

URBAN GROWTH MODELING OF PHNOM PENH, CAMBODIA USING SATELLITE IMAGERIES AND A LOGISTIC REGRESSION MODEL

Kompheak Mom and Suwit Ongsomwang*

Received: August 04, 2016; Revised: August 23, 2016; Accepted: August 24, 2016

Abstract

Phnom Penh City is facing rapid population growth with Cambodia having the second highest urban expansion rate in Asia, and it has encountered poor urban planning that results in urban sprawl and the loss of natural areas and agricultural land. To solve this problem, spatial and temporal dynamic driving factors for land use and land cover change should be well understood for enhancing urban planning. The specific objectives of the study are (1) to assess the land cover status and its change; (2) to employ a logistic regression (LR) model to discover the driving factors for urban growth; and (3) to predict the future urban growth pattern of Phnom Penh in 2030. Four main components of the research methodology are here conducted comprising (1) data collection; (2) data preparation; (3) model simulation and validation; and (4) urban growth prediction.

Results showed that the urban and built-up areas have continuously increased from 2002 to 2015 resulting in a major decline of arable land, vegetation and water bodies, while miscellaneous land was shown as fluctuating. Meanwhile, the pattern of urban growth expanded towards the southern, northern, and western areas of Phnom Penh during 2002-2009 with all types of growth including infill growth, expansion growth, linear branch, and isolated growth. However, during 2009-2015, the urban growth pattern occurred in all directions with expansion growth, clustered branch, and isolated growth. The driving factors for urban growth from the LR model for the 2002-2015, 2002-2009, and 2009-2015 periods varied according to the urban and built-up area pattern and time. Two common driving factors under the top 5 dominant factors, namely distance to the existing urban cluster and distance to an industrial area, showed a negative correlation with urban growth in the 3 periods. In addition, the final urban growth pattern from the LR model of the 3 periods showed a good result for overall accuracy at about 91%, 96%, and 94% a successful fit of urban allocation at about 58%, 54%, and 44%, and a relative operating characteristic at about 0.90, 0.95, and 0.90, respectively. Finally, the urban growth pattern prediction for Phnom Penh in 2030 from the optimum LR model of the 2002-2015 period with the estimated urban growth area using the Markov chain model revealed that urban

School of Remote Sensing, Institute of Science, Suranaree University of Technology, Nakhon Ratchasima 30000, Thailand. E-mail: suwit@sut.ac.th

** Corresponding author*

growth tends to take place in the north, south, and east around the existing urban clusters, and some are expected to occur along the major roads and ring roads.

In conclusion, the integration of the LR model, remote sensing, and GIS can be efficiently used as a tool to understand the driving factors and to predict the future urban growth pattern of Phnom Penh. It contributes significant information as a guideline for planners and decision makers.

Keywords: Urban growth modeling, logistic regression model, land cover classification, Phnom Penh, Cambodia

Introduction

Global urbanization has become a major issue in urban land use planning; as a result, much research in recent years has been conducted on urban growth modeling in order to achieve sustainable urban development (Achmad *et al.*, 2015). At the same time, there are many major problems related to the process of urbanization that result in reducing land use and in land cover becoming an impervious surfaces; the process of urbanization is caused by economic development and population growth (Shalaby and Tateishi, 2007). Bhatta (2010) stated that the impact of human population growth may result in a negative consequence that generally leads to uncontrolled urban sprawl and serious environmental problems. The outcome of urban sprawl may appear in the loss of precious land and resources, landscape alteration, environmental pollution, traffic congestion, infrastructure pressure, rising taxes, and neighborhood conflict (Allen and Lu, 2003; Aguayo *et al.*, 2007). Therefore, an urban planning policy should be conceived to control urban expansion in a sustainable way, to reduce the negative impact of urban sprawl, and to prevent the city from growing improperly (Shu *et al.*, 2014; Achmad *et al.*, 2015; Manu *et al.*, 2015). In addition, the driving factors behind urban expansion provide very important information, and they should be addressed and identified as a guideline for urban planners and decision makers for sustainable urban planning in the future. Since the driving factors can help to improve the understanding of the spatial and temporal dynamic process of land use and land cover change from the past to the future, by analyzing the relationship between them and the historical urban growth pattern, the future urban growth pattern can be predicted and the

direction of urban growth can be identified (Aguayo *et al.*, 2007; Manu *et al.*, 2015).

Concerning the models for identifying the driving factors of urban growth, the 3 types of them are the rule-based simulation, system dynamic simulation, and empirical estimation models (Hu and Lo, 2007). The rule-based simulation model that includes cellular automata (CA) will simulate the spatial pattern of urban growth based on the biophysical factors; however, it cannot integrate with human drivers (Hu and Lo, 2007). Meanwhile, the system dynamic simulation model can work well with CA, but it is poor in identifying the driving factors (Shu *et al.*, 2014). Conversely, the empirical estimation model which includes the logistic regression (LR) model, Markov chain analysis, multiple regression analysis, factor analysis, and principle component analysis, has been proved as an efficient model to figure out the influences that cause land use change. Particularly, the LR model allows researchers to find the relationship between dependent variables (urban and non-urban) and independent variables (factors relating to the environment, society, proximity and accessibility, and neighborhood) in order to understand influential factors in shaping urban growth patterns in the past in order to predict the future. Studies on urban growth using the LR model have been successfully implemented in many cities by many researchers including Allen and Lu (2003); Aguayo *et al.* (2007); Hu and Lo (2007); Duwal (2013); Alsharif and Pradhan (2014); Shu *et al.* (2014); and Achmad *et al.* (2015).

The developing country of Cambodia has recently been revealed as a country with one of the fastest urbanization processes in Asia,

according to the United Nations Population Fund (2014). A report from the National Institute of Statistics (2013) also revealed that over the last 2 decades, the urban population of Cambodia grew from 17.7% of the total population in 1998 to 21.4% of the total population in 2013. The annual urban growth rate which was 2.2% between 1998 and 2008 increased to 3.7% between 2008 and 2013. This indicates the trend of an increasing urban population on a large scale and it may be expected that the trend will accelerate in the future. A result from the United Nations Population Fund (2014) revealed that the urban population of Cambodia might reach 3.5 million or 22.1% of the total population in 2020 and will reach 5.4 million in 2030, particularly in the major city of Phnom Penh. Besides the population growth, the urban expansion rate of Cambodia was found to be the second fastest rate in Asia, about 4.3% annually (The World Bank, 2015). This shows that Cambodia's urbanization is expanding at a high rate, especially in Phnom Penh. Doyle (2012) stated that Phnom Penh has encountered a problem with poor planning as well as an ineffective master plan and lack of official land use zoning. Consequently, the city has faced many negative impacts, for instance the widespread expansion of the urban sprawl in natural areas (e.g. lakes) surrounding the core existing urban areas and agricultural land in

the suburban areas. The ineffective master plan and lack of land use zoning accompanied with the fastest expansion rate in Phnom Penh may inevitably lead to haphazard sprawl that will have more of an environmental impact on the city in the future. In addition, little attention has been paid to the modeling of the future urban growth pattern in Phnom Penh which is facing a high urban expansion rate. Therefore, the LR model is here chosen to identify the significant influential factors that have driven urban growth from the past to recent periods in order to predict the future Phnom Penh urban growth pattern. The specific objectives of this study are (1) to assess the land cover status and its change; (2) to employ the LR model to discover the driving factors of Phnom Penh's urban growth; and (3) to predict the future urban growth pattern of Phnom Penh in 2030.

Material and Method

Study Area

Phnom Penh which is the capital city of Cambodia and the largest urbanized area has been recognized as the main national economic and industrial center of the country; it is located on an embankment at the confluence of the Tonle Sap and Mekong Rivers (Figure 1). The city is situated at latitude $11^{\circ} 33' N$ and longitude $104^{\circ} 55' E$ with a total area of 678.5

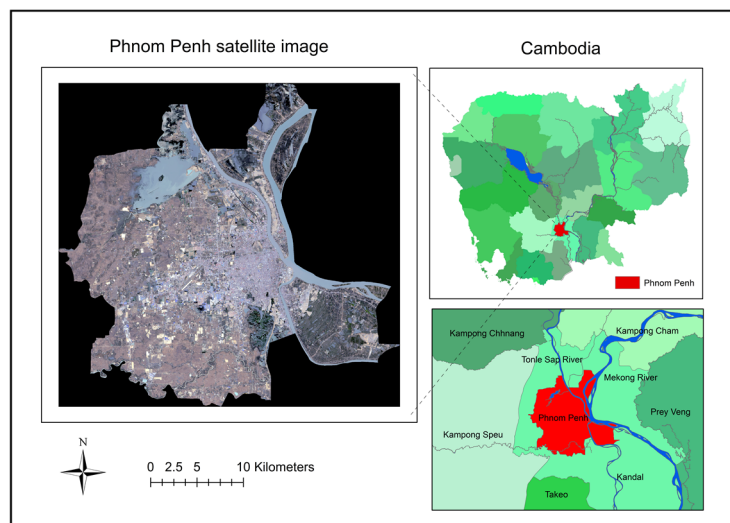


Figure 1. Map and location of Phnom Penh city

km² and provides homes for about 1.6 million of Cambodia's total population of 14.8 million (National Institute of Statistics, 2013).

