URBAN GROWTH PREDICTION USING LOGISTIC REGRESSION MODEL : Case Study in Bogor, West Java Province, Indonesia

(Prediksi Pertumbuhan Urbanisasi Menggunakan Model Regresi Logistik: Studi Kasus di Bogor, Provinsi Jawa Barat, Indonesia)

by/oleh:
Asep Karsidi[1] and **Antonius B. Wijanarto**[2]
[1,2] National Coordinating Agency for Surveys and Mapping
Jl Raya Jakarta Bogor KM 46 Cibinong
E-mail : akarsidi@gmail.com, anton@bakosurtanal.go.id

Diterima (received): 11 September 2011; Disetujui untuk dipublikasikan (accepted): 21 November 2011

ABSTRACT

Urbanization has been one of major issues in Indonesia, especially in big cities like Jakarta. Bogor as the supporting area of Jakarta has been affected by this urbanization. It did not only impact in increasing population, but it did also affect the carrying capacity of the environment. This research studied urban growth of Bogor, using logistic regression model. The model involve several driving factors that can be used to predict future impact of urbanization. The results indicate that the logistic regression model for urban growth in Bogor shows that R^2 = 0.18, which means that the model is relatively good. In this research ROC has been produced is 0.8429, which means the logistic regression model of urban growth of Bogor can be estimated well.

***Keywords**: Urban Growth, Logistic Model, Spatial Analysis*

ABSTRAK

Urbanisasi merupakan salah satu isu besar di Indonesia, khususnya di kota-kota besar seperti Jakarta. Bogor sebagai daerah pendukung Jakarta juga menjadi daerah yang mendapat dampak dari urbanisasi tersebut. Bukan hanya jumlah populasi yang makin bertambah, tetapi juga daya dukung lingkungan yang juga dipengaruhi. Penelitian ini mengkaji pertumbuhan urbanisasi di Bogor, dengan menggunakan model logistik. Model ini mengidentifikasi beberapa faktor pendorong yang digunakan untuk memprediksi pertumbuhan urbanisasi ke depan. Hasil dari model regresi logistik menunjukkan bahwa pertumbuhan urbanisasi di Bogor dapat diprediksi dengan R^2 = 0,18, yang dapat dikategorikan agak bagus. Dalam penelitian ini ROC = 0,8429, yang berarti urbanisasi di Bogor dapat diprediksi dengan baik menggunakan model regresi logistik.

***Kata Kunci:** Pertumbuhan Urbanisasi, Model Logistik, Analisis Spasial*

INTRODUCTION

Bogor Municipality and Regency are located in southern Jakarta, the Capital City of Indonesia. As a buffer zone, Bogor provides many services for Jakarta such as labors, fresh water from Ciliwung and Cisadane River streams, food from agriculture and farm, settlement and recreational areas.

Population of Bogor Regency increased from 3,004,444 in 1999 to 3,170,400 in 2001, and population of Bogor Municipality increased from 697,496 to 760,329 during 1999 – 2001. In providing services for Jakarta and to support people activities that live in Bogor itself, Bogor has been growing rapidly especially in providing agriculture and settlement area. Urban growth in any places occurs and is strongly related by increasing population. This situation causes many land clearing to fulfill human need of land. Disturbance in human population can cause problem like pollution and traffic jam, although can be overcome by technology and economy.

Urban growth has become global issue in capital cities. Consequently, land conversion will be occur rapidly, disturbing other existence of ecosystem like forest and leading to unstable environment. This problem has been occuring in Bogor as a buffer city from Jakarta city. This research was conducted to observe the trend of urban growth in Bogor from 2002 – 2005, which will be used as basis data to predict urban growth that might be happened in year 2008.

The objectives of this research are:

1. to identify urban area of Bogor in 2002, 2005 and 2008;
2. to analyze urban growth of Bogor in 2002 – 2005;
3. to produce the logistic regression model of urban growth based on the suggested driving factors during 2002 – 2005; and
4. to predict the urban growth in Bogor in 2008.

