Urban Landuse changes Prediction by using a CA–Markov and Logistic Regression Analysis

Boonchoo, K.
Department of Engineering Technology, Faculty of Engineering and Technology, Ragamangala University of Technology Srivijaya, 179 Maifad Sub-district, Sikao district, Trang 92150, Thailand, E-mail: krittawat.b@rmutsv.ac.th

Abstract
Phuket is the province with an island landscape and is world known for its tourist attraction. In the past, the development of the tourism and recreation industry has significantly increased. This has resulted in a dramatic increase of the population and the rapid growth of the city. In this study, identification and analysis of urban and built-up areas from land use and land cover (LULC) in 1995, 2002 and 2014 are done. Furthermore, remote sensing (RS), geographic information system (GIS) data and logistic regression analysis (LRA) were applied to find the correlation of driving factor spatial variation in determining change conditions. The results provided the possibility of a LULC change with cellular-automata (CA)-Markov and LRA, especially urban growth between 2014 and 2026 showed an increase of 25,064.45 Rai (4,010.31 Hectare). LULC prediction in 2026 of urban land use change area based on an optimum geospatial model is strong result recommendation can bring an urban planning to solve any problems that may occur in the future.

1. Introduction

Phuket Island, located in the southern tip of Thailand is among the most famous touristic destinations in Thailand. As a result of increased tourism, land use and land cover (LULC) of Phuket Island has changed rapidly, especially in the surrounding agricultural areas. The habitation areas are expanding in this region resulting in the drastic decrease in the agricultural areas. This change is causing a lot of problems such as the environment of the community, water shortage, waste management, deforestation planning and management etc. This is mainly because it's generally said known big cities have large physical complex systems such that they are interrelated with various subsystems so if one gets affected others are affected too. The understanding of this complex system can be done by creating a model which can be used to predict trends for urban growth. Jantz et al., (2004) classified the urban and built-up area by using satellite data to replicate the urban growth of the area.

Herold et al., (2003) applied the cellular-automata (CA) Markov theory to the urban and built-up expansion model under the Fundamental Grid Analysis. From the extensively applied urban CA models, they proposed four perspectives for improvements: adjustment of the model unit, incorporation of temporal contexts, the construction of a public platform for model comparison and scenario-based analyses. In this study, three main key issues are addressed (1) to assess the status and change of urban land use in Phuket Island; (2) to study the feasibility of driving factors for urban land use change; (3) to predict the trend of urban land use change in the future. In this study, we will identify and analyze urban areas from black and white orthophoto in 1995, color orthophoto in 2002 and Thaichote satellite imagery in 2014 to study the urban growth. We will use applied remote sensing (RS), geographic information system (GIS) data and logistic regression analysis (LRA) to correlate with driving factor variation in determining change conditions. It will be then compared with the non-driver variants derived from the Markov chain analysis, both methods were processed with the cellular automata model, and then validated accuracy with kappa hat. In the end, it will be compared with pairwise Z test to find an optimal model to predict the future expansion of the city in 2026.

2. Materials and Methods

2.1 Study Area

Phuket Island is located in the southern provinces (Changwat) of Thailand and it is country's largest island with an area of 522 sq. km. Phuket Island is one of Thailand's most popular tourist destinations and hence a large number of tourists from all over the world come here for vacations. Phuket has hence emerged as a booming market for hospitality sector

(Jones Lang LaSalle, 2013) which has dramatically increased the population in the last decade in this area (Department of Provincial Administration, 2013). Furthermore, this island has a significant amount of migration of local population, who come from mainland Thailand for jobs. Hence, tourism has a great impact on the livelihood of the local community as it is the large source of income on this Island. This growth of tourism has both positive impacts on the local and national economy and negative as the rapid expansion clearly comes with risks or side effects (Sae-Tan, 2013). Urban land use change has hence affected a number of aspects which should be studied for urban planning and management.

2.2 Research Framework

2.2.1 Data collection and preparation

The required input data were collected and prepared in for spatial analysis with raster format at the resolution of 25 m. We used the data by Ongsomwang and Boonchoo (2016) who applied black and white orthophotos in 1995, color orthophotos in 2002 and pan-sharpened THEOS images in 2014 which is used to visually interpreted LULC classes of the whole Phuket Island for assessing LULC data. In this study, LULC Classification is done following from Land Development Department in the first level with these classes: urban and built-up area, agriculture land, forest land, water body and miscellaneous land. In addition, LRA to use urban land use change driving factors as include: 1) Elevation, 2) Slope, 3) Soil fertility 4) Distance from the road, 5) Distance from the settlement, and 6) Distance from the water body. The conceptual framework of research is displayed in Figure 1 and it is separately described in the following sections.