Research Methodology

The research methodology and its workflow consists of 4 components: (1) data collection; (2) data preparation; (3) model simulation and validation; and (4) urban growth prediction (Figure 2).

Component 1: Data Collection

There are 2 types of data for urban growth modeling, satellite image data and the driving factors for urban growth which are

here collected as summarized in Tables 1 and 2, respectively.

Component 2: Data Preparation

Two major activities in data preparation are implemented; first, the land cover classification and change detection, and secondly, the extraction of the urban growth driving factors.

Land Cover Classification and Change Detection

The maximum likelihood classifier that has been widely used under the supervised classification approach (Jensen, 2005) is firstly

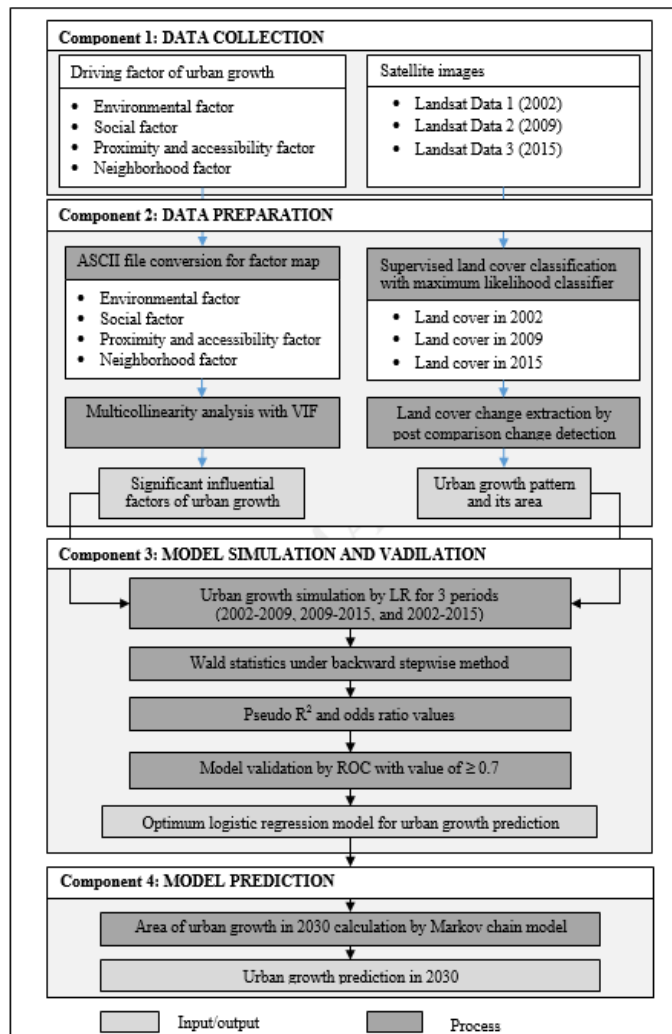


Figure 2. Schematic workflow of research methodology framework

Table 1. Basic information of satellite imageries

Satellite Images	Path	Row	Resolution	Source	Date	Application
Landsat 7 ETM+	126	52	30 m	USGS	3 January 2002	Land cover extraction
Landsat 5 TM	126	52	30 m	USGS	14 January 2009	Land cover extraction
Landsat 8 OLI	126	52	30 m	USGS	15 January 2015	Land cover extraction
Google Earth image			1.6 m	DigitalGlobe Satellite	3 February 2003	Accuracy assessment
Google Earth image			1.6 m	DigitalGlobe Satellite	7 January 2010	Accuracy assessment
Google Earth image			1.5 m	Astrium/ Spot image	31 October 2015	Accuracy assessment

Table 2. Driving factor dataset for urban growth modeling

No	Data collection	Year	Scale	Source
1	Phnom Penh boundary, District and Commune.	2011	1:125,000	Economic census of Cambodia in 2011, National Institute of Statistics (NIS) (2013)
2	National main roads and minor roads from topographic map.	2002	1:100,000	Cambodia National Geographic Department (NGD) (2002)
	Updated main roads and minor roads	2010 and 2015		Google Earth/ high resolution images
3	Water bodies consisting of Tonle Sap and Mekong rivers, lakes, and streams.	2002	1:100,000	Cambodia (NGD, 2002)
	Updated water bodies in the study area	2009 and 2015		Landsat 5 and 8 images
4	Public schools, health centers, and airport.	2002	1:100,000	Cambodia (NGD, 2002)
	Updated public schools, health centers, and airport.	2015		Openstreetmap.org
5	Urban and built up areas, vegetation, water bodies, and miscellaneous land	2002	1:100,000	Cambodia (NGD, 2002) and Landsat 7
	Updated land cover types	2009 and 2015		Landsat 5 and 8 images
6	Population density map	1998 and 2008		General population census 1998 and 2008, (NIS 2002; NIS, 2009)
7	Location of industrial areas in 2002	2002		White Book of Development and Management of Phnom Penh, 2020 (Bureau des Affaires Urbaines, 2007)
	Updated location of industrial areas in 2009 and 2015	2009 and 2015		Google Earth image interpretation

applied for land cover extraction in 2002, 2009, and 2015 from multi-temporal satellite imageries. In this study, the land cover classification scheme is modified based on the land use classification system of the Cambodia National Geographic Department which consists of (1) urban and built-up areas, (2) arable

land, (3) vegetation, (4) water bodies and (5) miscellaneous land (see the descriptions in Table 3). Then, each classified land cover map is evaluated for its accuracy (overall accuracy and Kappa hat coefficient) with correspondent reference maps: the Cambodia national topographic map, scale 1:100000, for the year

Table 3. Land cover classification scheme and description

Class	Land cover types	Description
1	Urban and built-up areas	Village, town, residential, commercial, and industrial area, communication and utility center and transportation infrastructure.
2	Arable land	Paddy field, abandoned paddy field, and open land.
3	Vegetation	Crop land, orchard land, grassland, shrub land, flooded grassland, flooded shrub land, garden, park, garden village, and forest.
4	Water bodies	River, stream, pond and lake.
5	Miscellaneous land	Bare land, marsh and swamp, sand bank, excavation site, and landfill

2002 data and the existing high spatial resolution Google Earth map for the years 2010 and 2015 data. The number of sample points is calculated based on the binomial probability distribution theory suggested by Fitzpatrick-Lins (1981) and these sample points are allocated using a stratified random sampling scheme.

After that, 3 periods of land cover change: 2002-2009, 2009-2015, and 2002-2015 are extracted using a post classification comparison algorithm. The results of the land cover change detection are further used to extract the area of urban growth and its pattern.

Urban Growth Driving Factors Extraction

The collected urban growth driving factors in the form of maps and statistical data, prepared as independent variables for the LR model, are firstly converted to ASCII files for multicollinearity analysis under the IBM SPSS statistics software. In principle, multicollinearity analysis is very important for logistic regression modeling in order to prevent the perfect linear relationships among independent variables as driving factors (Table 4). Therefore, the estimation of the LR coefficients possibly can be computed without bias (Odzemir, 2011). According to Rogerson (2010), the variance inflation factor (VIF) is here used to detect multicollinearity as:

$$VIF = \frac{1}{1 - r^2} \quad (1)$$

where r^2 is associated with the regression of the independent variable on other independent variables. To avoid multicollinearity, all

independent variables were calculated for the VIF to examine the correlation among them by linear regression analysis. As a result, if an independent variable gets a VIF value equal to or more than 10, it will be excluded from the LR model (Rogerson, 2010).

For running the simulation of the LR model, the input data (urban growth area and driving factors) for 3 different years - 2002, 2009, and 2015 - are prepared under the ArcGIS environment and converted into raster format at a resolution of 30 m, the same as the classified land cover map. Then, the driving factor which is continuous data is normalized into the range of 0-1 in order to get the distribution of the spatial value within the same scale (Table 4).