METHOD

Time and Research Location

The site is located in 6°18'10.70"-6°47'6.98" South and 106°41'59.98" - 106°56'1.32" East which neighbor with Depok Regency in northern, some districts of Bogor Regency in eastern and western part, and Sukabumi Regency in southern. Based on the Statistic Indonesia year 2001, the research site consist of 8 sub districts in Bogor Regency and 6 sub districts in Bogor Municipality which total area is 52,045 hectares. **Figure 1** below are the detail location of research site.

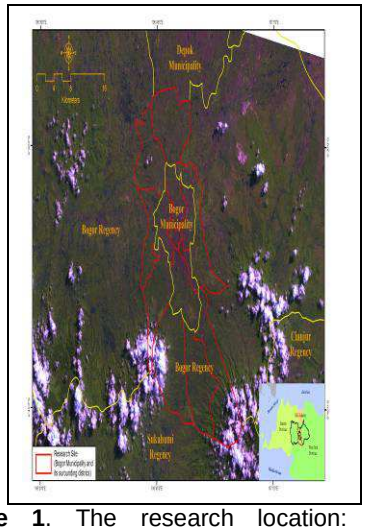


Figure 1. The research location: Bogor Municipality and Regency

Data Source

Spatial data will be used in this research are divided into two categories: raster and vector data. Raster data is used in this research is Landsat ETM+. Time series data of Landsat imagery will be processed to produce urban area map which will be analyzed the changes year by year. Vector data will be used in this research is road and river maps of Bogor which is produced by Ministry of Forestry. **Table 1** below is the detail list of spatial data will be used in this research.

Table 1. List of Spatial Data for the Research

No.	Category	Type of Data	Description	Source
1.	Raster	Landsat Imagery	- Path/ Row: 122/065 - Year: 2002, 2005, 2008 - Spatial resolution: 30x30 m	USGS, landsat.org, landcover.org.
2.	Vector	River map	- River map of Bogor	Ministry of Forestry
		Road map	- River map of Bogor	Ministry of Forestry

Land Cover Classification

Time-series data of Landsat images are used in this research to generate the information about time-series data of land cover types which exist in Bogor. The land cover types are generated by image classification process is done to each Landsat image. In order to have a good classification result, there are some processes should be done which can be divided into two categories: image pre-processing and image processing.

Image processing will be done in this activity is Landsat image classification, which will be conducted by Supervised Classification Technique using Maximum Likelihood method that use the reference data of land cover type year 2002 of the existing land cover as guidance for training sample. At the end of this activity, accuracy assessment will be done in order to determine that the land cover map is produced accurately.

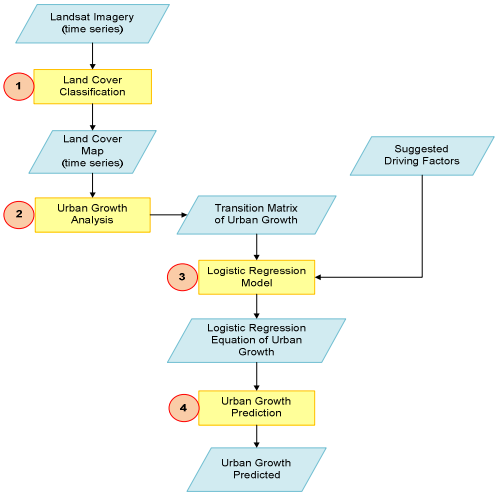


Figure 2. Four main activities in this research

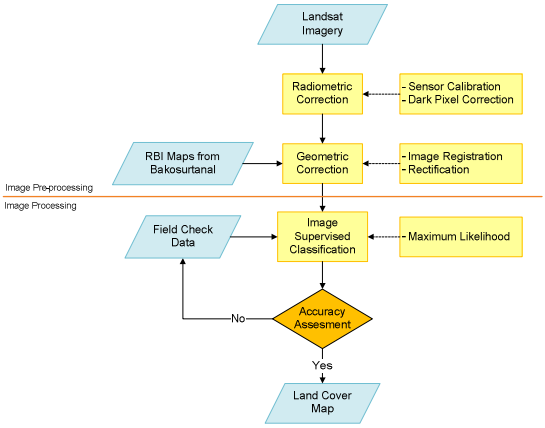


Figure 3. Flow diagram of land cover classification process

Urban Growth Analysis

Urban growth can be defined as the rate of growth of an urban population. It refers to the increasing or decreasing of urban area. Urban growth analysis processes two classified images from different time which the changes want to be analyzed. Time-series data of land cover maps that are produced from previous process should be reclassified into two main types: urban area and non-urban area. The reclassification process itself should exclude the land cover of water body, cloud, and shadow into their own type, because these land covers cannot be classified, either as urban area or non-urban area. Cloud and shadow should also be equal in area and distribution in each image will be analyzed in order to make the comparison of two images is equal.

