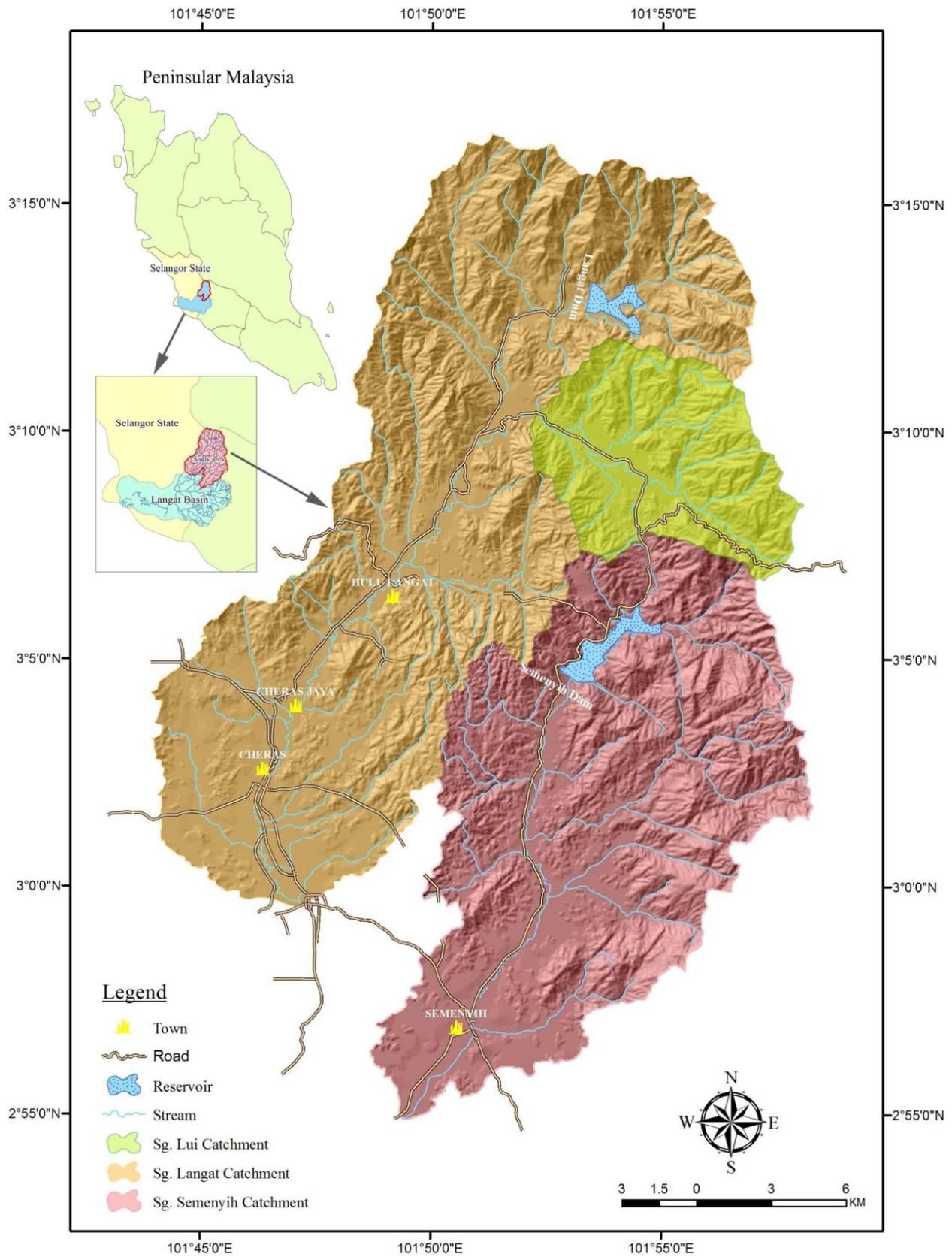


Figure 1. Geographic locations of the Lui, Hulu Langat and Semenyih sub basins within the State of Selangor, Peninsular Malaysia.

and map resolution (*i.e.* 30 m) were outlined based on the objective of a broad study, which was aimed at hydrological analysis of the Langat Basin during 1984-2020. The optimal pixel size was calculated based on the complexity of terrain method [26].

2.3. CA-Markov

A random process, $X(t)$, is a Markov process if for any moment in time, $t_1 < t_2 < \dots < t_n < t_{n+1}$ [27]. The random process satisfies (see Equation (1) below)

where: t_n is the present time so that t_{n+1} represents some points in the future and t_1, t_2, \dots, t_{n-1} represent various points in the past. Based on the present data, future is independent of the past. In other words, future of random process depends neither only on where it is now nor on how it got there [28].

If $X[k]$ is a Markov chain with the states $\{x_1, x_2, x_3, \dots\}$, then probability of transition from the state i to the state j in one time instant is,

$$p_{i,j} = Pr(X[k+1] = j | X[k] = i) \quad (2)$$

If Markov chain has a finite number of states, *i.e.* n , transition probability matrix can be defined as follows:

$$\begin{bmatrix} P_{1,1} & P_{1,2} & \dots & P_{1,n} \\ P_{2,1} & P_{2,2} & \dots & P_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ P_{n,1} & P_{n,2} & \dots & P_{n,n} \end{bmatrix} \quad (3)$$

If transition probabilities vary with the time, the above matrix need to be written explicitly as a function of k (e.g. $p_{i,j,k}$) [28].

CA-Markov in IDRISI program involves two techniques, *i.e.* Markov chain analysis and cellular automata [15]. Cellular automata underlie dynamics of the change events based on proximity concept so that the regions closer to existing areas of the same class are more probable to change to a different class. A cellular automaton is a cellular entity that independently varies its condition based on its previous state (according to a Markov transition rule) and adjacent neighbors [6]. The following 5×5 contiguity filter was utilized in this work:

$$\begin{bmatrix} 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 1 & 1 & 0 \\ 1 & 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 \end{bmatrix}$$

2.4. Modeling

Firstly, three calibration periods, *i.e.* 1990-2002, 1997-2002 and 1990-1997 were considered and correspondent simulation results were pre-analyzed in terms of goodness of fit. Results showed that the period 1990-1997 was more capable to simulate the future changes of land. Therefore, in this study, calibration data over the period 1990-1997 was used to extract transition probability matrix. Then two types of criteria (constraints and factors) were developed to determine which lands are suitable for future development. The constraint factors (dam reservoir and stream) were standardized into a Boolean value (0 and 1), while the parameter maps were standardized into a suitability continuous scale from 0 (minimum suitability) to 255 (maximum suitability) (Figures 2 and 3). Three types of fuzzy membership functions, *i.e.* sigmoidal, symmetrical, and user defined were used to re-scale parameter maps into the range 0 - 255.