Component 3: Model Simulation and Validation

Urban Growth Simulation by the Logistic Regression Model

The derived significant driving factors of urban growth as independent variables (x) and the urban growth area as a dependent variable (y) after multicollinearity analysis are used as input data for the LR model. The LR model is used to associate the urban growth with the driving factors and to generate an urban growth probability map (Cheng and Masser, 2003; Hu and Lo, 2007; Rogerson, 2010; Duwal, 2013; Alsharif *et al.*, 2013; Shu *et al.*, 2014). The general form of the LR equation can be calculated as below:

$$P = \frac{e^y}{1 + e^y} = \frac{e^{a+b_1x_1+b_2x_2+\dots+b_mx_m}}{1 + e^{a+b_1x_1+b_2x_2+\dots+b_mx_m}} \quad (2)$$

Table 4. Driving factors used in logistic regression model

Driving Factor	Abbrev.	Types	Reference
Environmental factor			
1:vegetation, 0:not vegetation ^(a)	Veget	Category	Hu and Lo (2007); Shu <i>et al.</i> (2014)
1:miscellaneous land, 0:not miscellaneous land ^(a)	Miscell	Category	Hu and Lo (2007); Shu <i>et al.</i> (2014)
1:arable land, 0:not arable land ^(a)	Arable	Category	Hu and Lo (2007); Shu <i>et al.</i> (2014)
Distance to river ^(b)	DistRiver	Continuous	Cheng and Masser (2003); Braimoh and Onishi (2007); Shu <i>et al.</i> (2014); Achmad <i>et al.</i> (2015)
Social factor			
Population density ^(a)	DenPop	Continuous	Hu and Lo (2007); Braimoh and Onishi (2007); Dewan and Yamaguchi (2009); Alsharif and Pradhan (2013); Achmad <i>et al.</i> (2015)
Proximity and accessibility factor			
Distance to CBD ^(c)	DistCBD	Continuous	Cheng and Masser (2003); Hu and Lo (2007); Alsharif and Pradhan (2013); Achmad <i>et al.</i> (2015)
Distance to health center ^(a)	DistHealth	Continuous	Eyoh <i>et al.</i> (2012); Duwal (2013)
Distance to public school ^(a)	DistSchool	Continuous	Eyoh <i>et al.</i> (2012); Duwal (2013); Alsharif and Pradhan (2013)
Distance to main road ^(c)	DistMroad	Continuous	Cheng and Masser (2003); Hu and Lo (2007); Braimoh and Onishi (2007); Alsharif and Pradhan (2013); Shu <i>et al.</i> (2014); Achmad <i>et al.</i> (2015)
Distance to road ^(a)	DistRoad	Continuous	Cheng and Masser (2003); Eyoh <i>et al.</i> (2012)
Distance to railway line ^(b)	DistRailLine	Continuous	Braimoh and Onishi (2007); Eyoh <i>et al.</i> (2012)
Distance to Phnom Penh airport ^(b)	DistAirport	Continuous	Cheng and Masser (2003)
Distance to industrial area ^(a)	DistIndustry	Continuous	Braimoh and Onishi (2007); Eyoh <i>et al.</i> (2012); Shu <i>et al.</i> (2014)
Distance to industrial river port ^(b)	DistIndPort	Continuous	
Neighborhood factor			
Number of urban cells within a 7x7 cell window ^(a)	Num_Urban	Continuous	Cheng and Masser (2003); Hu and Lo (2007); Shu <i>et al.</i> (2014)
Distance to existing urban cluster ^(a)	DistUrban	Continuous	Cheng and Masser (2003); Hu and Lo (2007); Alsharif and Pradhan (2013)

Note (a) These independent variables are different for all dates.

(b) These independent variables remain the same for all dates.

(c) These independent variables are different only in 2002 and 2009, while independent variables of 2015 remain the same as the variables of 2009.

where $x_1, x_2, \dots, x_m$ are independent variables, y is the urban growth area as a dependent variable in non-linear function form, $b_1, b_2, \dots, b_m$ are the regression coefficients, and P is the probability of occurrence (0-1).

After running the LR model, the Wald statistic, which is used to test the significance of every LR coefficient (b) to the model's outcome (Menard, 2002; Rogerson, 2010), is examined as:

$$Wald = \frac{b^2}{SE_b^2} \quad (3)$$

where b is a coefficient of the independent variable (x) and SE_b is a standard error. If the value of b is more than twice the value of SE_b , it may be regarded as significantly different from zero. Therefore, the independent variable

(x) can be considered as important to explain the relationship with the dependent variable (y) (Rogerson, 2010). In this study, the backward stepwise method is applied to identify the significance of independent variables based on the P-value obtained at 0.05 level of confidence.

Likewise, the odds ratio (OR), which is applied to interpret the degree of influence of independent variables, is calculated by the exponential of coefficient (b) as below:

$$OR = e^b \quad (4)$$

Moreover, the pseudo Nagelkerke R^2 , which is used to interpret the predictive ability and fitting degree of the model, is also examined from the statistical report. Basically, if the pseudo R^2 value ranges between 0.2 and 0.4, it is considered as a good fit (Clark and Hosking, 1986; Hu and Lo 2007; Shu *et al.*, 2014). The Nagelkerke pseudo R^2 ($f_{(x)}$) has the formula as below:

$$f_{(x)} = \frac{1 - \exp\left(\frac{-2(LL_m - LL_0)}{N}\right)}{1 - \exp\left(\frac{2LL_0}{N}\right)} \quad (5)$$

where LL_m and LL_0 are the log-likelihood of the model and intercept, respectively, and N is the sample size.

Model Validation

Two of the main techniques include the relative operating characteristic (ROC) and contingency table that are recommended by many researchers (Pontius and Schneider, 2001; Hu and Lo, 2007; Duwal, 2013; Arsanjani *et al.*, 2013) and were here used to evaluate the performance of the LR model.

Basically, the ROC compares the simulated urban map with the reference urban map by measuring the relationship of agreement of the pair of images that are urbanized. The area under the ROC curve varies from 0.5 to 1, where 0.5 is a random assignment of probability and 1 is a perfect assignment of probability (Hu and Lo, 2007; Alsharif *et al.*, 2013).

Meanwhile, the consistency table technique compares the simulated urban growth map, which is created by allocating the urban

growth areas to the probability map of the LR model from the highest probability cell to the lowest probability cells, with the actual urban growth map for accuracy assessment (overall accuracy, omission and commission errors) (Duwal, 2013).

Component 4: Urban Growth Prediction

The weak points of the LR model are quantification of change and temporal analysis since the model provides only the probability area and direction of the growth. Therefore, measuring the change demand for urban growth is required in order to provide the pattern of urban growth for the future. According to Arsanjani *et al.* (2013), there are 2 methods to estimate the change demand area for the future, the Markov chain model and statistical extrapolation based on the numbers of the population. However, it is not quite logical and reasonable for predicting the future urban pattern based only on the numbers of the population due to the preference of property developers to build high-rise constructions rather than one-family houses for maximum profit (Arsanjani *et al.*, 2013). For this reason, this study uses the Markov chain model to estimate the predicted demand area of urban growth in 2030 based on the 13 years' period of transition area between the 2002 and 2015 land cover classifications. After that the predicted demand area is further used for allocation to the derived future urbanization probability map from the LR model in order to generate the urban growth pattern of Phnom Penh in 2030.

Results and Discussion

Land Cover Classification and Its Change and Accuracy Assessment

The derived land cover information of 2002, 2009, and 2015 is comparatively summarized in Table 5 while the distributions of the land cover maps of the 3 years are presented in Figure 3. From the results, arable land (the majority is paddy fields) in 2002 was the predominant land cover class that consisted of 292.39 km² while the urban and built-up areas covered about 38.13 km². However, the urban and built-up areas in 2009 had increased by about 2 times the urban area in 2002 and

expanded westward along the main road to the airport and to the north and south of Phnom Penh; there were few such areas to the east because of frequent flooding in rainy the season over the wetland. In fact, the increased urban and built-up areas mainly were created from arable land and vegetation. In addition, miscellaneous land (most of it is bare land for construction sites and development activity) increased from 8.29 km² in 2002 to 45.16 km² in 2009 because of the future urban development like residential buildings and industrialization. These activities can offer more jobs and attract more rural immigrants.