The result of the urban growth analysis is urban growth image which happened in during 2002-2005 and its transition matrix and probability of change. Since the

research focusing on urban growth, only the transition matrix and change probability of urban growth will be used as dependent variable to build the logistic regression of urban growth that will be processed in the next activities.

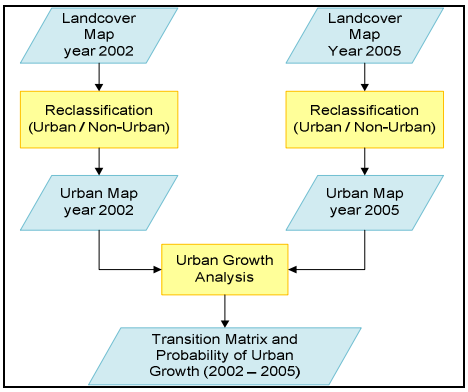


Figure 4. Flow diagram of urban growth analysis

Logistic Regression Model

Prediction of urban growth may be useful for government to plan the action should be done if the prediction become real. The prediction of urban growth in Bogor will help the government to set up the policy which can be managed the urban growth will be happened, so its growth will not exceed the environmental carrying capacity.

The prediction of urban growth in this research will be done using the statistical logistic regression method which uses the urban growth year 2002-2005 as dependent variable and the driving factors as independent variables. Independent variables are used in this research are distance from road, distance from river, distance from agriculture area, and distance from the existing urban area. The logistic regression model will be done by using IDRISI 15.0 which will result the logistic regression equation that will be used in predicting urban growth in 2008. **Figure 5** shows the flow diagram of logistic regression model process.

Urban Growth Prediction

In this research, urban growth prediction in Bogor is deal with statistical method (logistic regression). The flow diagram of urban growth prediction is shown in **Figure 6**.

The logistic regression equation which is resulted from logistic regression model analysis is used to predict the urban growth year 2008 by inputting the existing condition of independent variables year 2005 into logistic regression equation. The result of urban prediction is predicted urban area year 2008 which will be validated by the classified image of urban area year 2008.