Analytic Hierarchy Process (AHP) was used to determine the weights of driving factors. AHP is a measurement theory founded on expert judgment to drive priority scales using pair wise comparisons [29]. In this method, comparisons are made based on a scale of absolute judgment that shows how much more one element prevails over the other for a given attribute (Table 1). However, the judgment may be inconsistent [30]. Considering n elements to be compared ($C_1 \dots C_n$), a_{ij} denotes the relative weight or priority of C_i over C_j . $A = (a_{ij})$ is a square matrix of order n with the constraints, $a_{ij} = 1/a_{ji}$, for $i \neq j$, and $a_{ii} = 1$. Such a matrix is considered as a reciprocal matrix. When the weights are transitive, they will be consistent and $a_{ik} = a_{ij} \cdot a_{jk}$ for all i, j and k . In this condition, $A\omega = \lambda\omega$, ω is an eigenvector (of order n) and λ is an eigenvalue. For a consistent matrix, $\lambda = n$. Human judgments are inconsistent to a greater or lesser degree, therefore in matrices concerning human judgment, the condition $a_{ik} = a_{ij} \cdot a_{jk}$ will not be reliable. In such a case, $A\omega = \lambda_{\max} \times \omega$ and $\lambda_{\max} \geq n$. The difference between λ_{\max} and n is an indicator for inconsistency of the judgment. Consistency Index (CI) can be measured by $(\lambda_{\max} - n)/(n-1)$. Consistency Ratio (CR) is calculated as follows:

$$CR = \frac{CI}{RI} \quad (4)$$

where: RI is the Random Consistency Index which is a function of number of elements. If the CR exceeds 0.1, set of judgments may be too inconsistent to be reliable and the CR of zero means that the judgments are perfectly consistent [31].

SINMAP⁶ approach [32] under ArcView GIS was

$$F_X(X(t_{n+1}) \leq x_{n+1} | X(t_n) = x_n, X(t_{n-1}) = x_{n-1}, \dots, X(t_1) = x_1) = F_X(X(t_{n+1}) \leq x_{n+1} | X(t_n) = x_n), \quad (1)$$

⁶Stability Index Mapping.

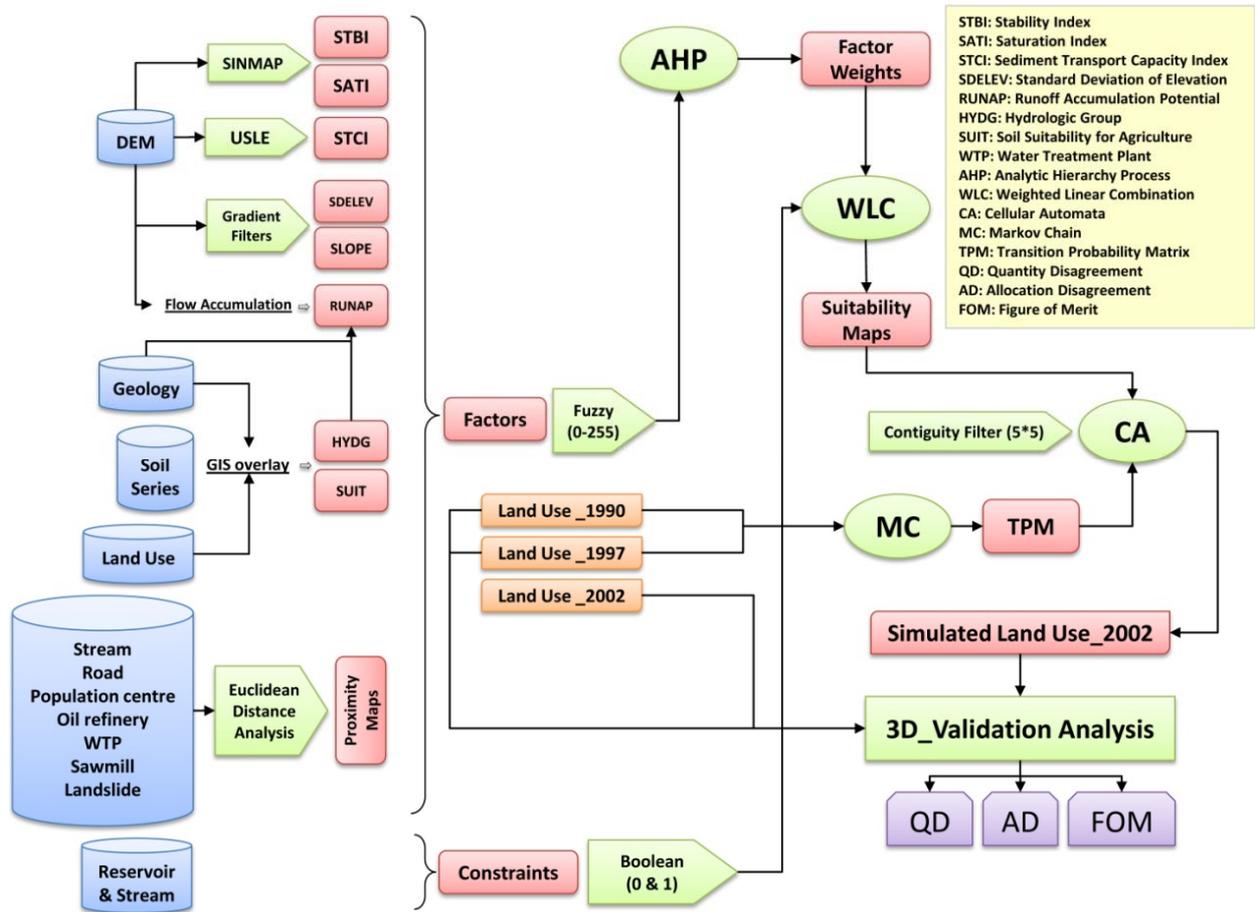


Figure 2. Computational framework of the study.