Consequently, the urban and built-up areas during the period 2009 to 2015 increased,

largely from the miscellaneous land and vegetation in 2009, with an area about 38.42 km². However, miscellaneous land decreased and remained at about 26.91 km² which indicated that the development of residential buildings, satellite cities, and industrial enterprises was in progress by transforming some arable land, vegetation, and water bodies. Meanwhile, vegetation decreased from 213.13 km² to 186.56 km², and water bodies decreased greatly from 132.18 km² to 88.54 km² due to land being infilled for development projects such as satellite cities, residential areas, and recreation centers. According to Doyle (2012), these enormous projects caused the loss of a network of wetlands: Boeung Kak

Table 5. Area and percentage of land cover classes during 2002-2015

Land cover type	2002		2009		2015	
	km ²	%	km ²	%	km ²	%
Urban and built up areas	38.13	6	70.31	10	108.73	16
Arable land	292.39	43	225.88	33	275.92	40
Vegetation	215.28	31	213.13	31	186.56	27
Water bodies	132.57	19	132.18	19	88.54	13
Miscellaneous land	8.29	1	45.16	7	26.91	4
Total area	686.66	100	686.66	100	686.66	100

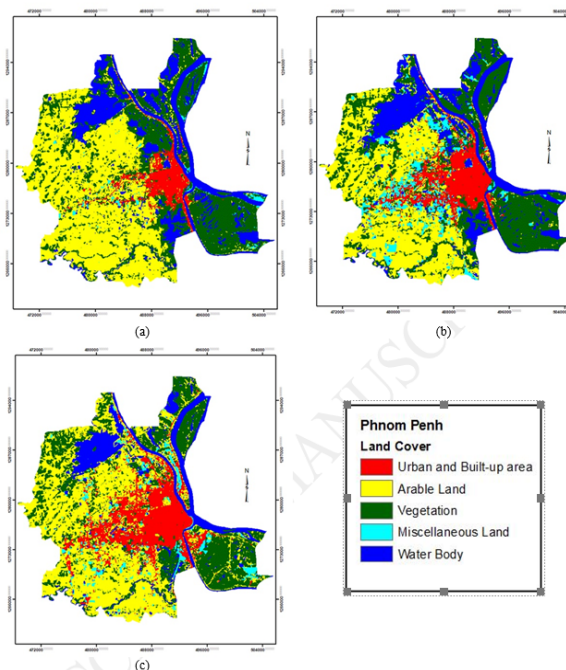


Figure 3. Distribution of land cover pattern (a) in 2002, (b) in 2009, and (c) in 2015

Lake, streams, and ponds surrounding Phnom Penh. Arable land showed an increase due to the El Nino effect with a dry and hot climate on in 2015. There was no water in the rice fields in the northern part of Phnom Penh, as presented in the Landsat images in 2015. In fact, when water flows out of the wetland into the large lake, the rice fields become dry and give a spectral value similar to arable land. Moreover, the increase of arable land in 2015 also came from the disappearance of water that was caused by infilling the lake for the development of satellite cities and residential buildings in the low lying areas of central and eastern Phnom Penh.

In addition, the accuracy assessment result of 3 classified images with 203 stratified random points is summarized in Table 6 and included the overall accuracy and the Kappa hat coefficient. It revealed that the Kappa hat coefficient of the classified land cover map in 2002, 2009, and 2015 is 88.60%, 88.11%, and 89.13%, respectively. Landis and Koch (1977) stated that Kappa hat coefficient values $> 80\%$

represent a strong agreement between the classification and reference points, and so are acceptable for performing a quantitative analysis.

Development of Urban Growth Pattern

The urban growth pattern that occurred in the periods between 2002 and 2009 and 2009 and 2015 was here detected using overlay analysis. According to Wilson *et al.* (2003), there are 3 main types of spatial pattern of urban growth: infill growth, expansion growth, and outlying growth (linear branch, clustered branch, and isolated growth).

It was found that Phnom Penh's urban growth pattern between 2002 and 2009 illustrated all 3 types of growth patterns, as shown in Figure 4(a). Most of the infill growth areas were situated inside the existing urban area of 2002 and located in some parts to the north of Phnom Penh where most of the conversion came from arable land. Expansion growth dispersed at the existing urban outer edge and in some areas near the airport and industrial areas. Moreover, the linear branch

Table 6. Accuracy assessment of classified land cover maps

Land cover images	2002	2009	2015
Overall accuracy (%)	92.16	91.18	92.12
Kappa hat coefficient (%)	88.60	88.10	89.10

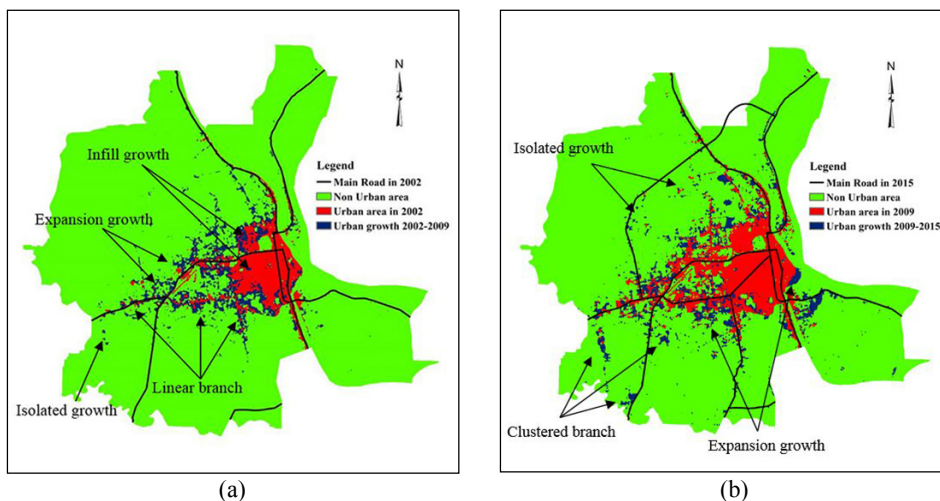


Figure 4. Spatial pattern of urban growth during: (a) 2002 to 2009 and (b) 2009 to 2015

was depicted as a significant growth pattern of Phnom Penh since it was noticeably found along the main road to the airport and national road Number 4 to Kompong Speu province. Isolated growth was also found in some specific areas that were located far from the existing urban clusters, since they were mostly industrial sites and the Phnom Penh Special Economic Zone (PPSEZ) that are located in the outskirts and rural areas.

Figure 4(b) revealed that the major pattern of urban growth between 2009 and 2015 can be identified into 3 types: expansion growth, clustered branch, and isolated growth. The expansion growth pattern was found in the whole of Diamond Island satellite city and some areas around industrial sites. Furthermore, it should be noticed that some isolated growth areas during 2002-2009 became clustered branches because construction of some industrial sites was completed, particularly in the PPSEZ and other industrial factories. Isolated growth also occurred on the northern part of the ring road where factories accounted for most of the growth.

Multicollinearity Effects and Input Data for LR Model

The result showed that the VIF value of 2 independent variables of 2002 and 2009, distance to industrial river port (DistIndPort) and distance to river (DistRiver) were greater than 10; therefore, these 2 variables were removed from the LR model due to being highly correlated with other independent variables. Table 7 shows the remaining 14 variables and their VIF values that indicate no multicollinearity effect. Thus, all 14 variables are safely used for further analysis in the LR model.

Figure 5 displays urban growth areas as a dependent variable (y) for the LR model of the 2002-2009, 2009-2015, and 2002-2015 periods, while Figure 6 demonstrates an example of the driving factors as independent variables (x) for the LR model of the 2002-2015 period.

Calibration of LR Model on Urban Growth in 3 Periods

The results of 3 LR calibration equations for the long-term period of 2002-2015 and short-term periods of 2002-2009 and 2009-2015 are summarized in Table 7 which indicates

Table 7. LR calibration equation and its statistics of 3 model periods: 2002-2009, 2009-2015, and 2002-2015

No	Variables	Model 2002-2015					Model 2002-2009					Model 2009-2015				
		b	SE _b	Wald	OR	VIF	b	SE _b	Wald	OR	VIF	b	SE _b	Wald	OR	VIF
1	Arable	1.656	0.019	7,569.18	5.241	3.047	3.725	0.042	8,030.98	41.462	3.047	1.321	0.025	2,823.380	3.747	2.586
2	Veget	1.690	0.018	9,323.73	5.419	2.020	3.801	0.042	8,358.28	44.761	2.020	1.834	0.023	6,335.010	6.260	1.943
3	Miscell	1.984	0.036	3,107.97	7.274	1.067	3.465	0.06	3,340.28	31.987	1.067	2.979	0.024	15,008.18	19.668	1.388
4	DenPop	1.123	0.055	415.49	3.075	2.253	1.378	0.073	358.709	3.968	2.253	-----	-----	-----	-----	2.013
5	DistCBD	-3.351	0.065	2,666.87	0.035	9.830	-1.468	0.109	180.162	0.230	9.830	-2.973	0.061	2,408.050	0.051	6.704
6	DistHealth	-0.448	0.059	57.646	0.639	7.005	-3.175	0.107	873.835	0.042	7.005	1.943	0.066	857.692	6.982	6.305
7	DistSchool	-0.439	0.044	98.855	0.645	1.364	-1.686	0.068	616.011	0.185	1.364	-0.322	0.068	22.118	0.725	1.487
8	DistAirport	-4.155	0.044	8,857.58	0.016	3.758	-4.829	0.070	4,698.22	0.008	3.758	-0.159	0.048	11,040	0.853	4.126
9	DistIndustry	-3.624	0.091	1,600.03	0.027	7.620	-6.32	0.170	1,389.51	0.002	7.620	-2.782	0.081	1,193.390	0.062	2.726
10	DistMroad	-2.012	0.050	1,602.16	0.134	1.609	0.502	0.084	35.88	1.652	1.609	-0.478	0.062	58.669	0.620	3.587
11	DistRoad	-3.965	0.052	5,817.87	0.019	1.473	-4.506	0.091	2,471.61	0.011	1.473	-1.095	0.082	176.184	0.335	1.427
12	DistRailLine	0.662	0.039	289.141	1.938	2.233	0.333	0.068	23.827	1.395	2.233	-1.304	0.049	711.739	0.272	3.179
13	DistUrban	-2.268	0.090	634.178	0.103	5.745	-10.565	0.232	2,069.84	0.000026	5.745	-20.095	0.223	8,146.130	0	5.471
14	Num Urban	-2.129	0.032	4,412.37	0.119	2.290	0.049	0.049	43.008	1.383	2.290	-1.099	0.032	1,180.700	0.333	2.392
	Constant	0.920	0.026	1,260.48	2.510		-1.500	0.054	764.741	0.223		-1.499	0.032	2,154.480	0.223	