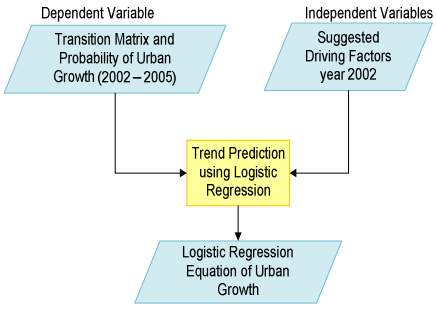


Figure 5. Flow diagram of logistic regression model process

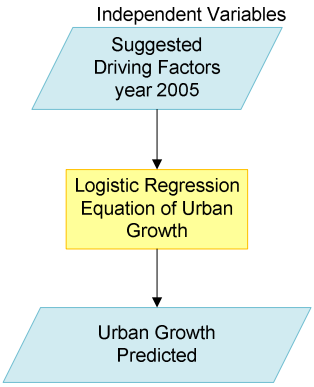


Figure 6. Flow diagram of urban growth prediction

RESULT AND DISCUSSION

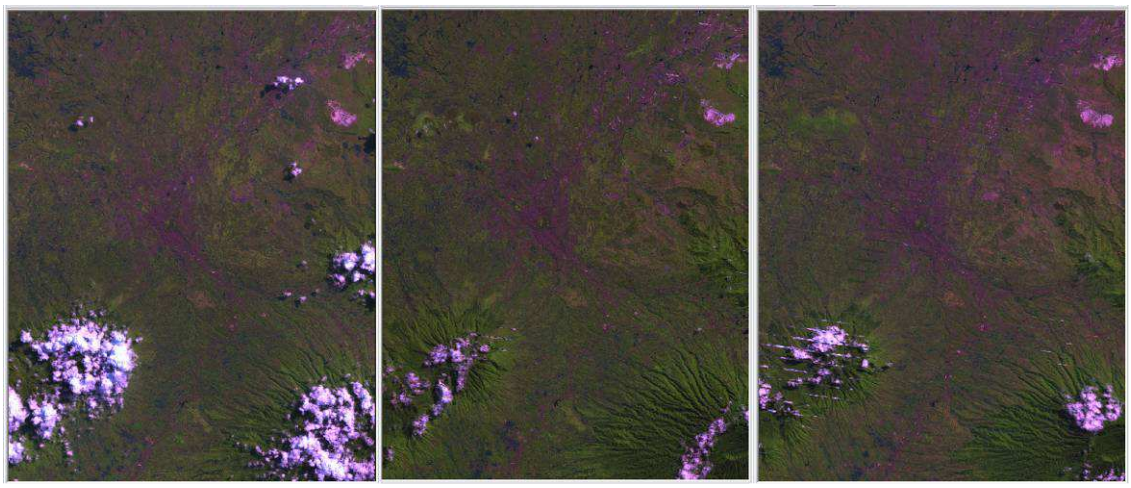
Land Cover Classification

Land cover classification is the first step in predicting urban growth in Bogor. The classification is done to three Landsat ETM+ imageries year 2002, 2005, and 2005 (**Figure 7**). ERDAS Imagine 9.1 is used to do the supervised classification using maximum likelihood method. Landsat ETM+ is used is classified into six land cover types: (1) Forest, (2) Agriculture, (3) Bareland, (4) Urban Area, (5) Water body, and No Data. Land cover map year 2002 which produced by Ministry of Forestry is used as reference data when creating sample of land cover types will be classified. The number of samples are taken is at least 30 samples each land cover.

Land cover classification has been done in three time series resulting overall accuracy 80%, 78% and 76% in year 2002, 2005, and 2008, respectively. The land cover classification shows that agriculture area in three time series is the largest area than the other land cover types even the area was decreased time by time. In 2002, forest cover is the second largest of land cover types which the area for about 11,750 Ha, but in 2005 and 2008 the forest cover was decreased. In that time, urban area has taken the second places of the largest land cover types with area up to 14,424 ha and 17,215 ha in 2005 and 2008 respectively. Other land cover also increasing is bare land which was increased from 1,104 ha in 2002 to 1,867 ha in 2005 and 2,615 ha in 2008. The detail result of land cover classification is shown in **Table 3**, and the land cover maps of Bogor year 2002, 2005 and 2008 is shown in **Figure 8**.

Urban Growth Analysis

Urban growth analysis is conducted by developing comparison matrix of land cover year 2002 and 2005. The matrix of land cover change shows that urban area has been shifted other land covers such as forest, agriculture, bare land and water body. Forest cover has been changed to urban area is only 6.39 ha and bare land is 175.77 ha. The largest contributor for urban growth is agriculture for about 4,900.41 ha or 34% of total urban area in 2005 (see **Table 4**).

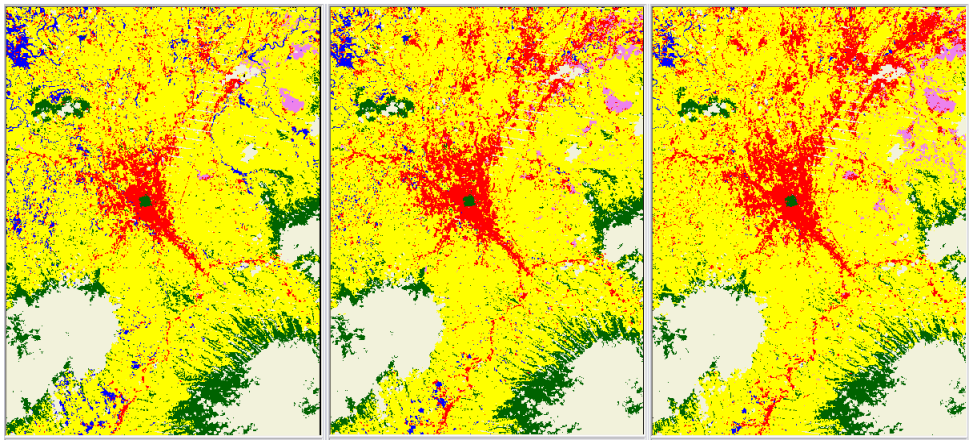


(a) (b) (c)