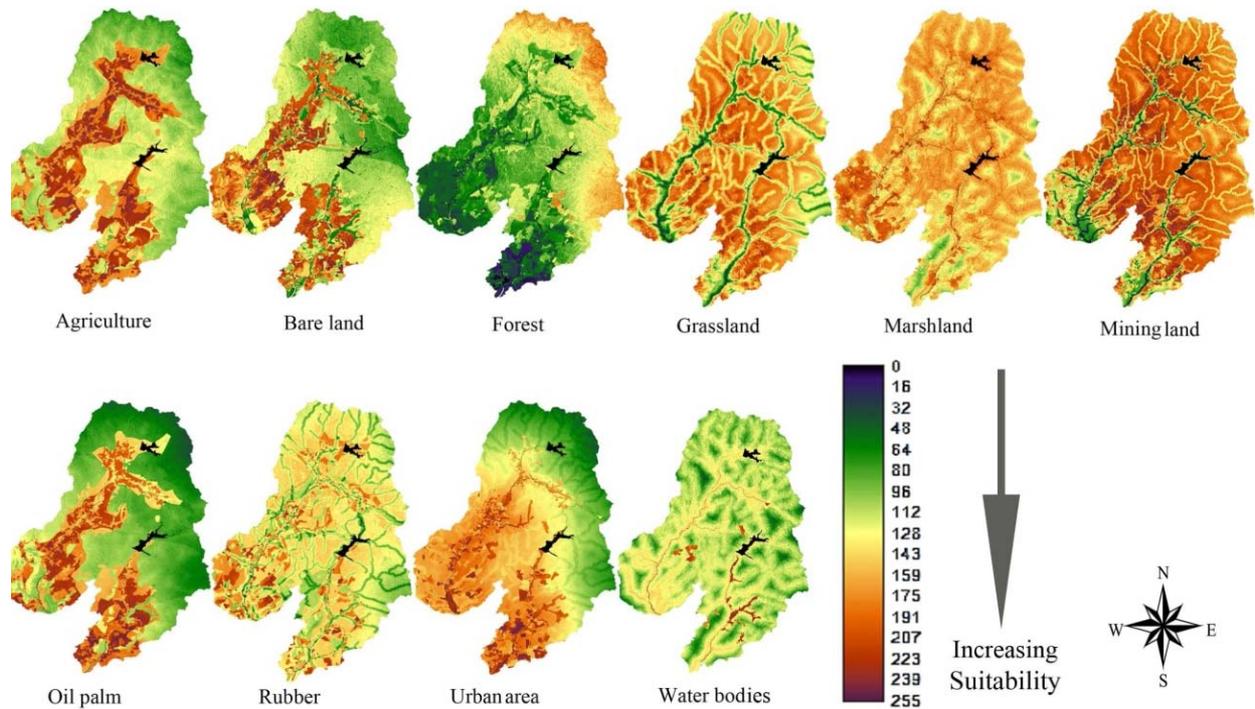


Figure 3. Transition suitability maps for different land covers generated using MCE.

applied to derive stability and saturation indices. Stability index was used in the extraction of suitability maps for agriculture, bare land, mining land, and urban areas, while saturation and roughness indices were used in suitability mapping for marshland and water bodies. Runoff accumulation map was produced based on impervious areas and Digital Elevation Model (DEM). This map was utilized to determine suitable areas for marshland and water bodies. Sediment transport capacity index (topography factor in USLE⁷ model) was used to extract suitability maps for grassland, marshland, and agriculture (Figure 2). Distance to sawmill and oil refinery centers were important parameters in the extraction of suitability maps for forest and oil palm stands, respectively. Distance to Water Treatment Plant (WTP) was used as a driving factor in AHP and MCE processes due to its impact on development of agriculture and urbanization (Figure 2).

As illustrated in Figure 2, Weighted Linear Combination (WLC) method was used in multi criteria evaluation of the drivers and constraints. WLC multiplies each standardized factor map by its factor weight and then sums the results. This result is then multiplied by each constraint to mask out unsuitable areas [6].

2.5. Model Validation

Land use map dated 2002 was used in the model valida-

tion, which was calibrated using the data of 1990-1997.

Disagreement parameters determine the disagreement between simulated map and observed map [17,18,33]. Quantification error (quantity disagreement) happens when the quantity of cells of a category in the simulated map is different from the quantity of cells of the same category in the reference map. Location error (allocation disagreement) occurs where location of a class in the simulated map is different from location of that class in the reference map [17]. These metrics were calculated based on the methods presented by Pontius and Millones (2011) [17] as follows.

2.5.1. Disagreement Components

These calculations and indices were derived directly from Table 2. *J* refers to the number of categories and number of strata in a typical stratified sampling design. Each category in the comparison map is indexed by *i*, which ranges from 1 to *J*. The number of pixels in each stratum is denoted by *N_i*. Each observation is recorded based on its category in the comparison map (*i*) and the reference map (*j*). The number of these observations is summed as the entry *n_{ij}* in row *i* and column *j* of the contingency matrix. Proportion of the study area (*P_{ij}*) that is category *i* in the simulated map and category *j* in the observed map is estimated by the following equation [17]:

Table 1. Rating scale utilized in AHP (adopted from Saaty, 1980).

Intensity of importance	Definition	Explanation
1	Equal importance	Two elements contribute equally to the objective
3	Moderate importance	Experience and judgment slightly favor one element over another
5	Strong Importance	Experience and judgment strongly favor one element over another
7	Very strong importance	One element is favored very strongly over another, its dominance is demonstrated in practice
9	Extreme importance	The evidence favoring one element over another is of the highest possible order of affirmation

Table 2. Format of estimated population matrix (adopted from Pontius and Millones, 2011).

		Reference				Comparison Total
		<i>j</i> = 1	<i>j</i> = 2	...	<i>j</i> = <i>J</i>	
Comparison	<i>i</i> = 1	<i>P</i> ₁₁	<i>P</i> ₁₂	...	<i>P</i> _{1<i>J</i>}	$\sum_{j=1}^J P_{1j}$
	<i>i</i> = 2	<i>P</i> ₂₁	<i>P</i> ₂₂	...	<i>P</i> _{2<i>J</i>}	$\sum_{j=1}^J P_{2j}$

	<i>i</i> = <i>J</i>	<i>P</i> _{<i>J</i>1}	<i>P</i> _{<i>J</i>2}	...	<i>P</i> _{<i>JJ</i>}	$\sum_{j=1}^J P_{Jj}$
Reference Total		$\sum_{i=1}^J P_{i1}$	$\sum_{i=1}^J P_{i2}$...	$\sum_{i=1}^J P_{iJ}$	1

⁷Universal Soil Loss Equation.

$$p_{ij} = \left(\frac{n_{ij}}{\sum_{j=1}^J n_{ij}} \right) \left(\frac{N_i}{\sum_{i=1}^J N_i} \right) \quad (5)$$

The Equation (6) calculates quantity disagreement (q_g) for an arbitrary category g .