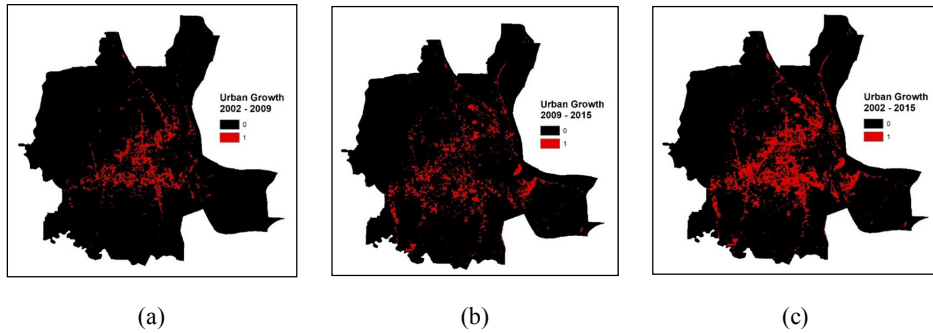


Figure 5. Urban growth areas as dependent variable for LR model in each period: (a) 2002-2009, (b) 2009-2015, and (c) 2002-2015

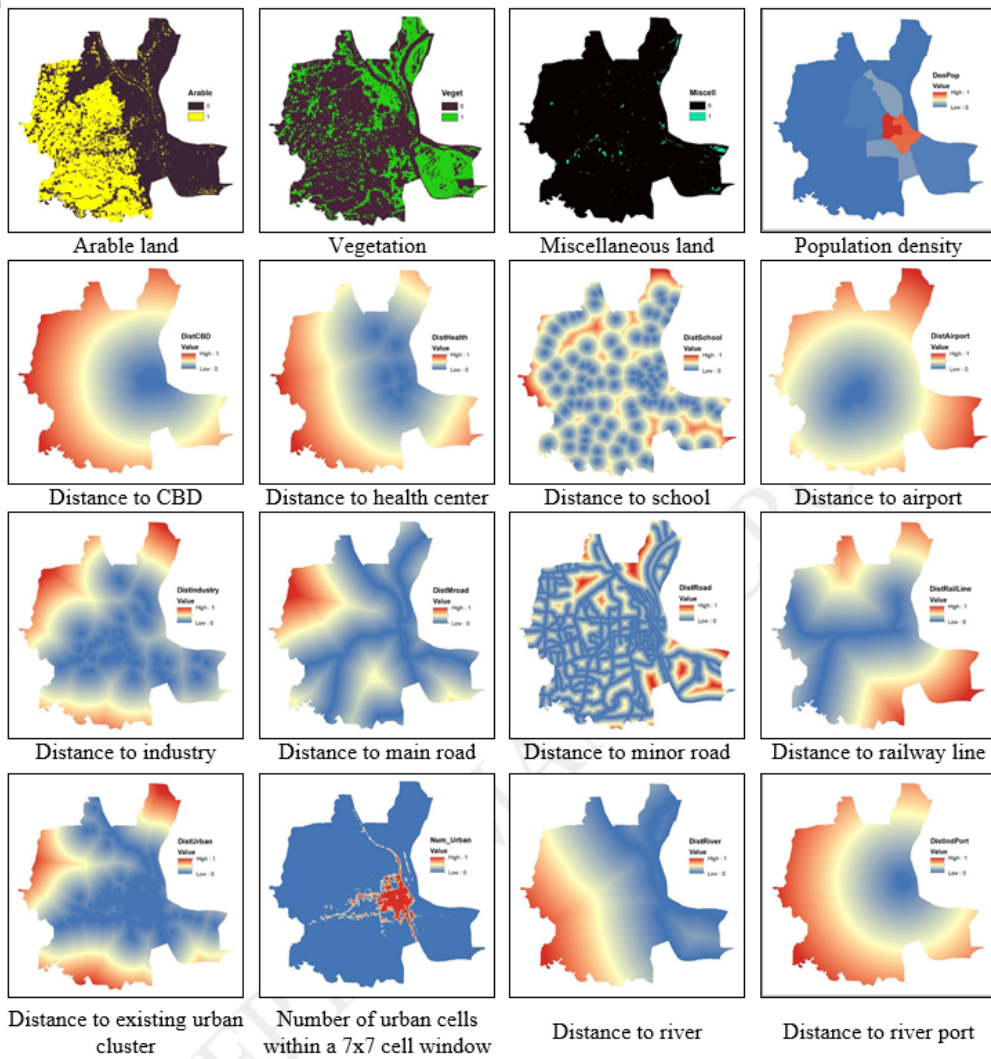


Figure 6. Driving factors for LR model of 2002-2015 period

the value of the coefficient b , SE_b , Wald, OR, and VIF statistics.

According to the Wald statistic for the significance of a single predictor in Table 7, the calibrated LR model of the 2002-2015 and 2002-2009 periods was obtained in the first iteration of the backward stepwise method with all 14 variables. However, the LR model of the 2009-2015 period was obtained in the second iteration of the backward stepwise method with only 13 variables. In addition, the pseudo Nagelkerke R^2 , which is calculated using Equation 5 under SPSS software, showed that the LR model of the 2002-2015 period provided a value of 0.439. Meanwhile, the pseudo Nagelkerke R^2 values of the LR model of the 2002-2009 and 2009-2015 periods were 0.5 and 0.37, respectively. These results indicate that all 3 models perform well, and all significant independent variables could be efficiently used to interpret the relationship with urban growth for the 3 periods.

Furthermore, the influence of the driving factors on urban growth based on their coefficients (b) of the LR model for the 3 periods was slightly different. For the LR model of the long-term period, arable land, vegetation, miscellaneous land, population density, and distance to railway line had positive influences on urban growth and had coefficient values equal to 1.656, 1.690, 1.984, 1.123, and 0.662 respectively, while the remaining 9 variables, distance to the central business district (CBD), health centers, public schools, Phnom Penh airport, industrial areas, main roads, roads, existing urban clusters, and number of urban cells within a 7×7 window cell had negative influences on urban growth with coefficient values of -3.351, -0.448, -0.439, -4.155, -3.624, -2.012, -3.965, -2.268, and -2.129, respectively.

For the LR model of the 2002-2009 period, there were 7 independent variables that showed positive influences on urban growth, and they had coefficients equal to 3.725, 3.801, 3.465, 1.378, 0.502, 0.333, and 0.324 for arable land, vegetation, miscellaneous land, population density, distance to main road, distance to railway line and number of urban cells within a 7×7 cell window, respectively. All other proximity factors including distance to the CBD, health centers, public schools, airport, industrial areas, roads, and the existing urban clusters had negative influences on urban growth, and had coefficient values equal to -1.468, -3.175, -1.686, -4.829, -6.32, -4.506, and -10.565, respectively. Meanwhile, the LR model of the 2009-2015 period had 4 factors that had positive influences on urban growth which were arable land, vegetation, miscellaneous land, and distance to health center with coefficient values of 1.321, 1.834, 2.979, and 1.943, respectively. There were 9 factors showing the negative influences on urban growth which were distance to the CBD, public schools, airport, industrial areas, main roads, roads, railway lines, and the existing urban clusters and number of urban cells within a 7×7 cell window with coefficient values of -2.973, -0.322, -0.159, -2.782, -0.478, -1.095, -1.304, -20.095, and -1.099, respectively.