2.2.2 LULC Prediction in 2014 and optimum geospatial models

We used two geospatial models for LULC prediction including CA-Markov model, and CA-Markov with LRA were applied to predict LULC in 2014. The CA-Markov model can be predicted by the probability of only LULC change.

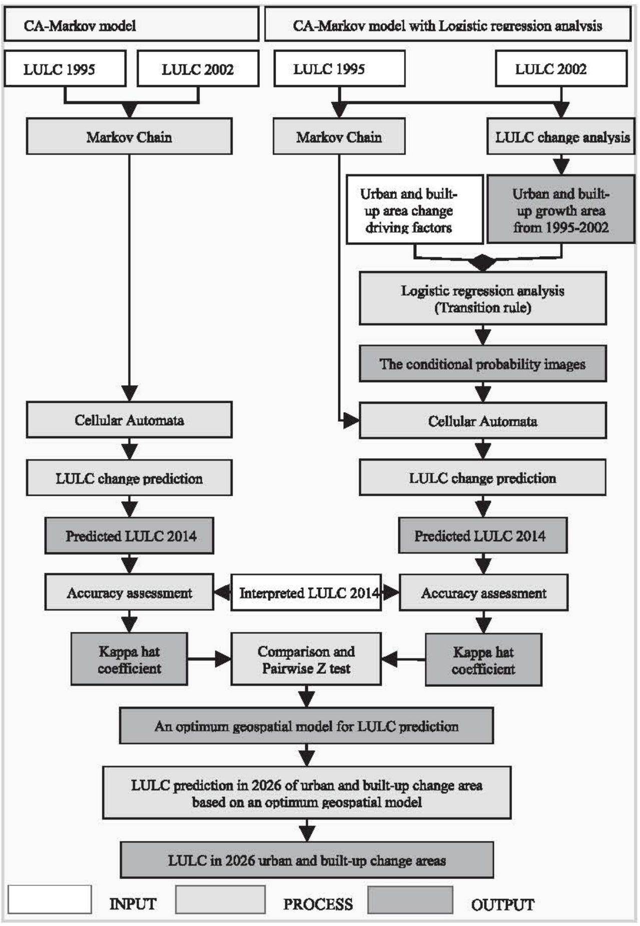


Figure 1: The conceptual framework of research

On the other hand, another method used LRA to create condition map from driving factors for urban land use change prediction in the future. For CA-Markov model, two main input data: LULC in 1995 and 2002 are used to predict LULC in 2014 with Markov Chain and Cellular Automata. Meanwhile, CA-Markov with LRA requires LULC in 1995 and 2002 as basic input with LULC change driving factors (see Table 3) for LULC change is applied to predict LULC in 2014 based on LULC change analysis (Urban and built-up growth area from 1995-2002) with the urban land use change vulnerability driving factors.

Additionally, urban and built-up area change driving factors were used to identify the urban land use location preference by logistic regression analysis (Transition rule) to create the conditional probability images for allocating the urban area in 2014. After that, Cellular Automata was used to predict LULC in 2014. Subsequently, the predicted LULC in 2014 from each model is compared with the interpreted LULC in 2014 to identify an optimum model for LULC prediction based on Kappa hat coefficient with pairwise Z test. Herewith, Kappa hat coefficient (KHAT) among two geospatial models for LULC prediction are examined significantly different as suggested by (Congalton and Green, 2009) as below in equation 1:

$$Z = \frac{|\widehat{K_1} - \widehat{K_2}|}{\sqrt{\widehat{var}(\widehat{K_1}) + \widehat{var}(\widehat{K_2})}}$$

Equation 1

Where Z is normalized and standard normal distribution,

$\widehat{K_1}$ is KHAT of the first geospatial model,
$\widehat{K_2}$ is KHAT of the second geospatial model,
$\widehat{var}(\widehat{K_1})$ is variance of KHAT of the first geospatial model, and
$\widehat{var}(\widehat{K_2})$ is variance of KHAT of the second geospatial model.