$$q_g = \left| \left(\sum_{i=1}^J p_{ig} \right) - \left(\sum_{j=1}^J p_{gj} \right) \right| \quad (6)$$

The overall quantity disagreement (Q) which incorporates all J categories is computed by Equation (7):

$$Q = \frac{\sum_{g=1}^J q_g}{2} \quad (7)$$

Allocation disagreement (a_g) for an arbitrary category g is calculated using Equation (8). The first argument within minimum function is the omission of category g while the second argument is the commission of category g .

$$a_g = 2 \min \left[\left(\sum_{i=1}^J p_{ig} \right) - p_{gg}, \left(\sum_{j=1}^J p_{gj} \right) - p_{gg} \right] \quad (8)$$

The overall allocation disagreement (A) is computed by the following equation:

$$A = \frac{\sum_{g=1}^J a_g}{2} \quad (9)$$

The proportion of agreement C is estimated by Equation (10):

$$C = \sum_{j=1}^J p_{jj} \quad (10)$$

Based on Equation (11), total disagreement D is a sum of overall quantity of disagreement and overall allocation of disagreement:

$$D = 1 - C = Q + A \quad (11)$$

2.5.2. Figure of Merit

The figure of merit equals to the intersection of observed change and simulated change divided by the union of observed change and simulated change. The figure of merit ranges from 0, *i.e.* no overlap between real and predicted changes, to 100%, *i.e.* perfect overlap between real and predicted changes.

$$\text{Figure of Merit} = B / (A + B + C + D) \quad (12)$$

where: A is the area of error due to observed change simulated as persistence, B is the area of correct due to observed change simulated as change, C is the area of error due to observed change simulated as change to wrong category, and D is the area of error due to observed persistence simulated as change [33].

3. Results and Discussion

3.1. Land Use Change

Calibration of CA-Markov was done based on the changes in land cover over the period 1990-1997 (**Figure 4**).

According to transition probability matrix (**Table 3**), the biggest change occurred on grassland. Field and desk investigations demonstrated that grassland with bare land and mining land are transit categories and hence exposed to more changes during the time. Agriculture land⁸, oil palm and rubber stands also changed over the time. Agriculture land shifted mostly to oil palm and rubber plantings and urban areas while older rubber stands were transformed to agricultural and urban areas. Bare land changed mostly to grassland, forest and oil palm plantings. Forest stand primarily changed to grassland, mining land, urban areas, and rubber planting. Grassland and marshland mainly shifted to rubber planting, agricultural and urban areas. Mining land and oil palm stand changed

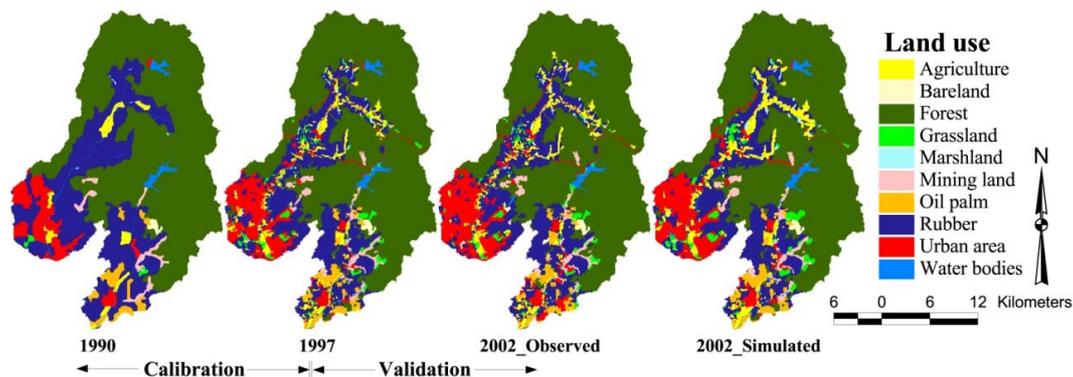


Figure 4. Observed land use map versus simulated land use map.

⁸Refers to cultivation of annual field crops and horticultural crops, and does not include perennial industrial crops, *i.e.* oil palm and rubber.

Table 3. Transition probability matrix for land use change modeling under the 1990-1997 calibration period.

Newer \ Older	Agriculture	Bare land	Forest	Grassland	Marshland	Mining	Oil palm	Rubber	Urban area	Water bodies
Agriculture	0.6947	0.0000	0.0000	0.0050	0.0130	0.0000	0.0634	0.1116	0.1122	0.0001
Bare land	0.0000	0.9003	0.0606	0.0230	0.0000	0.0000	0.0140	0.0022	0.0000	0.0000
Forest	0.0023	0.0003	0.9668	0.0073	0.0004	0.0063	0.0000	0.0112	0.0054	0.0001
Grassland	0.0478	0.0000	0.0976	0.6524	0.0069	0.0072	0.0000	0.1252	0.0228	0.0402
Marshland	0.0415	0.0000	0.0000	0.0000	0.9184	0.0000	0.0000	0.0401	0.0000	0.0000
Mining	0.0083	0.0000	0.1008	0.0356	0.0000	0.7126	0.0068	0.0620	0.0533	0.0206
Oil palm	0.0431	0.0006	0.0472	0.0068	0.0003	0.0389	0.7662	0.0861	0.0109	0.0000
Rubber	0.0733	0.0000	0.0478	0.0255	0.0038	0.0092	0.0160	0.7486	0.0741	0.0018
Urban area	0.0634	0.0000	0.0086	0.0175	0.0003	0.0080	0.0035	0.0701	0.8241	0.0046
Water bodies	0.0000	0.0000	0.0013	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.9987

primarily to forest and rubber plantings, bare land, and urban areas. In this region, water bodies did not change much. However, there were some transitions of water bodies to marshland and urban areas.

However, some of these changes in land use as indicated in the land use maps appear as a mismatch in reality. For example, the transitions of bare land and grassland to forest over a period of 7 years seem unreasonable. Also, the transition of urban areas to agricultural and rubber plantings seem unlikely under the Malaysian policy of land development at the time [34,35]. This mismatch is not caused by the model, but rather it could have been caused by the use of different land use classification criteria or inconsistent resolution of satellite imagery over the 12-year period.

As shown in **Table 4**, twenty factors were entered into AHP to determine the weights proportional to their importance in land allocation process. Consistency ratio in all land cover categories is lower than 0.1, which infers a consistent and reliable judgment. **Figure 3** depicts the transition suitability of different land uses within the Langat Basin.