In summary, the dominant driving factors of urban growth from the 3 LR models are here again synthesized and ranked for the top 5 important factors based on the coefficient value (b), as presented in Table 8. It was found that there were 2 common driving factors, namely distance to the existing urban clusters and distance to industrial areas that had high negative influences on urban growth in all 3 periods. However, the distance to airport and distance to roads factors were not stable

Table 8. Importance of influence of factors on urban growth in 3 different periods

Top five influential factors	Period of logistic regression model analysis		
	2002–2015	2002–2009	2009–2015
First rank	DistAirport	DistUrban_2002	DistUrban_2009
Second rank	DistRoad_2002	DistIndustry_2002	DistCBD_2009
Third rank	DistIndustry_2002	DistAirport	Miscell_2009
Fourth rank	DistCBD_2002	DistRoad_2002	DistIndustry_2009
Fifth rank	DistUrban_2002	Veget_2002	DistHealth_2009

influences for all periods, since they had an effect only in the 2002-2015 and 2002-2009 periods, and the influences changed to distance to the CBD and miscellaneous land in the 2009-2015 period.

In addition, the influences of the dominant variables on urban growth can be more explained by the OR value, as shown in Table 7. It revealed that the influence of distance to the existing urban clusters on urban growth for the short-term periods were higher than for the long-term period. Herein, the existing urban cluster of the LR model for the 2002-2009 period has the OR of 0.000026 or 1/38,461.5. This indicates that the area closer to the existing urban clusters has a higher possibility of urban growth by about 38,461.5 times than the areas farther from the existing urban clusters. This can be observed in Figure 4(a) where most of the urban growth pattern is the expansion growth type that is situated around the neighborhood of the urban areas in 2002. Meanwhile, the odds ratio of urban growth occurrence for the LR model of the 2009-2015 period in an area farther from the existing urban clusters in 2009 were almost zero, which also reflects that the odds of urban growth in an area closer to the existing urban cluster were extremely high. However, for the LR model of the long-term period, distance to the existing urban clusters was the fifth rank and did not have a high influence on the probability of urban growth, as indicated by the OR value of 0.103 or 1/9.7. The result is true for the long-term period, because the urban growth extensively spread from the existing urban clusters in 2002 to the airport that is situated further from the neighborhood of the urban areas in 2002.

Similarly, the distance to industrial areas had the OR equal to 0.027 or 1/37, 0.002 or 1/500, and 0.062 or 1/16.12 for the long- and short-term periods of 2002-2015, 2002-2009, and 2009-2015, respectively. These results indicate that the possibility of urban growth in areas 1 km or closer to the industrial areas is 37 times, 500 times, and 16.12 times greater than the possibility of urban growth in areas further than 1 km from industrial areas. Since the economics and labor force depend on the industrial sector the major part of which comprises garment factories, many urban housing facilities, other accommodation, shops,

restaurants, parks, and residential buildings for laborers were built around the garment factories along the main road in the suburban areas and in the region near the Phnom Penh international airport, and they became larger than the industry itself. This has demonstrated the pulling factor of urbanization in Phnom Penh, since it attracts more rural migration for job employment and quality of life.

During the long- and short-term periods of 2002-2015 and 2002-2009, urban growth was significantly influenced by the distance to the airport that indicated the OR of 0.016 or 1/62.5 and 0.008 or 1/125, respectively. This means that the odds of urban growth in an area 1 km or closer to the airport was 62.5 times and 125 times greater than the odds of urban growth in an area further than 1 km from the airport. This result obviously confirms that the trend of the urban growth pattern expanded westward to the airport which is located in the middle of Phnom Penh and in the low population density areas.

The result of the LR model of the 2002-2015 period also reveals that urban growth was influenced by road accessibility. The OR for distance to road was 0.019 or 1/52.63, which indicates that an area 1 km or closer to a road is estimated to be 52.63 times more likely to have urban growth than an area more than 1 km from the road. This contributed to the spatial pattern of the linear branch development, especially along the main road to the airport and some minor roads outside the existing urban areas in 2002 (Figure 4(a)).

The LR models of the 2002-2015 and 2009-2015 periods also show that urban growth was notably influenced by the distance to the CBD with the OR equal to 0.035 or 1/28.57 and 0.051 or 1/20, respectively. This result is true because the Phnom Penh CBD that is located in the core urban area had some available vacant land for development; hence, the vacant land closer to the CBD had a high probability to transform into urban and built-up areas, and it is estimated that the probability is 28.57 and 20 times higher than an area more than 1 km from the CBD in the period of 2002-2015 and 2009-2015, respectively. The conversion of vacant land to urban area can be observed by the occurrence of the infill growth, as shown in Figure 4(a). Moreover, another reason behind the CBD's influence on urban

growth is the development of the satellite city on Diamond Island that is situated close to the CBD. Hence, urban growth has occurred on the whole island by transforming vegetation land; the pattern of expansion growth on Diamond Island can be observed in Figure 4(b).

For the LR model of the 2009-2015 period, miscellaneous land became a very significant influence on urban growth, as shown in Table 7. It was found that the OR of each land cover are 3.75 for arable land, 6.26 for vegetation, and 19.70 for miscellaneous land. This means that the probability of urban growth occurrence on miscellaneous land is particularly high because miscellaneous land is bare land that is prepared for construction sites and urban development. However, it should be noticed that the OR of miscellaneous land during 2009-2015 is highly different from the OR of arable land and vegetation when comparing the difference between the odds ratio of each land cover type for the short- and long-term periods model of 2002-2009 and 2002-2015. This means that the development of the construction sector in Phnom Penh has boomed since 2009 resulting in the increase of the construction sites, like the PPSEZ, industrial parks, residential buildings (Borey), and satellite cities.

However, during the 2002-2009 period, the degree of influence of the vegetation, arable land, and miscellaneous land were almost the same, as indicated in Table 7 by the result of the OR values of 44.761, 41.462, and 31.987, respectively. This means that the probability of urban growth occurrence on the 3 land cover type are rather equal. Health centers in 2009 referred to national and private hospitals,

which had increased and were located inside the downtown area, while only 2 or 3 were located in the suburban areas. For this reason, urban growth tends to occur outside the downtown area which makes it farther from a health center and gives the OR of 7 which means the possibility of urban growth in an area more than 1 km away from a health center is 7 times greater than urban growth in an area 1 km or nearer to a health center.

Validation of the LR Model on Urban Growth in 3 Periods

All the LR models for the 3 periods on urban growth have to evaluate the performance of the simulation results using a contingency table and the ROC methods by comparing the probability map of each LR model (Figure 7) with the actual urban growth map (see also Figure 5).

For the LR model of the 2002-2015 period validation, the urban growth area between 2002 and 2015 that was allocated to the urbanization probability image was about 70.6 km². The model provided an overall accuracy of 91.32%. This high figure for overall accuracy comes from the successful fit of the non-urban area of about 95%, but the successful fit of the urban area is only 58% (Figure 8(a)). At the same time, the area under the ROC curve of the calibrated LR model of the 2002-2015 period was 0.903 (Table 9).

Likewise, for the validation of LR model of the 2002-2015 period, the urban growth area between 2002 and 2009 that was allocated to the urbanization probability image was about 32.18 km². The model gave an overall accuracy of 95.67%. This high figure for overall accuracy comes from the successful fit

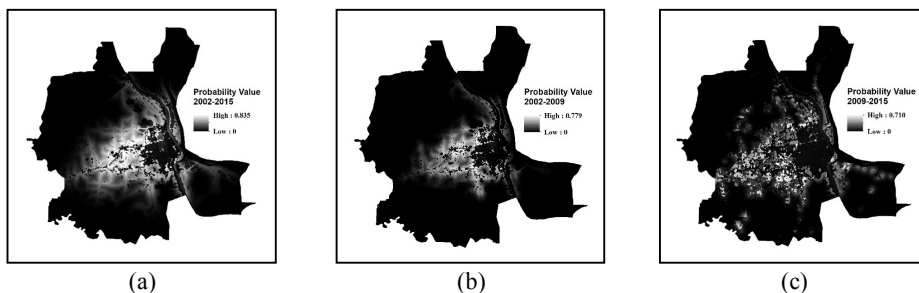


Figure 7. Urban growth probability map by LR model in each period: (a) 2002 to 2015, (b) 2002-2009, and (c) 2009-2015

of the non-urban area of about 98%, but the successful fit of the urban area is only 54% (Figure 8(b)). At the same time, the area under the ROC curve of the calibrated LR model of the 2002-2009 period was 0.948 (Table 9).

Similarly, for the validation of the LR model of 2002–2015 period, the urban growth area between 2009 and 2015 that was allocated to the urbanization probability image was about 38.41 km². The model gave an overall accuracy of 93.67%. This high figure for overall accuracy comes from the successful fit of the non-urban area of about 97%, but the successful fit of the urban area is only 44% (Figure 8(c)). Meanwhile, the ROC result reveals that the LR model of the 2009-2015 period provides a high ROC value of 0.901 (Table 9).