Figure 7. Landsat ETM+: (a) 2002, (b) 2005, and (c) 2008

Table 3. Land cover of Bogor year 2002, 2005 and 2008

ID	Land Cover	Color	Area (ha)		
			2002	2005	2008
1	Forest		11,750.04	10,494.45	8,319.96
2	Agriculture		83,497.32	81,232.38	81,071.19
3	Bareland		1,104.12	1,867.86	2,615.31
4	Urban Area		9,163.44	14,424.57	17,215.11
5	Water		4,788.54	2,303.10	1,099.35
6	No Data		21,922.92	21,922.92	21,922.92
Overall Accuracy: 2002 = 80.00%, 2005 = 78%, 2008 = 76%					



(a) (b) (c)

Figure 8. Land cover: (a) 2002, (b) 2005, and (c) 2008

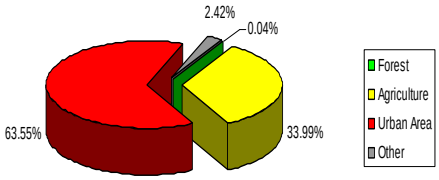


Figure 9. Land cover contributors for urban growth in Bogor during 2002 –2005.

Urban area has been shifted other land covers dramatically during 2002 – 2005, and it is happened mostly in north side of Bogor. This situation might be driven by the need of people in Jakarta which need settlement area, trade center and other buildings to support their activities. The location of urban area in Bogor during 2002 – 2005 can be seen in **Figure 10**.

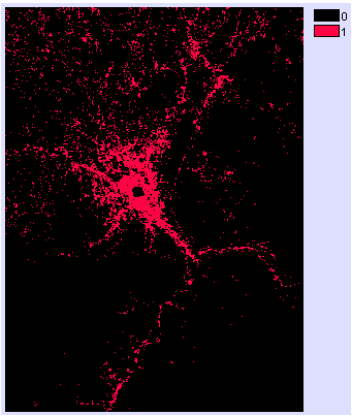
Logistic Regression Model

Urban growth prediction in this research is done using logistic regression model which calculates dependent and independent variables of urban growth. Urban growth data during 2002–2005 acts as dependent variable of logistic regression model, and suggested driving factors which drive urban growth as independent variables. The suggested driving factors of urban growth are distance from agriculture area, distance from existing urban area, distance from roads, and distance from rivers.

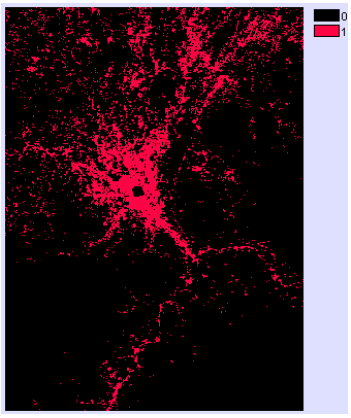
Urban growth map during 2002 – 2005 is resulted using overlay operation to urban area map year 2002 and 2005 to find the difference between urban areas in both time series. The map of urban growth during 2002–2005 can be seen in **Fig. 11**.

The preparation of independent variables are used to develop logistic regression model are made using buffer operation which calculates the distance from its features of independent variables. The independent variables of urban growth is taken from year 2002 which is used in develop logistic regression model is in Bogor are shown in **Figure 12, Figure 13, Figure 14** and **Figure 15.**

The logistic regression equation is computed from calculation of dependent and independent variables of urban growth. The result of calculation of logistic regression model is shown in **Table 5** and **Table 6**.