3.2. Validation

Based on error analysis (**Tables 5** and **6**), change simulation of bare land and grassland categories resulted in the lowest figure of merit, *i.e.* 47.39% and 49.45%, respectively. This could be due to the transit nature of these categories. Conversely, simulation of forest and water bodies resulted in the highest figure of merit, *i.e.* 94.39% and 80.32%, respectively. This could be caused by high percentage of persistence in these categories, *i.e.* 97% and 99% for forest and water bodies, respectively. Due to protection policy of the Hulu Langat Forest Reservoir, transition rate in the forest area was very low, which caused a high persistence. Figure of merit in change simulation of marshland and agriculture land was

51.03% and 57.42%, respectively. In simulation of agriculture land, quantity error was nearly similar to allocation error. This indicates that simulation of agriculture land is impacted by both types of errors. However, in simulation of marshland, quantity error was significantly higher than the allocation error. In this work, simulations of marshland and agriculture land were mostly impacted by transition inconsistency between calibration and validation periods. Figure of merit in change simulation of mining land and oil palm was 60.85% and 66.01%, respectively. In simulations of mining land and oil palm stands, transition consistency between calibration and validation periods was relatively high, which resulted in lower quantity error than allocation error. Figure of merit in change simulation of urban area and rubber stands was 76.37% and 67.71%, respectively. Change simulation of urban area was more affected by allocation error rather than quantity error. However, in simulation of rubber stands, due to transition inconsistency between calibration and validation periods, quantity error reached 2.39%.

Table 5 shows total quantity error, total allocation error, and total figure of merit calculated using a validation analysis over three time periods which are: reference T1, reference T2, and simulated T2. Using the same approach, components of agreement and disagreement as a fraction of landscape area are given in **Table 6**.

Total figure of merit, total allocation error, and total quantity error were 5.62%, 6.13%, and 3.53%, respectively. According to **Table 6** and **Figures 5** and **6**, the components “persistence simulated as change” and “change simulated as persistence” were 5.63% and 3.46%, respectively. The model was able to correctly simulate only 0.57% of the changes. Simulated change is the sum of “persistence simulated as change”, “change simulated as change to wrong category” and “change simulated correctly.” Observed change is the sum of “change simulated as change to wrong category”, “change

Table 4. Type of fuzzy membership function, eigenvectors of weight (values in italic) and AHP consistency ratio for each land use type.

Land cover Parameter	Agriculture	Bare land	Forest	Grassland	Marshland	Mining land	Oil palm	Rubber	Urban area	Water bodies
Proximity to stream	sig. MD 0.0648			sig. SYM 0.2296	sig. MD 0.2462	sig. SYM 0.2466	sig. MI 0.0646	sig. SYM 0.1364	sig. SYM 0.0455	sig. MD 0.3110
Proximity to road	sig. MD 0.1180	sig. MI 0.1698	sig. MI 0.1603	sig. SYM 0.2125		sig. SYM 0.0720	sig. MD 0.1487	sig. SYM 0.1581	sig. MD 0.1141	
Proximity to population centre		sig. MD 0.1900	sig. MI 0.0991	sig. SYM 0.1153		sig. SYM 0.0677		sig. SYM 0.0473	sig. MD 0.1885	
Proximity to sawmill			sig. MI 0.0891							
Proximity to oil refinery							sig. MD 0.0920			
Proximity to WTP	sig. SYM 0.0607						sig. MD 0.0818		sig. MD 0.1885	
Proximity to landslide									sig. MD 0.0271	
Runoff accumulation potential					sig. MI 0.1107					sig. MI 0.0874
Rainfall	sig. SYM 0.0275		sig. MI 0.0535		sig. MI 0.0483		sig. MD 0.0427			
Sediment transport capacity index	sig. MD 0.0232			sig. SYM 0.0623	sig. MD 0.0607					
Stability index	sig. MI 0.0333	sig. SYM 0.0371				sig. SYM 0.0769			sig. MI 0.0150	
Saturation index					sig. MI 0.1046					sig. MI 0.0497
Standard deviation of elevation					sig. MD 0.0946					sig. MD 0.0874
Slope	sig. MD 0.1153	sig. SYM 0.1328	sig. MI 0.2432	sig. SYM 0.0930		sig. SYM 0.1848	sig. SYM 0.0856	sig. MI 0.1341	sig. MD 0.0478	sig. MD 0.0874
Elevation	sig. MD 0.0422		sig. MI 0.0305	sig. SYM 0.0479	sig. SYM 0.0392	sig. SYM 0.0486	sig. MD 0.0319		sig. MD 0.0271	
Soil series			sig. SYM 0.0791							
Soil suitability for agriculture	sig. MD 0.2442						sig. MD 0.2399	sig. MD 0.1559		
Soil hydrologic group		sig. SYM 0.0729		sig. SYM 0.0283	sig. SYM 0.0330				sig. MI 0.0305	sig. MI 0.0256
Geology		sig. MD 0.0575				sig. MD 0.0562	sig. MD 0.0240	sig. MD 0.0368		
Land cover (available lands to develop)	UD 0.2708	UD 0.3398	UD 0.2451	UD 0.2112	UD 0.1631	UD 0.2472	UD 0.1887	UD 0.3314	UD 0.3158	UD 0.3515
Consistency ratio	0.10	0.05	0.03	0.04	0.04	0.02	0.04	0.05	0.05	0.04

Note: sig. MI sigmoidal-Monotonically Increasing; sig. MD sigmoidal-Monotonically Decreasing; sig. SYM sigmoidal-Symmetrical; UD User Defined (Linear Interpolation).

simulated correctly” and “change simulated as persistence”. In this work, due to a high percentage of “persistence simulated as change”, simulated change (6.69%) was higher than the observed change (4.52%) (**Table 6**, **Figures 5 and 6**).

Validation results showed that CA-Markov did not accurately simulate land use change dynamics in the

Langat Basin. The reasons are discussed in the following paragraphs.

As expressed by Pontius and Malanson (2005) [11], several LUCC models assume that land use change would be projected accurately using spatial dependency presented in a contiguity rule. However, **Figure 4** shows that the new patches of bare land and grassland categories

Table 5. Validation results for each category and total landscape.