Meanwhile, variance of KHAT is calculated using following equation:

$$\widehat{var}(\widehat{K}) = \frac{1}{n}\left\{\frac{\theta_1(1-\theta_1)}{(1-\theta_2)^2} + \frac{2(1-\theta_1)(2\theta_1\theta_2-\theta_3)}{(1-\theta_2)^3} + \frac{(1-\theta_1)^2(\theta_4-4\theta_2{}^2)}{(1-\theta_2)^4}\right\}$$

Equation 2

where $\theta_1 = \frac{1}{n}\sum_{i=1}^{k} n_{ii}$,

$\theta_2 = \frac{1}{n^2}\sum_{i=1}^{k} n_{i+}n_{+i}$,

$\theta_3 = \frac{1}{n^2}\sum_{i=1}^{k} n_{ii}(n_{i+} + n_{+i})$ and

$\theta_4 = \frac{1}{n^3}\sum_{i=1}^{k}\sum_{j=1}^{k} n_{ij}(n_{j+} + n_{+i})^2$

In practice, given the null hypothesis H_0: $(\widehat{K_1} - \widehat{K_2}) = 0$ and the alternative H_1 : $(\widehat{K_1} - \widehat{K_2}) \neq 0$, H_0 is rejected if $Z \geq Z_{\alpha/2}$, where $\alpha/2$ is the confidence level of the two-tailed Z test and the degrees of freedom are assumed to be ∞ (infinity). This result was used to identify an optimal model for LULC prediction in the future.

2.2.3 LULC prediction in 2026

The derived optimum geospatial model was used for LULC prediction for 2026 to predict future urban and built-up change area.

3. Results and Discussion

3.1 LULC Assessment in 1995, 2002, and 2014

The areas and percentages of the LULC classes on Phuket Island in 1995, 2002, and 2014 are comparatively summarized in Table 1. The distribution of LULC pattern of Phuket Island in 1995, 2002, and 2014 are shown in Figure 2. Furthermore, 756 randomly stratified sampling points based on multinomial distribution theory with the desired level of confident 95 percent and a precision of 5 percent are used for accuracy assessment of LULC data in 2014. The accuracy assessment of the interpreted LULC in 2014 by the ground survey in 2017 is 98.28% for overall accuracy and 97.57% for Kappa hat coefficient was shown in Table 2. The overall urban growth rate of the study area continuously increased. Conversely, areas of agricultural land and forest land continuously decreased during the same period.

These basic LULC data were further used to extract the urban growth on Phuket Island. However, urban areas change between 1995 and 2014 was expanded from 17132.03 Rai (2741.13 Hectare), 2673.05 Rai (427.69 Hectare), 388.67 Rai (62.19 Hectare) and 16996.88 Rai (2719.50 Hectare) for agriculture land, forest land, water body and miscellaneous land, respectively.

3.2 Optimum Geospatial Model for LULC Prediction

Two main set of input data were used from the study of Ongsomwang and Boonchoo (2016) who applied the CA-Markov model where the LULC in 1995 and 2002, were used to predict the LULC in 2014 with the Markov chain and cellular automata. An application of this model was described in more detail by (Paegelow and Camacho Olmedo, 2005) and its application can be found in (Ongsomwang and Saravisutra, 2011). Meanwhile, the CA-Markov with

LRA required the LULC in 1995 and 2002 as basic input with driving factors on the LULC change (Table 3) which was applied to predict the LULC in 2014 based on the LULC change analysis and transition rule of the LULC change with logistic regression analysis. Basically, the LRA can be applied to identify trends in the LULC change, urban growth, and habitat modelling (Gontier et al., 2009).

Table 1: Area and percentage of LULC classes on Phuket Island in 1995, 2002, and 2014

LULC	**1995**		**2002**		**2014**	
	sq. km	**%**	**sq. km**	**%**	**sq. km**	**%**
Urban and built-up area	66.91	12.80	78.82	15.07	126.28	24.15
Agriculture land	233.89	44.74	226.97	43.41	209.12	40.00
Forest land	121.52	23.24	116.08	22.20	106.64	20.40
Water body	13.95	2.67	14.23	2.72	14.54	2.78
Miscellaneous land	86.56	16.56	86.74	16.59	66.26	12.67
Total	522.84	100.00	522.84	100.00	522.84	100.00

Table 2: Error matrixes and accuracy assessment of LULC for Phuket Island in 2017

LULC 2014 by Visual interpretation	**Ground truth data**						
	U	**A**	**F**	**W**	**M**	**Total**	**User accuracy**
Urban and built-up area (U)	184					184	100.00
Agricultural land (A)		295	6			301	98.01
Forest (F)		2	192			194	98.97
Water body (W)				11		11	100.00
Miscellaneous land (M)		1	4		61	66	92.42
Total	184	298	202	11	61	756	
Procedure accuracy	100.00	98.99	95.05	100.00	100.00		