From the results, the overall accuracy of the LR model for the 3 periods is slightly different and ranges between 91.32% and 95.67%. The high figure for overall accuracy comes from the successful fit of the non-urban areas in the study area. However the accuracy of the predicted urban area is significantly different and ranges between 43.64% and 57.91% with the LR model of 2002-2015 providing the highest value. In addition, the ROC of the LR model for the 3 periods is slightly different and ranges between 0.901 and 0.948. The result showed that the LR model of the 3 periods is nearly perfect for the assignment of the probability of urban expansion, and it can be accepted for the urban growth prediction.

As was seen in the contingency table, the 2 types of errors are omission and commission. Herei, the omission error as an underprediction

(omission error) was allocated to the wrong location of the urban area in the model (i.e. 29.71, 14.83, and 21.65 km² for the LR model of the 3 periods), and these areas were replaced by overprediction (commission error). These errors indicate that the influence of neighboring areas of the existing urban areas have higher probability values than scattered urban areas. This means that most of the underprediction areas are isolated growth areas and in the linear branch, while overprediction areas are laid closely to the infill growth and expansion growth around the neighboring areas of the existing urban clusters. This phenomenon indicates that the LR model may not be appropriate for prediction of the outlying growth type, and it agrees with the result from previous work of Duwal (2013) who also found the weakness of the LR model in predicting the outlying urban growth.

Optimum LR Model and Future Urbanization Pattern

According to the spatial and statistical pattern validation for the LR model of the 3 periods, as reported earlier and presented in Figure 8 with the contingency table in Table 9, the LR model of the 2002-2015 long-term period was chosen as the optimum LR model for the Phnom Penh urban growth pattern in the 2030 prediction because it provides a high ROC value and high percentage of overall accuracy with the satisfactory result of fitting urban allocation. Moreover, it is a long-term period model, so the latest driving factors of 2015 could be updated according to city development, for example the existing urban cluster in 2015, and the result would be more

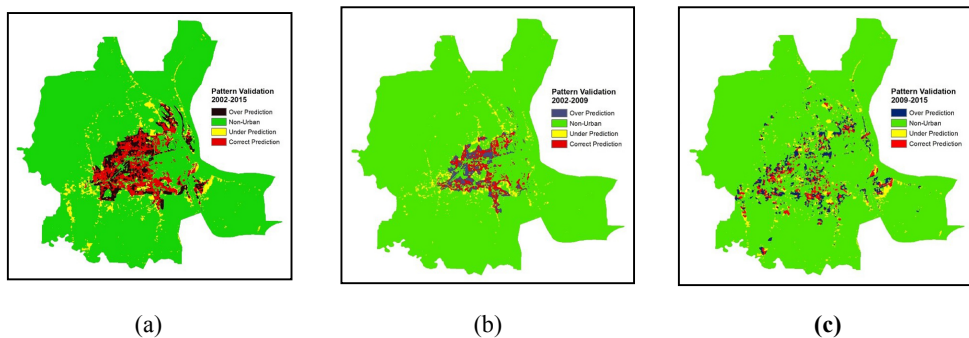


Figure 8. Validation pattern of LR model for period: (a) 2009-2015, (b) 2009-2015, and (c) 2009-2015

reliable. This finding also agreed with the previous work of Nduwayezu (2015), who also got a good calibration result from the LR model of a long-term period.

Urban Growth Pattern Prediction in 2030

Due to the limitation of the binary LR model to quantify the urban growth area in the future as mentioned above, the Markov chain model was here used to predict the urban area in 2030 based on the Markov transition probability matrix of the land cover change between 2002 and 2015 (Table 10) and the transition area matrix of the land cover in 2030 (Table 11). Theoretically, both the transition probability matrix and transition area matrix can be further used to predict the urban growth pattern in 2030 using the CA-Markov model.

As a result from Table 11, the predicted urban area in 2030 was 185.83 km². This means that the increase in urban area in 2030 from the existing urban area of 2015 is 77.1 km². This area was used for allocation on the future urbanization probability map (Figure 9) which was derived from an optimal equation of the LR model of the 2002-2015 period with updated driving factors in 2015. Finally, the predicted urban growth pattern of Phnom Penh in 2030 based on the LR model is produced and displayed in Figure 10.

It illustrates that the pattern of urban growth occurred in the north, south and east directions around the existing urban clusters in 2015, and some are expected to occur along the major roads and ring roads. Furthermore, most of the miscellaneous land that comprises

Table 9. Validation of LR model of 3 periods: 2000-2015, 2002-2009, and 2009-2015

Unit in km ² Model	Reality 2002-2015			Reality 2002-2009			Reality 2009-2015		
	Urban area	Non-urban area	Total	Urban area	Non-urban area	Total	Urban area	Non-urban area	Total
Urban area	40.88	29.72	70.60	17.35	14.83	32.18	16.76	21.66	38.43
Non-urban area	29.71	584.01	613.72	14.83	637.31	652.14	21.65	624.24	645.89
Total	70.59	613.73	684.32	32.18	652.14	684.32	38.41	645.91	684.32
Overall accuracy (%)			91.32			95.67			93.67
Non-urban area accuracy (%)			95.16			97.73			96.65
Urban area accuracy (%)			57.91			53.92			43.64
ROC			0.903			0.948			0.901

Table 10. Markov transition probability matrix of land cover change between 2002 and 2015

Land cover type in 2002	Land cover type in 2015				
	U	A	V	W	M
Urban and built-up areas (U)	1	0	0	0	0
Arable land (A)	0.1391	0.6538	0.1551	0.0151	0.0369
Vegetation (V)	0.1499	0.2726	0.4894	0.0548	0.0332
Water bodies (W)	0.0512	0.1721	0.2332	0.4802	0.0632
Miscellaneous land (M)	0.2416	0.3126	0.3208	0.0904	0.0346

Table 11. Transition area matrix by Markov chain model for land cover in 2030

Land cover type in 2015	Land cover type in 2030 (in km ²)				
	U	A	V	W	M
Urban and built-up areas (U)	108.72	0.00	0.00	0.00	0.00
Arable land (A)	38.25	179.74	42.65	4.15	10.14
Vegetation (V)	27.88	320.71	91.04	10.20	6.18
Water bodies (W)	4.49	15.07	20.43	42.06	5.54
Miscellaneous land (M)	6.49	8.40	8.62	2.43	0.93
Total	185.83	523.92	162.75	58.84	22.80

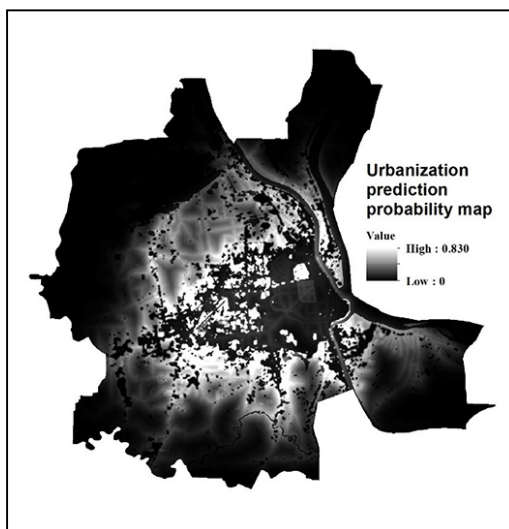


Figure 9. Future urbanization probability map of Phnom Penh

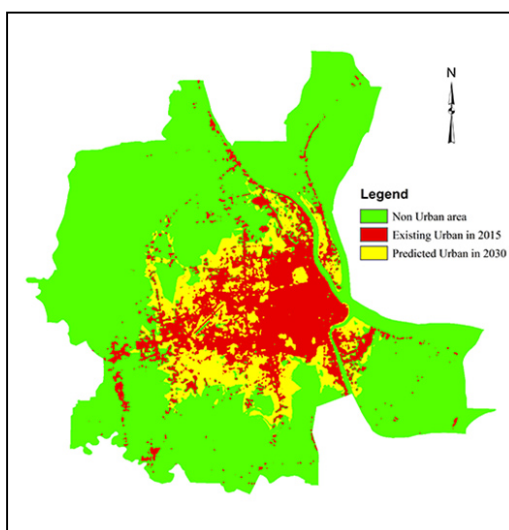


Figure 10. Prediction of urban growth pattern of Phnom Penh in 2030

construction sites for the development of satellite cities in 2015 is also expected to have had the construction sites finished and will become urban and built-up areas by 2030.