Category	Loss (commission)	Persistence (agreement)	Gain (omission)	Quantity error	Allocation error	Figure of merit
	% of study area					
Agriculture	1.91	3.70	0.83	1.08	1.66	57.42
Bare land	0.01	0.27	0.29	0.28	0.01	47.39
Forest	2.13	54.88	1.13	1.00	2.27	94.39
Grassland	1.02	1.39	0.40	0.62	0.81	49.45
Marshland	0.21	0.24	0.02	0.19	0.05	51.03
Mining land	0.63	1.51	0.34	0.28	0.69	60.85
Oil palm	0.58	1.94	0.42	0.16	0.84	66.06
Rubber	1.47	17.23	3.86	2.39	2.94	76.37
Urban area	1.43	7.80	2.29	0.86	2.86	67.71
Water bodies	0.27	1.38	0.06	0.21	0.13	80.32
Total	9.66	90.34	9.66	<u>3.53</u>	<u>6.13</u>	<u>5.62</u>

Note: Underlined values were calculated using the three-dimensional approach.

Table 6. Validation results for simulations of total landscape using three-dimensional approach.

Component	Landscape area (%)
Persistence simulated correctly	89.85
Persistence simulated as change	5.63
Change simulated as change to wrong category	0.49
Change simulated correctly	0.57
Change simulated as persistence	3.46
Total	100.00
Simulated change	6.69
Observed change	4.52

do not expand from existing patches. In the present study, CA-Markov imposes spatial dependency by the contiguity rule. As such, CA-Markov was unable to accurately simulate land use change for the bare land and grassland categories.

When the main signal of landscape is persistence, it is important that the model focus on the most important signal of change in the landscape. In this region, there were some noises resulting from the transitions of small patches specifically in agriculture and urban categories. These noises were projected by CA-Markov. However, studies have shown that the predictive power of CA-Markov is higher for cases where it focuses on major signal and ignores noises [11].

Land use maps utilized in this work were generated from satellite imagery, topographic maps, field surveys, and manual digitization. The digitized land use maps clearly showed some mismatch among the land use polygons, especially that of grassland, bare land and ur-

ban area, due to different interpretation of category definition. This led to some errors in the digitization process and consequently affected the simulation accuracy.

In this study, as confirmed by Pontius and Neeti (2010) [36], there were three sources of uncertainty, viz. the data, the model, and future land change processes. Calibration and validation processes in this study required land use maps in three points of time. The duration of time intervals was dependent on data availability. As mentioned previously, there were some inconsistencies between the calibration and validation periods in terms of change intensity that was mostly due to data scarcity and lack of freedom in data selection. This study also showed that CA-Markov was unable to simulate accurately based on marginal changes in land use over the given time period. CA-Markov simulations using the 1990-1997 calibration data set had an overall Kappa of 89% and a standard Kappa of 85%. These Kappa agreement statistics infer high capability of the model in land use change simulation based on the given calibration period. However, the results of three-dimensional validation approach as proposed by Pontius *et al.* (2011) [18] demonstrated that these Kappa accuracies are mostly caused by high land persistence over time, as shown in this study (**Figure 6** and **Table 6**). Therefore, the Kappa agreement statistics are unable to measure the strength of CA-Markov in simulating land use change. Typically, Kappa statistics compare accuracy to a randomness baseline. According to Pontius and Millones (2011) [17], however, randomness is not a logical option for mapping. In addition, several Kappa indices suffer from basic theoretical errors. Thus, it is recommended that quantity disagreement and allocation disagreement be used for accuracy assessment, instead of Kappa statistics.

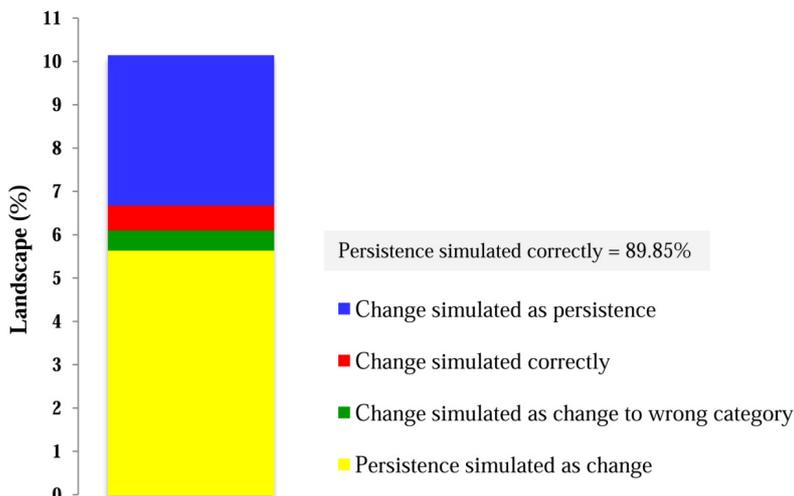


Figure 5. Components of agreement and disagreement.

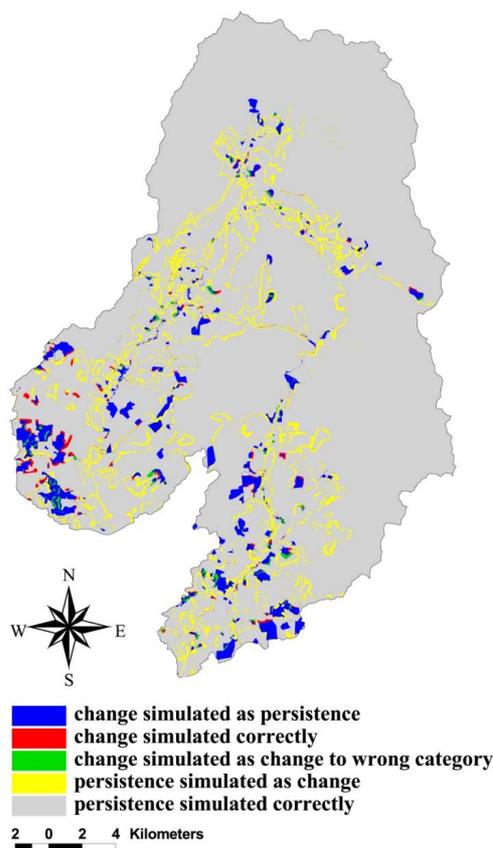


Figure 6. Map representation of agreement and disagreement components.

Each LUCC model has its own capabilities and limitations. A single model is not capable to capture all of key processes in land use change projection [5]. In this study, the apparent differences in land use change between urban and non-urban areas were not well projected by CA-Markov. Therefore, according to the results of this study and other recent works [37,38], and with regard to

data scarcity and the limitation of the applied resolution in this work, the following refinements are suggested as improvement to the simulation accuracy by CA-Markov:

- 1) Modeling urban and non-urban areas separately;
- 2) Integration of logistic regression and CA-Markov to prepare more realistic suitability maps;
- 3) Incorporation of more deterministic socio-economic parameters in land change simulation, such as population growth.

4. Conclusions

In this work, the validity of CA-Markov was studied using the 1990-1997 calibration period.