(a) year 2002



(b) year 2005

Figure 10. Urban area of Bogor: (a) year 2002 and (b) year 2005

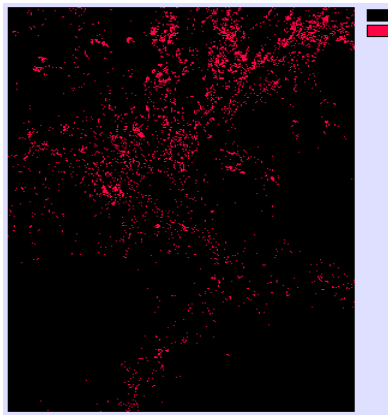
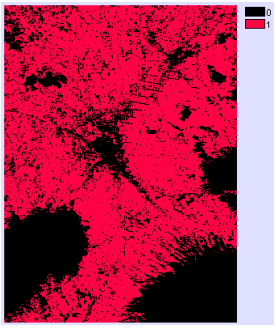


Figure 11. Urban growth of Bogor during 2002 – 2005



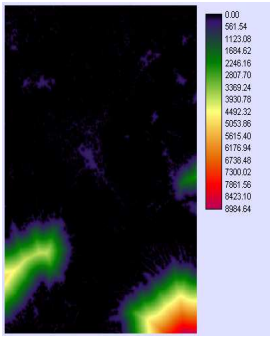
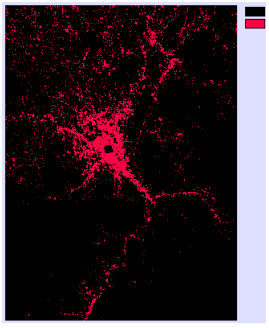


Figure 12. Distance (buffer) of agriculture area year 2002



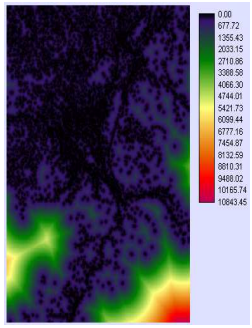
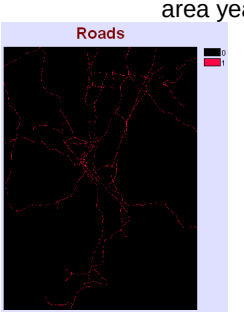


Figure 13. Distance (buffer) of urban aArea year 2002



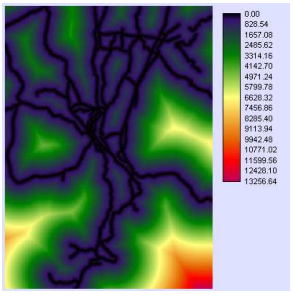
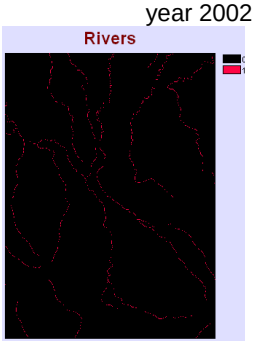


Figure 14. Distance (buffer) of roads year 2002



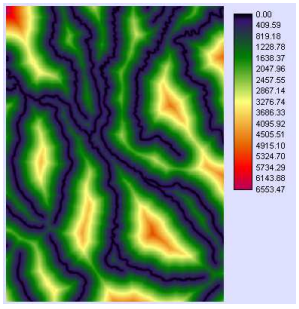


Figure 15. Distance (buffer) of rivers year 2002

Table 4. Land cover change matrix in Bogor during 2002-2005

Land Cover		2005						Total (2002)
		Forest	Agriculture	Bareland	Urban Area	Water	No Data	
2002	Forest	10,494.45	1,244.88	2.61	6.39	1.62	0.00	11,749.95
	Agriculture	0.00	76,494.33	1,110.06	4,900.41	985.32	0.00	83,490.12
	Bareland	0.00	232.11	690.75	175.77	4.95	0.00	1,103.58
	Urban Area	0.00	0.00	0.00	9,163.44	0.00	0.00	9,163.44
	Water	0.00	3,242.34	64.17	172.71	1,308.51	0.00	4,787.73
	No Data	0.00	0.00	0.00	0.00	0.00	21,922.92	21,922.92
	Total (2005)	10,494.45	81,213.66	1,867.59	14,418.72	2,300.40	21,922.92	132,217.74

Table 5. Individual regression coefficient

Factors	Variables	Coeficient
Intercept	Intercept (β_0)	-1.17298759
Distance from urban area	Urbandist (β_1)	-0.00545200
Distance from agriculture	Agricdist (β_2)	-0.03534407
Distance from roads	Roaddist (β_3)	-0.00039130
Distance from rivers	Riverdist (β_4)	-0.00014849