Over all accuracy = 0.9828 and Kappa hat accuracy = 0.9757

Table 3: List of variables included in the logistic regression model

Variable	**Meaning**	**Nature of variable**
Dependent Y	 0 – no urban growth; 1 – urban growth	 Discrete
Independent X1 X2 X3 X4 X5 X6	 Elevation (m) Slope (%) Soil fertility Distance from road (m) Distance from settlement (m) Distance from water body (m)	 Continuous Continuous Continuous Continuous Continuous Continuous

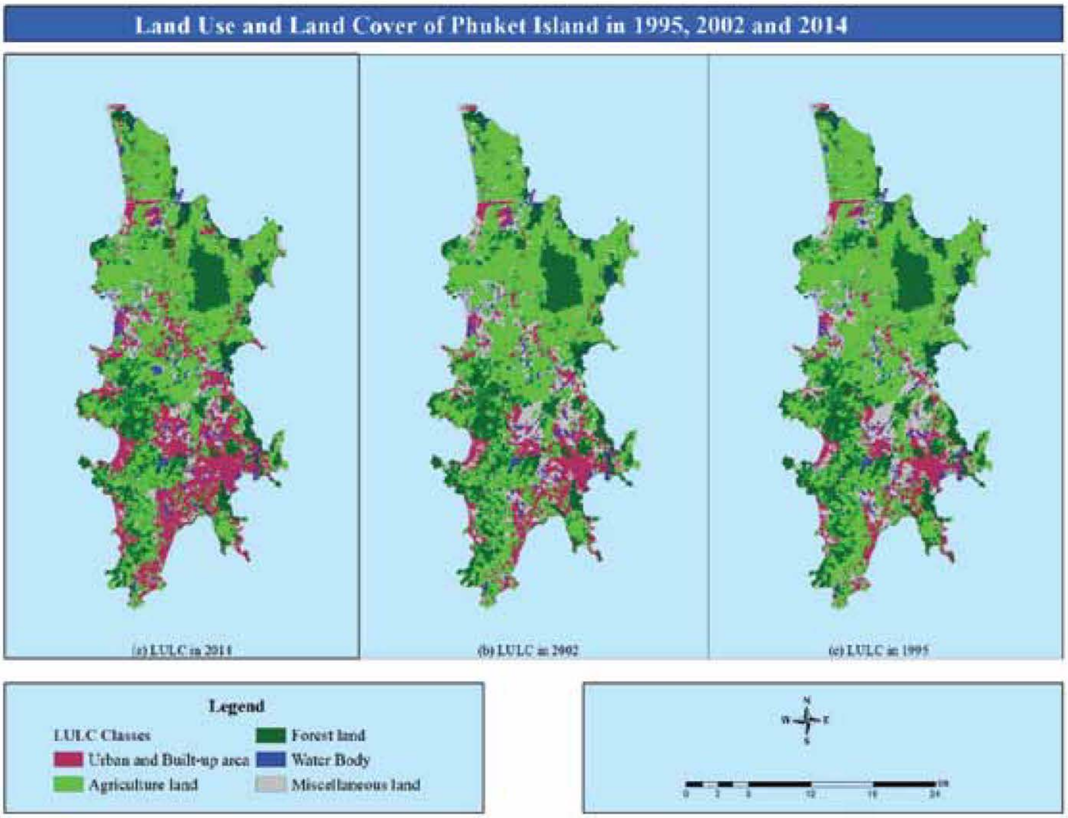
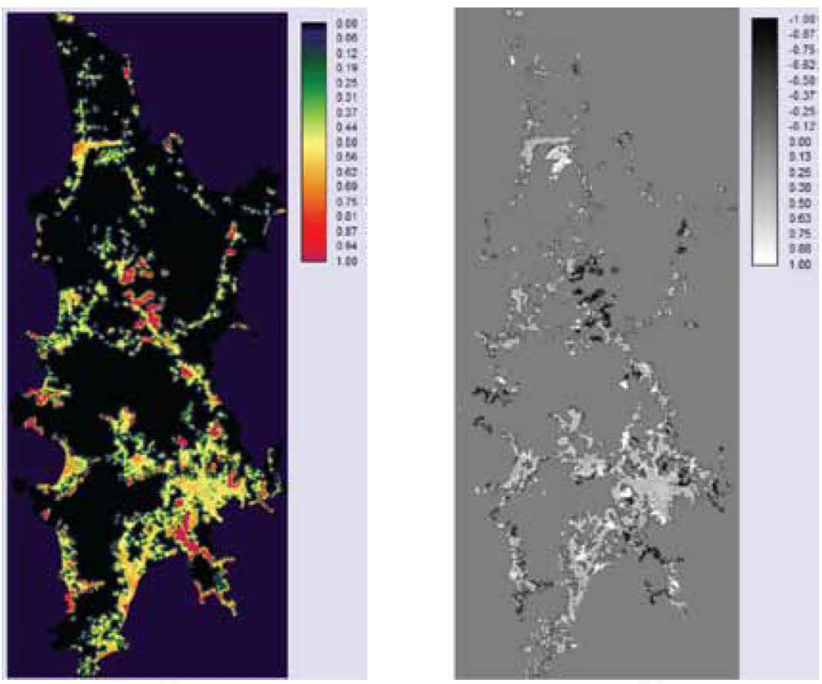