CA-Markov did not accurately simulate land use change dynamics in this area which resulted in the high values of quantity and allocation disagreements and low value of figure of merit computed using a three-dimensional approach.

Results showed that the new patches of grassland and bare land did not expand from existing patches. CA-Markov imposed spatial dependency by the contiguity rule, which caused inaccurate simulation of land use change for bare land and grassland. Moreover, CA-Markov simulation included noises which were dominant in agriculture and urban area categories. However, the projecting power of CA-Markov was higher in situations where major signals dominated over noises. Mismatching among land cover polygons due to different interpretations of land use definitions contributed to inaccuracies in the simulation of land use change. Data scarcity and lack of freedom in data selection caused some transition inconsistencies between calibration and validation intervals. In addition, CA-Markov exhibited difficulty in change simulation due to marginal changes in this area over time.

This study demonstrates the utility of three-dimensional analysis and disagreement components in validating land

use change models.

5. Acknowledgements

The authors acknowledge Universiti Putra Malaysia for procuring the necessary data for this study.

REFERENCES

- [1] P. H. Verburg, W. Soepboer, A. Veldkamp, R. Limpiada, V. Espaldon and S. S. Mastura, "Modeling the Spatial Dynamics of Regional Land Use: The CLUE-S Model," *Environmental Management*, Vol. 30, No. 3, 2002, pp. 391-405. [doi:10.1007/s00267-002-2630-x](https://doi.org/10.1007/s00267-002-2630-x)
- [2] C. Kamusoko, M. Aniya, B. Adi and M. Manjoro, "Rural Sustainability under Threat in Zimbabwe—Simulation of Future Land Use/Cover Changes in the Bindura District Based on the Markov-Cellular Automata Model," *Applied Geography*, Vol. 29, No. 3, 2009, pp. 435-447. [doi:10.1016/j.apgeog.2008.10.002](https://doi.org/10.1016/j.apgeog.2008.10.002)
- [3] P. H. Verburg, P. P. Schot, M. J. Dijst and A. Veldkamp, "Land Use Change Modelling: Current Practice and Research Priorities," *GeoJournal*, Vol. 61, No. 4, 2004, pp. 309-324. [doi:10.1007/s10708-004-4946-y](https://doi.org/10.1007/s10708-004-4946-y)
- [4] H. Chen and R. G. Pontius Jr., "GEOMOD Modeling," Clark Lab, Clark University, Worcester, 2006.
- [5] G. Luo, C. Yin, X. Chen, W. Xu and L. Lu, "Combining System Dynamic Model and CLUE-S Model to Improve Land Use Scenario Analyses at Regional Scale: A Case Study of Sangong Watershed in Xinjiang, China," *Ecological Complexity*, Vol. 7, No. 2, 2010, pp. 198-207. [doi:10.1016/j.ecocom.2010.02.001](https://doi.org/10.1016/j.ecocom.2010.02.001)
- [6] J. R. Eastman, "IDRISI Kilimanjaro, Guide to GIS and Image Processing," Clark Lab, Clark University, Worcester, 2003.
- [7] C. Agarwal, G. M. Green, J. M. Grove, T. P. Evans and C. M. Schweik, "A Review and Assessment of Land-Use Change Models: Dynamics of Space, Time, and Human Choice," Northeastern Research Station, USDA Forest Service, Delaware, 2002.
- [8] J. F. Mas, M. Paegelow, B. de Jong, O. Masera, G. Guerrero, M. Follador, *et al.*, "Modelling Tropical Deforestation: A Comparison of Approaches," 2008. www.csr.ufmg.br/dinamica/publications
- [9] M. Paegelow and M. T. C. Olmedo, "Possibilities and Limits of Prospective GIS Land Cover Modelling—A Compared Case Study: Garrotxes (France) and Alta Alpujarra Granadina (Spain)," *International Journal of Geographical Information Science*, Vol. 19, No. 6, 2005, pp. 697-722. [doi:10.1080/13658810500076443](https://doi.org/10.1080/13658810500076443)
- [10] S. Srinivasan, "Land Use Change as a Tool: A Framework to Link Transportation and the Environment in New Delhi, India," *The Workshop on Transportation, Land Use and the Environment*, Pune, 3-4 December 2001. http://www.deas.harvard.edu/TransportAsia/workshop_papers/Srinivasan.pdf
- [11] R. G. Pontius Jr. and J. Malanson, "Comparison of the Structure and Accuracy of Two Land Change Models," *International Journal of Geographical Information Science*, Vol. 19, No. 2, 2005, pp. 243-265. [doi:10.1080/13658810410001713434](https://doi.org/10.1080/13658810410001713434)
- [12] N. Samat, "Integrating GIS and CA-Markov Model in Evaluating Urban Spatial Growth," *Malaysian Journal of Environmental Management*, Vol. 10, No. 1, 2009, pp. 83-99.
- [13] N. Samat, R. Hasni and Y. A. E. Elhadari, "Modelling Land Use Changes at the Peri-Urban Areas Using GEOGRAPHIC Information Systems and Cellular Automata Model," *Journal of Sustainable Development*, Vol. 4, No. 6, 2011, pp. 72-84. [doi:10.5539/jsd.v4n6p72](https://doi.org/10.5539/jsd.v4n6p72)
- [14] J. J. Arsanjani, W. Kainz and M. Azadbakht, "Monitoring and Spatially Explicit Simulation of Land Use Dynamics: From Cellular Automata to Geosimulation—A Case Study of Tehran, Iran," 2011 *International Symposium on Image and Data Fusion (ISIDF)*, Tengchong, 9-11 August 2011, pp. 1-4. [doi:10.1109/ISIDF.2011.6024203](https://doi.org/10.1109/ISIDF.2011.6024203)
- [15] Y. H. Araya and P. Cabral, "Analysis and Modeling of Urban Land Cover Change in Setúbal and Sesimbra, Portugal," *Remote Sensing*, Vol. 2, No. 6, 2010, pp. 1549-1563. [doi:10.3390/rs2061549](https://doi.org/10.3390/rs2061549)
- [16] L. Wang, H. Hu, X. Zheng, J. Deng and G. Ning, "Study on LUCC Based on Vector Date Source Using the CA-Markov Model," 2010. www.ieee.org
- [17] R. G. Pontius Jr. and M. Millones, "Death to Kappa: Birth of Quantity Disagreement and Allocation Disagreement for Accuracy Assessment," *International Journal of Remote Sensing*, Vol. 32, No. 15, 2011, pp. 4407-4429. [doi:10.1080/01431161.2011.552923](https://doi.org/10.1080/01431161.2011.552923)
- [18] R. G. Pontius Jr., S. Peethambaram and J. C. Castella, "Comparison of Three Maps at Multiple Resolutions: A Case Study of Land Change Simulation in Cho Don District, Vietnam," *Annals of the Association of American Geographers*, Vol. 101, No. 1, 2011, pp. 45-62. [doi:10.1080/00045608.2010.517742](https://doi.org/10.1080/00045608.2010.517742)
- [19] A. F. Mohamed, W. Z. Wan Yaacob, M. R. Taha and A. R. Samsudin, "Groundwater and Soil Vulnerability in the Langat Basin, Malaysia," *European Journal of Scientific Research*, Vol. 27, No. 4, 2009, pp. 628-635.
- [20] H. Memarian, S. K. Balasundram, J. Talib, C. B. Teh, A. M. Sood, K. C. Abbaspour and A. Haghizadeh, "Hydrologic Analysis of a Tropical Watershed Using KINEROS2," *Environment Asia*, Vol. 5, No. 1, 2012, pp. 84-93.
- [21] K. R. Ayub, L. S. Hin and H. Abd Aziz, "SWAT Application for Hydrologic and Water Quality Modeling for Suspended Sediments: A Case Study of Sungai Langat's Catchment in Selangor," *International Conference on Water Resources (ICWR 2009)*, Langkawi, 26-27 May 2009. http://redac.eng.usm.my/html/publish/2009_09.pdf
- [22] H. Juahir, S. M. Zain, M. K. Yusoff, T. I. Hanidza, A. S. Armi, M. E. Toriman, *et al.* "Spatial Water Quality Assessment of Langat River Basin (Malaysia) Using Environmental Monitoring Assessment," *Environmental Monitoring Assessment*, Vol. 173, No. 1-4, 2010, pp. 625-641. [doi:10.1007/s10661-010-1411-x](https://doi.org/10.1007/s10661-010-1411-x)
- [23] H. Memarian, S. K. Balasundram, J. Talib, A. M. Sood and K. C. Abbaspour, "Trend Analysis of Water Discharge and Sediment Load during Past Three Decades of Development at the Langat Basin, Malaysia," *Hydrological Science*