Table 6. Logistic regression equation

The logistic function: $f(z) = \frac{1}{1+e^{-z}}$
which variable z is usually defined as: z = -1.1730 - 0.005452*urbandist - 0.035344*agricdist - 0.000391*roaddist - 0.000148*riverdist

The calculation of logistic regression model produces Pseudo R^2 which indicates the fitness of relationship of the model. Thus, pseudo R_square = 1 indicates a perfect fit, where as pseudo R_square = 0 indicates no relationship. Pseudo R^2 greater than 0.2 is considered a relatively a good fit (Clark and Hosking 1986 *in* IDRISI Andes Help Contents version 1990 - 2000). The statistics of logistic regression model for urban growth in Bogor shows that Pseudo R^2 = 0.1832, which means the model is slightly a relatively a good fit..

Beside the Pseudo R^2, IDRISI Andes also calculated Relative Operation Characteristic (ROC) which is an excellent method to compare a Boolean map of "reality" versus a suitability map. Thus, ROC is included here as an excellent statistic for measuring the goodness of fit of logistic regression. The ROC value ranges from 0 to 1, where 1 indicates a perfect fit and 0.5 indicates a random fit. A ROC value between 0.5 and 1 indicates some association between the X variables and Y. The larger the ROC is better the fit. In this research ROC has been produced is 0.8429, which means the logistic regression model of urban growth of Bogor is a good fit.

Urban Growth Prediction

The prediction of urban growth for next 3 years (in 2008) is resulted by inputting new values of independent variables into logistic regression equation has been produced from previous process. There are four independent variables are involved in the model: distance from agriculture area, distance from existing urban area, distance from roads, and distance from rivers. In this research, distance from agriculture area and distance from existing urban area are generated from buffer operation of its features are taken from land cover year 2005. Meanwhile, distance from roads and rivers variables is assumed still in the same condition which means still using the previous data year 2002.

The calculation of urban growth prediction is conducted using IDRISI

Andes by inputting the suggested independent variables of urban growth. Urban growth prediction of Bogor year 2008 using logistic regression model has been resulted in previous process produces new prediction image of urban growth year 2008. The images shows that the urban growth seems will be happened when the area near the roads and existing urban area which will shift the agriculture area. Although, based on the previous discussion the logistic regression model has a good fit, the range of probability of urban growth is happened in 2008 only 0 – 0.21. It means only small probability it will happen. This situation may be caused by two independent variables (distance from roads and rivers) are used in prediction is the same data when developing logistic regression model. It also means both variables (distance from roads and rivers) have strong relationship in driving the urban growth. Whereas, when there are no changes in both variables, the probability of urban growth is also small