Figure 2: Distribution of LULC pattern of Phuket Island in 1995, 2002, and 2014



(a) (b)

Figure 3: (a) Urbanization probability maps of Phuket Island prediction in 2014. Lighter tones indicate higher probabilities of urban growth (b) Residual of regression

Meanwhile, the LRA is a method to discover the empirical relationships between a binary dependent and several independent categorical and continuous variables (Arsanjani et al., 2013). The LRA was used here to associate the urban expand areas between 1995 and 2002 with the driving factors on urban growth for generating the urban land use change vulnerability index (UVI). In practice, a multivariate regression analysis was first used to identify the linear relationship (Z) between the urban expansion areas and their driving factors (x_1, x_2, x_3, ⋯ , x_n) as:

$$Z = b_0 + b_1x_1 + b_2x_2 + b_3x_3 + \cdots + b_nx_n.$$

Equation 3

After that, the probability values of the driving factors were transformed into a nonlinear relationship by a logistic curve to create the UVI as a probability value as:

$$UVI = \frac{1}{1+e^{-Z}}$$

Equation 4

Examples of the LRA is application on urban growth can be found in Hu and Lo, 2007. Logistic regression of urban land use change between 1995 and 2002 results were regression coefficients, means and standard deviations for the logistic regression model containing the 6 independent variables as shown in Table 3. The probability map was used for recognizing probability of future urban growth areas and quantity. In the study area, urban area accounted for 12.80% in 1995 while for 2002 the same is 15.07%.

To produce the spatial pattern of urban distribution in a certain amount of urban area, the increase of the number of urban cells is compared to the 2002 base urban map. Then the number of urbanized cells was represented as base on equation 1 to the probability map (Figure 3a). This generated a growth map. The residuals were not standardized and are bounded between -1 and 1. The residuals are calculated as in Figure 3b:

$$Residual\ [i] = bserved\ (Yi) - Predicated\ probability\ (i)$$

Equation 5

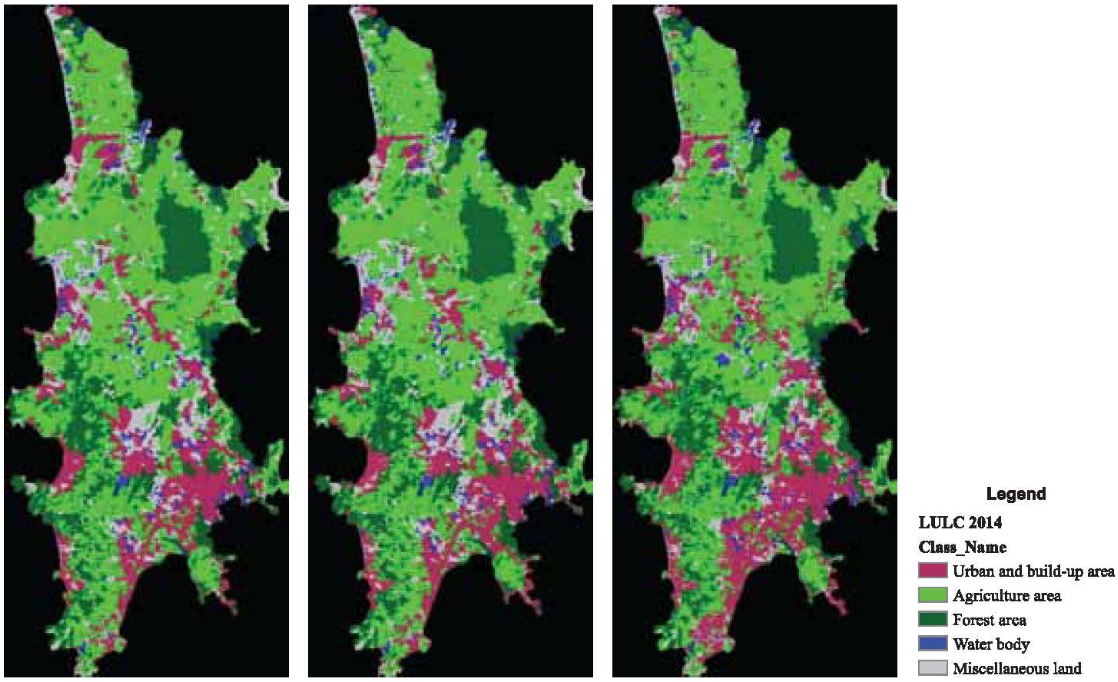
Urban areas which tend to grow close to slope (x_2) has a coefficient of -0.06126431. Distance from settlement (x_5) has a coefficient of -0.02235446. On the other hand, soil (x_3) has a coefficient of 3.93954082 as shown in Table 4, which means that the urban areas expand in medium soil fertility level area.