- ences Journal*, Vol. 57, No. 6, 2012, pp. 1-16.
[doi:10.1080/02626667.2012.695073](https://doi.org/10.1080/02626667.2012.695073)
- [24] M. H. Noorazuan, *et al.* "GIS Application in Evaluating Land Use—Land Cover Change and Its Impact on Hydrological Regime in Langat River Basin, Malaysia," *Proceedings of the Conference Map Asia 2003*, Kuala Lumpur, 2003.
http://www.geospatialworld.net/index.php?option=com_content&view=article&id=15467
- [25] Z. Liu, D. Ostrenga, W. Teng and S. Kempler, "Tropical Rainfall Measuring Mission (TRMM) Precipitation Data and Services for Research and Application," *Bulletin of the American Meteorological Society*, Vol. 93, No. 9, 2012, pp. 1317-1325. [doi:10.1175/BAMS-D-11-00152.1](https://doi.org/10.1175/BAMS-D-11-00152.1)
- [26] T. Hengl, "Finding the Right Pixel Size," *Computers and Geosciences*, Vol. 32, No. 9, 2006, pp. 1283-1298.
[doi:10.1016/j.cageo.2005.11.008](https://doi.org/10.1016/j.cageo.2005.11.008)
- [27] A. A. Markov, "Extension of the Limit Theorems of Probability Theory to a Sum of Variables Connected in a Chain," *The Notes of the Imperial Academy of Sciences of St. Petersburg*, VIII Series, Physio-Mathematical College XXII, 1907.
- [28] S. L. Miller and D. Childers, "Markov Processes," In: *Probability and Random Processes*, Academic Press, Burlington, 2004, pp. 323-367.
- [29] T. L. Saaty, "The Analytic Hierarchy Process," McGraw-Hill, New York, 1980.
- [30] T. L. Saaty, "Decision Making with the Analytic Hierarchy Process," *International Journal of Services Sciences*, Vol. 1, No. 1, 2008, pp. 83-98.
[doi:10.1504/IJSSCI.2008.017590](https://doi.org/10.1504/IJSSCI.2008.017590)
- [31] G. Coyle, "The Analytic Hierarchy Process (AHP)," 2004.
www.booksites.net
- [32] R. T. Pack, D. G. Tarboton and C. N. Goodwin, "SINMAP, a Stability Index Approach to Terrain Stability Hazard Mapping," UTAH State University, Logan, 2001.
- [33] R. G. Pontius Jr., W. Boersma, J. C. Castella, K. Clarke, T. de Nijs, C. Dietzel, *et al.*, "Comparing the Input, Output, and Validation Maps for Several Models of Land Change," *The Annals of Regional Science*, Vol. 42, No. 1, 2008, pp. 11-37. [doi:10.1007/s00168-007-0138-2](https://doi.org/10.1007/s00168-007-0138-2)
- [34] R. Leete, "Malaysia, Achieving the Millennium Development Goals, Successes and Challenges," The United Nations Country Team, Kuala Lumpur, 2005.
- [35] I. Omar, "Rules Affecting the Land Development Process in Malaysia—A Review on Regulation of Environmental Impact Assessment (EIA)," *8th Pacific Rim Real Estate Society Conference*, Lincoln University, Christchurch, 21-23 January 2002, pp. 1-21.
- [36] R. G. Pontius Jr. and N. Neeti, "Uncertainty in the Difference between Maps of Future Land Change Scenarios," *Sustainability Science*, Vol. 5, No. 1, 2010, pp. 39-50.
[doi:10.1007/s11625-009-0095-z](https://doi.org/10.1007/s11625-009-0095-z)
- [37] L. Poelmans and A. Van Rompaey, "Complexity and Performance of Urban Expansion Models," *Computers, Environment and Urban Systems*, Vol. 34, No. 1, 2009, pp. 17-27. [doi:10.1016/j.compenvurbsys.2009.06.001](https://doi.org/10.1016/j.compenvurbsys.2009.06.001)
- [38] J. J. Arsanjani, M. Helbich, W. Kainz and A. D. Bolorani, "Integration of Logistic Regression, Markov Chain and Cellular Automata Models to Simulate Urban Expansion," *International Journal of Applied Earth Observation and Geoinformation*, Vol. 24, 2012, pp. 265-275.
[doi:10.1016/j.jag.2011.12.014](https://doi.org/10.1016/j.jag.2011.12.014)