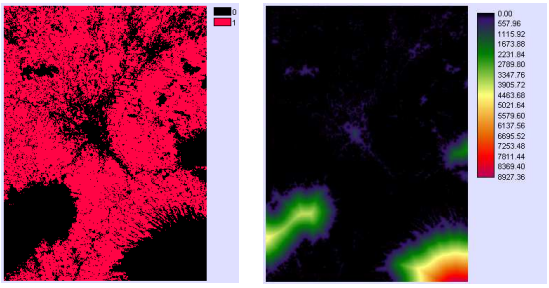


Figure 16. Distance (buffer) of agriculture area year 2005

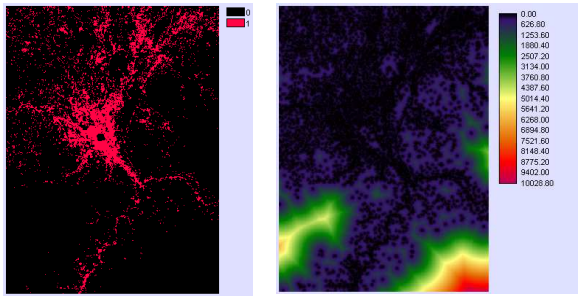


Figure 17. Distance (buffer) of urban area year 2005

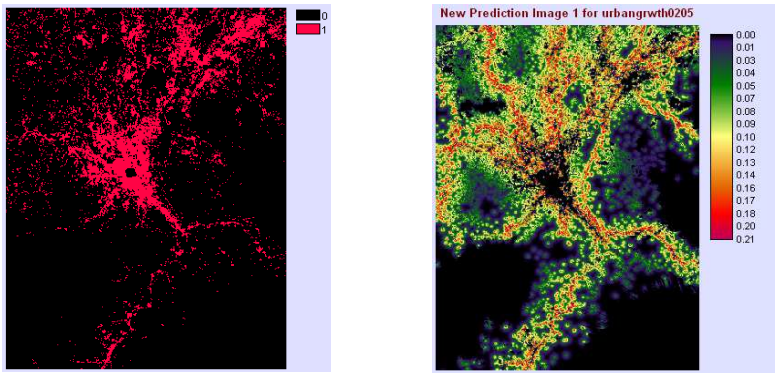


Figure 18. Urban area generated from land cover classification (left) and predicted urban area (right) year 2008

CONCLUSION

Urban growth has become global issue in capital cities, that supported by urbanization. In consequence land conversion will be happen rapidly, that disturb to the other existence of ecosystem like forest and other conversion which head for unstable of environment.

Urban growth prediction in this research is done using logistic regression model which calculates dependent and independent variables of urban growth. Urban growth data during 2002 – 2005 acts as dependent variable of logistic regression model, and suggested driving factors which drive urban growth as independent variables. The suggested driving factors of urban growth are distance from agriculture area, distance from existing urban area, distance from roads, and distance from rivers.

The statistics of logistic regression model for urban growth in Bogor shows that Pseudo R^2 = 0.1832, which means the model is slightly a relatively a good fit. In this research ROC has been produced is 0.8429, which means the logistic regression model of urban growth of Bogor is a good fit.

Although, the logistic regression model has a good fit, the range of probability of urban growth was only 0 – 0.21. This situation may be caused by two independent variables (distance from roads and rivers) are used in prediction is the same data when developing logistic regression model. It means both variables (distance from roads and rivers) have strong relationship in driving the urban growth. Whereas, when there are no changes in both variables, the probability of urban growth is also small.

REFERENCES

Bernhardsen, T. 1992. *Geographic Information Systems*. Norwegian: VIAK IT and Norwegian Mapping Authority. Norwegian.

Hosmer, D.W. and S. Lemeshow. 2000. *Applied Logistic Regression*. Second Edition. New York: John Wiley, 375 p.

Karsidi,.A. 2004. *Spatial Analysis of Land Use/Land Cover Change Dynamics Using Remote Sensing and GIS: A Case Study in the Down Stream and Surrounding of the Ci Tarum Watershed*. PhD Thesis, Department of Geographical and Environmental Studies, The University of Adelaide, South Australia.

Lillesand, T. M. and R. W. Kiefer. 1987. *Remote Sensing and Image Interpretation*. Second Edition. Canada: John Wiley & Sons.

_____. 1990. *Penginderaan Jauh dan Interpretasi Citra*. Penerjemah: Dulbahri, Prapto Suharsono, Hartono and Suharyadi. Yogyakarta: Gadjah Mada University Press.

McCullagh, P. and J.A. Nelder. 1989. *Generalized Linear Models*. London: Chapman and Hall.

Mather, Paul M. 2004. *Computer Processing of Remotely-Sensed Images: An Introduction*. Third Edition. Nottingham. John Wiley.

Montgomery, M. R., *et al*. Panel on Urban Dynamics, National Research Council (eds.). 2003. *Cities Transformed: Demographic Change and Its Implications in the Developing World*, pp. 128-153.

NASA. 2002. *Urban Growth in the Baltimore-Washington Corridor.* Earth Science Enterprise. Washington.

Truett, J., J. Cornfield, and W. Kannel. 1967. A multivariate analysis of the risk of coronary heart disease in Framingham. *Journal of Chronic Diseases*, 20, 511-524.

UNFPA. 1996. *The State of World Population 1996: Changing Places: Population, Development and the Urban Future*, p. 1. New York: UNFPA.

UNFPA 2007. *State of world population 2007: Unleashing the Potential of Urban Growth.* New York: UNFPA.