Relative operating characteristic (ROC) was used to validate the logistic regression model. ROC method is an excellent method to evaluate the validity of a model that predicts the occurrence of an event by comparing a probability image depicting the probability of that event occurring and a binary image showing where that class actually exists. In this study, the result with 100 thresholds (Sample-based computation when applicable) was a ROC value of 0.9427. A ROC value of 1 indicates that there is a perfect spatial agreement between the actual urban growth map and the predicted probability. A ROC value of 0.5 is the agreement that would be expected due to chance, i.e., the cells values on the predicted probability image were assigned to random locations. However, the accuracy of the two geospatial models for the LULC prediction was summarized to show an overall 83.24% with kappa 76.90% for CA-Markov and overall 83.43% with kappa 77.17% for CA-Markov with LRA. The distribution of LULC pattern from CA-Markov, CA-Markov with LRA and On Screen Digitizing of Phuket Island in 2014 was shown in Figure 4.

Table 4: Estimated coefficients, means and standard deviations for the logistic regression model to predict the LULC in 2014

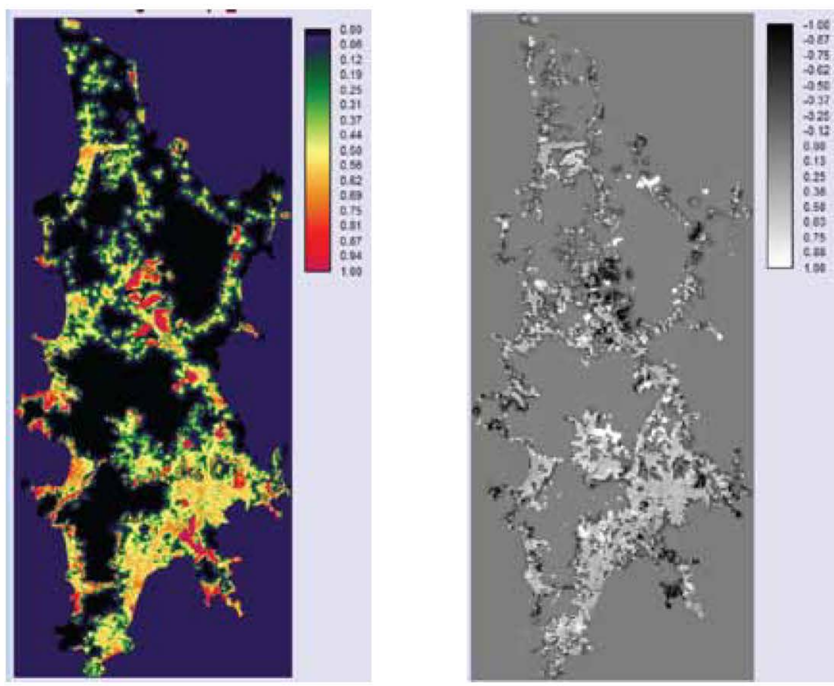
Variables	Coefficient	Mean	Standard Deviation
Constant	-3.789	0.078	0.268
x1	-0.015	34.527	72.463
x2	-0.061	9.076	16.266
x3	3.940	0.706	0.765
x4	0.001	128.773	229.384
x5	-0.022	249.671	431.702
x6	0.001	305.136	475.651

Z=Urban Growth, x_1=Elevation (m), x_2=Slope (%), x_3=Soil fertility, x_4=Distance from road (m), x_5=Distance from settlement (m), x_6=Distance from water body (m)