Conclusions

The results of the land cover classification from remotely sensed data in 2002, 2009, and 2015 using the maximum likelihood classifier

showed that arable land for rice production is the predominant land cover type in Phnom Penh in those 3 years. Since the Cambodian economy relies on the agricultural sector, this result can be used as evidence for proof in terms of the spatial pattern from the past upto recent years. Meanwhile, arable land and vegetation were the key land cover classes for urbanization in the 3 periods, and the urban and built-up areas continuously increased from 38.13 km² in 2002 to 70.31 km² in 2009 and to

108.73 km² in 2015; this caused the decline of arable land, vegetation, and water bodies. However, the change for miscellaneous land that was bare land for construction and development sites was shown as fluctuating. This demonstrates the fast development of urbanization that reflects the increase of the construction sector, the growth of the economy and industrialization, and the demand for housing and satellite cities for people living in Phnom Penh. In addition, the pattern of urban growth demonstrates that urban areas expanded only toward the southern, northern and western sides of Phnom Penh during 2002-2009 with all types of growth, while the eastern part of Phnom Penh did not show much of a growth pattern due to that area being mostly wetland. However, the expansion of urban growth during 2009-2015 occurred in all directions, and the types of urban growth were changed to be expansion growth, clustered branch, and isolated growth, whereas infill growth did not occur during 2009-2015.

The analysis of the driving factors on urban growth using the LR models of the 2002-2015, 2002-2009, and 2009-2015 periods found that the relative influences of the driving factors varied according to the urban and built-up areas' pattern and time. In this study, there were 2 common driving factors which were the distance to the existing urban clusters and the distance to industrial areas that had negative influences on urban growth in all 3 periods from observation of the top 5 driving factors. The distance to the airport and roads are the dominant influences on urban growth in the long-term period (2002-2015) as well as in short-term period (2002-2009). However, the influence of the driving factors changed to distance to the CBD and miscellaneous land during the 2009-2015 short-term period.

The final urban growth pattern from the LR model of the 3 periods showed a good result for overall accuracy with about 91%, 96%, and 94% for the successful fit for urban allocation with values of about 58%, 54%, and 44%, and the ROC of 0.90, 0.95, and 0.90, respectively. These results show that LR model of the 2002-2015 period provided the best LR calibration equation and it was chosen as an optimum LR model to predict the urban growth pattern of Phnom Penh in 2030. As a result, it revealed that Phnom Penh tends to

expand to the north, south, and east. This expansion can be influenced by the nearest existing urban clusters in 2015, and the influence from roads.

In conclusion, the results from the findings showed that the LR model incorporated with satellite imageries' data and the GIS technique is an efficient approach for application to understand the influence of the driving factors for the urban growth of Phnom Penh. It contributes significant information as a guideline for planners and decision makers.

References

- Achmad, A., Hasyim, S., Dahlan, B., and Aulia, D. (2015). Modeling of urban growth in tsunami-prone city using logistic regression: analysis of Banda Aceh, Indonesia. *Appl. Geogr.*, 62:237-246.
- Aguayo, M.I., Wiegand, T., Azocar, G.D., Wiegand, K., and Vega, C.E. (2007). Revealing the driving forces of mid-cities urban growth patterns using spatial modeling: a case study of Los Angeles, Chile. *Ecol. Soc.*, 12(1):13.
- Allen, J. and Lu, K. (2003). Modeling and prediction of future urban growth in the Charleston region of South Carolina: a GIS-based integrated approach. *Conserv. Ecol.*, 8(2):2.
- Alsharif, A.A.A. and Pradhan, B. (2014). Urban sprawl analysis of Tripoli Metropolitan City (Libya) using remote sensing data and multivariate logistic regression model. *J. Indian Soc. Remote Sens.*, 42(1):149-163.
- Arsanjani, J.J., Helbich, M., Kainz, W., and Boloorani, A.D. (2013). Integration of logistic regression, Markov chain and cellular automata. *Int. Appl. Earth Obs.* 21:265-275.
- Bureau des Affaires Urbaines. (2007). *Livre Blanc du Développement et de l'Aménagement de Phnom Penh* (White Book of Development and Management of Phnom Penh). Municipalité de Phnom Penh, Bureau des Affaires Urbaines, Phnom Penh, Cambodia, 330p.
- Bhatta, B. (2010). *Analysis of Urban Growth and Sprawl from Remote Sensing Data*. Springer-Verlag, Berlin, 172p.
- Braimoh, K.A. and Onishi, T. (2007). Spatial determinants of urban land use change in Lagos, Nigeria. *Land Use Policy*, 24:502-515.
- Cheng, J. and Masser, I. (2003). Urban growth pattern modeling: a case study of Wuhan city, PR China. *Landscape Urban Plan.*, 62:199-217.
- Clark, W.A.V. and Hosking, P.L. (1986). *Statistical Methods for Geographers*. Wiley, New York, NY, USA, 528p.
- Dewan, A.M. and Yamaguchi, Y. (2009). Land use and land cover change in Greater Dhaka, Bangladesh: using remote sensing to promote sustainable urbanization. *Appl. Geogr.*, 29:390-401.

- Doyle, S.E. (2012). City of water: architecture, urbanism and the floods of Phnom Penh. *Nakhara: Journal of Environmental Design and Planning*, 8:135-154.
- Duwal, S. (2013). Modeling urban growth in Kathmandu Valley, [MSc. thesis]. Faculty of Geo-Information Science and Earth Observation, University of Twente, Enschede, Netherlands, 80p.
- Eyoh, A., Olayinka, D.N., Okwuashi, P.N.O., Isong, M., and Udoudo, D. (2012). Modelling and predicting future urban expansion of Lagos, Nigeria from remote sensing data using logistic regression and GIS. *International Journal of Applied Science and Technology*, 2(5):116-124.
- Fitzpatrick-Lins, K. (1981). Comparison of sampling procedures and data analysis for a land-use and land cover map. *Photogramm. Eng. Rem. S.*, 47(3):343-351.
- Hu, Z. and Lo, C.P. (2007). Modeling urban growth in Atlanta using logistic regression. *Comput. Environ. Urban*, 31:667-688.
- Jensen, J.R. (2005). *Introductory Digital Image Processing*, 3rd ed. Prentice Hall, Upper Saddle River, NJ, USA, 544p.
- Landis, J. and Koch, G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 33:159-174.
- Manu, A., Twumasi, Y.A., Lu, K.S., and Coleman, T.L. (2015). Predicting urban growth of a developing country city using a statistical modeling approach. *Int. J. Geoma. Geosci.*, 5(4):603-613.
- Menard, S.W. (2002). *Applied Logistic Regression Analysis*, 2nd ed. SAGE Publications Inc., Thousand Oaks, CA, USA, 128p.
- National Geographic Department. (2002). *Cambodia National Topographic Map*. National Geographic Department, Ministry of Land Management, Urban Planning and Construction, Phnom Penh, Cambodia.
- National Institute of Statistics. (2013). *Cambodia Inter-Censal Population Survey 2013 Final Report*. National Institute of Statistics, Ministry of Planning, Phnom Penh, Cambodia, 142p.
- National Institute of Statistics. (2009). *General Population Census of Cambodia 2008 National Report on Final Census Results*. National Institute of Statistics, Ministry of Planning, Phnom Penh, Cambodia.
- National Institute of Statistics. (2002). *General Population Census of Cambodia 1998 Final Census Result*, (2nd ed.). National Institute of Statistics, Ministry of Planning Phnom Penh, Cambodia.
- Nduwayezu, G. (2015). Modeling urban growth in Kigali City Rwanda. [MSc. thesis]. Faculty of Geo-Information Science and Earth Observation, University of Twente, Enschede, Netherlands, 77p.
- Odzemir, A. (2011). Using a binary logistic regression method and GIS for evaluating and mapping the groundwater spring potential in the Sultan Mountains (Aksehir, Turkey). *J. Hydrol.*, 405:123-136.
- Pontius, R.G. Jr. and Schneider, L.C. (2001). Land-cover change model validation by an ROC method for the Ipswich watershed, Massachusetts, USA. *Agr. Ecosyst. Environ.*, 85:239-248.
- Rogerson, P.A. (2010). *Statistical Methods for Geography*, 3rd ed. SAGE Publications, London, UK, 368p.
- Shalaby, A. and Tateishi, R. (2007). Remote sensing and GIS for mapping and monitoring land cover and land-use changes in the Northwestern coastal zone of Egypt. *Appl. Geogr.*, 27:28-41.
- Shu, B., Zhang, H., Li, Y., Qu, Y., and Chen, L. (2014). Spatiotemporal variation analysis of driving forces of urban land spatial expansion using logistic regression: a case study of port towns in Taicang City, China. *Habitat Int.*, 43:181-190.
- The World Bank. (2015). *East Asia's Changing Urban Landscape: Measuring a Decade of Spatial Growth*. The World Bank, Washington, DC, USA, 184p.
- United Nations Population Fund. (2014). *Urbanization and its linkage to socio-economic and environmental issues*. United Nations Population Fund (UNFPA), Phnom Penh, Cambodia, 38p.
- Wilson, E.H., Hurd, J.D., Civco, D.L., Prisloe, M.P., and Arnold, C. (2003). Development of a geospatial model to quantify, describe and map urban growth. *Remote Sens. Environ.*, 86:275-285.