LULC in 2014 (CA-Markov) LULC in 2014 (CA-Markov with LRA) LULC in 2014

Figure 4: Distribution of LULC pattern from CA-Markov, CA-Markov with LRA and On Screen Digitizing of Phuket Island in 2014



(a) (b)

Figure 5: (a) Urbanization probability maps of Phuket Island prediction in 2026. Lighter tones indicate higher probabilities of urban growth (b) Residual of regression

Table 5: Estimated coefficients, means and standard deviations for the logistic regression model to predict the LULC in 2026

Variables	Coefficient	Mean	Standard Deviation
Constant	-3.341	0.125	0.331
x_1	-0.027	34.527	72.463
x_2	-0.051	9.076	16.266
x_3	3.693	0.706	0.765
x_4	0.001	128.773	229.384
x_5	-0.009	220.906	396.930
x_6	0.001	293.100	460.080

Z=Urban Growth, x_1=Elevation (m), x_2=Slope (%), x_3=Soil fertility, x_4=Distance from road (m),x_5=Distance from settlement (m), x_6=Distance from water body (m)

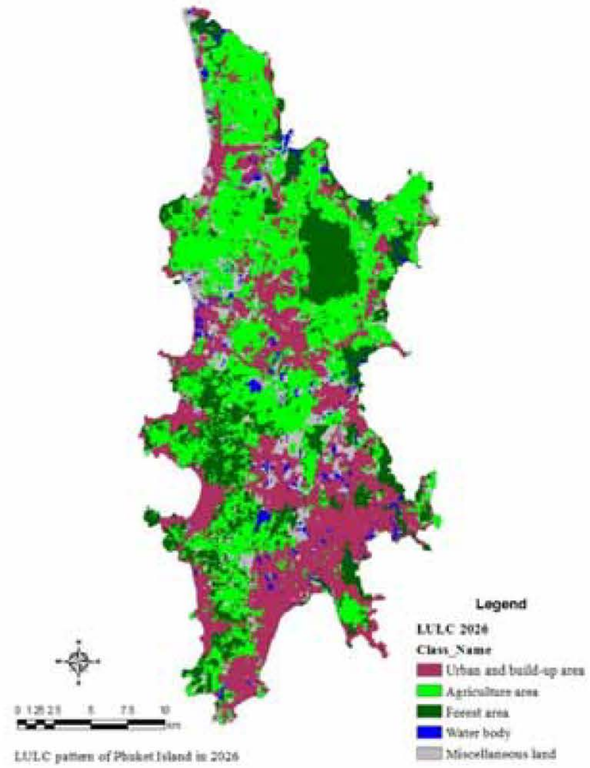


Figure 6: Distribution of LULC pattern of Phuket Island in 2026

However, the results of the pairwise Z-test between the CA-Markov with LRA and CA-Markov for the LULC prediction for urban growth were not significantly different at the 90%, 95%, or 100%. It was accepted two-tail confidence level as 1.65, 1.96 and 2.58, respectively. In this study, the pairwise Z statistic value as 3.42 was accepted two-tail confidence level at 100%. As a result, the confidence level from the CA-Markov with LRA model was the optimum geospatial model.

3.3 LULC Prediction in 2026

The derived optimum geospatial model was the CA-Markov with LRA model used to predict the LULC in 2026 as shown in Figure 5. The LRA was used here to associate the urban expansion areas between 2002 and 2014 with the driving factors on urban growth for generating the UVI. Logistic regression results was present regression equation as follow as

$$Z = -3.3412 - 0.027367\times x_1 - 0.050623\times x_2 + 3.693246\times x_3 + 0.000507\times x_4 - 0.009432\times x_5 + 0.000970\times x_6$$

Equation 6

Where Z=Urban Growth, and driving factor were x_1=Elevation (m), x_2=Slope (%), x_3=Soil fertility, x_4=Distance from road (m), x_5=Distance from settlement (m) and x_6=Distance from water body (m). However, urban areas tend to grow close to slope (x_2) has a coefficient of - 0.050623 and elevation (x_1) has a coefficient of -0.027367. In contrast, soil fertility (x_3) has a coefficient of 3.693246, which means that urban expand in moderate soil fertility area due to it has two level include as moderate and low soil fertility area. Moderate soil fertility was spread in area with elevation which is mostly in agricultural land and forest areas. On the other hand, the other classes in study area have low soil fertility level. Moreover, their results were then used to predict the LULC in 2026 as shown in Table 5 and its result was a ROC value of 0.9160.

The results provide the possibility of a LULC change, especially urban growth between 2014 and 2026 was increased 25,064.45 Rai (4,010.31 Hectare). The distribution of LULC pattern of Phuket Island in 2026 was shown in Figure 6. It can be seen that urban growth has shown a pattern similar to the change between 1995 and 2002 due to population growth, the continued increase in tourism. The settlements are built on the low level of elevation and slope areas because of the flat area and near the beach are more expensive but it was the tourist attraction.

4. Conclusions

The optimum geospatial models for the LULC prediction was successfully examined and identified based on an accuracy assessment. Furthermore, it appears that integration of geospatial model for LULC prediction and LRA method for the urban land use change vulnerability analysis can be used as an effective tool for urban growth and expansion areas. A possible solution would be to combine a logistic regression model with CA-Markov model such that the predicted probability values can participate in the formulation of the rules in the dynamic simulation. LULC prediction maps of Phuket Island clearly demonstrate the future trend of urban growth. It can be seen that urban growth will occur around slope which has less angle and elevation is flat with moderate soil fertility. Other demographic trends affecting urban growth worldwide include the relatively fast population growth, the continued increase in tourism. The settlements are built on the low level of elevation and slope areas because of the flat area and near the beach are more expensive but it was the tourist attraction. In this study, the CA-Markov with LRA was recognized as the optimum geospatial models for specific urban growth and it was applied to predict the LULC in 2026. The results provide the possibility of urban land use change, especially used to fulfil for urban planning of the Phuket Office of Public Work and Town and Country Plan.

Acknowledgements

I am particularly thankful to Faculty of Science and Fisheries Technology, Rajamangala University of Technology Srivijaya, Trang Campus for providing the scholarship for this study.

References

Arsanjani, J. J., Helbich, M., Kainz, W. and Boloorani, A. D., 2013, Integration of Logistic Regression, Markov Chain and Cellular Automata Models to Simulate Urban Expansion. *Int. J. Appl. Earth Obs.*, Vol. 21, 265-275.

Congalton, R. G. and Green, K., 2009, Assessing the Accuracy of Remotely Sensed Data. *Principles and Practices. 2nd ed. CRC Press,* Boca Raton, FL, USA, 192.

Department of Provincial Administration, 2013, Number of the Population of Phuket Province between 2003 and 2012. Bangkok, Thailand: *Ministry of Interior.* Available from: http://service.nso.go.th/nso/web/statseries/tables/58300_Phuket/1.1.3.xls.

Gontier, M., Mortberg, U. and Balfors, B., 2009, Comparing GIS-based Habitat Models for Applications in EIA and SEA. *Environ. Impact Assess.*, Vol. 30(1), 8-18.

Herold, M., Gunter, M. and Keith, C. C., 2003, Spatiotemporal Form of Urban Growth: Measurement, Analysis and Modeling. *Remote Sensing of Environment*, Vol. 86, No. 3, 286-302.

Hu, Z. and Lo, C. P., 2007, Modeling Urban Growth in Atlanta using Logistic Regression. *Comput., Environ. and Urban Systems*, Vol. 31, 667–688.

Jantz, C. A., Scott, J. G. and Mary, K. S., 2004, Using the SLEUTH Urban Growth Model to Simulate the Impacts of Future Policy Scenarios on Urban Land Use in the Baltimore-Washington Metropolitan Area. *Environment and Planning B: Planning and Design*, Vol. 31(2), 251-271.

Jones Lang LaSalle, 2013, OnPoint • Spotlight on Thailand - Hotel Investment Market. *Hotels & Hospitality Group, Jones Lang LaSalle IP, Inc.*, 8.

Ongsomwang, S. and Boonchoo, K., 2016, Integration of Geospatial Models for the Allocation of Deforestation Hotspots and Forest Protection Units. *Surranaree J. Sic. Technol.* Vol. 23(3), 283-307.

Ongsomwang, S. and Saravisutra, A., 2011, Optimum Predictive Model for Urban Growth Prediction. *Suranaree J. Sci. Technol.* Vol. 18(2), 141-152.

Paegelow, M. and Camacho Olmedo, M. T., 2005, Possibilities and Limits of Prospective GIS Land Cover Modelling - A Compared Case Study: Garrotxes (France) and Alta Alpujarra Granadina (Spain). *Int. J. Geogr. Inf. Sci.* Vol. 19(6), 697-722.

Sae-Tan, A., 2013, Phuket Tourism: Is it Sustainable for Phuketians? Available from: http://csrasia.com/csr-asia-weekly-news-detail.php?id=12108. Accessed date: 28 May 2014.