Modelling and Predicting Future Urban Expansion of Lagos, Nigeria from Remote Sensing Data Using Logistic Regression and GIS

Aniekan Eyoh[a], Dupe Nihinlola Olayinka[b]
Peter Nwilo[c] Onuwa Okwuashi[d], Mfon Isong[e], and Daniel Udoudo[f]

Department of Geoinformatics & Surveying[a, d, & e]
Faculty of Environmental Studies
University of Uyo
Uyo, Nigeria

Department of Surveying & Geoinformatics[b, c, & f]
Faculty of Engineering
University of Lagos
Lagos, Nigeria

Abstract

This research explores the implementation of a loosely coupled logistic regression model and geographic information systems in modelling and predicting future urban expansion of Lagos from historical remote sensing data (Landsat TM images of Lagos acquired on 1984, 2000 and 2005). ArcGIS and MATLAB software are used for the modelling. The three Landsat images are classified using the k-means unsupervised algorithm in MATLAB. Ten salient explanatory land use variables are extracted for the calibration of the model. The model is calibrated by running a simulation for period 1984 to 2000. The computed logistic coefficients of the 10 explanatory variables show that all the 10 explanatory variables are significant at 95% confidence level based on a two-tailed test, since all the 10 variables yields p-values <0.05. The simulated map in 2000 is compared with the reference data in 2000; and evaluated using the Kappa statistic. The computed Kappa statistic is 0.7640; which implies a substantial agreement between the predicted and the reference data. The calibration model for 1984-2000 is used to predict 2005 map. A comparison of the predicted and reference data in 2005 yields Kappa statistics estimate of 0.6998; which indicates a substantial agreement between the predicted and the reference data. A prediction of 2030 is derived upon satisfactory result obtained for the 2005 prediction based on the 1984-2000 calibrated model. An urban expansion of 129.49% is predicted between 1984 and the forecasted 2030.

Keywords: Urban Expansion; Remote Sensing; GIS; Logistic Regression

1. Introduction

There has been a renewed focus on the study of urban systems in the last few years, as urbanisation remains a major development challenge exerting awesome pressure on social, economic and environmental sustainability (Pickett et al., 2001). Urbanisation may be the benchmark for measuring economic growth and development, however in the case of sprawling cities of developing countries it has been accompanied by poverty, unemployment, environmental degradation, decaying infrastructure, and uncontrollable growth of informal settlements (Angotti, 1993). Urban sprawl in Lagos (see Figure 1) has put profound pressure on housing, infrastructure, and the environment (Abiodun, 1974; Braimoh & Onishi, 2007). The social and environmental repercussion of loosely planned urban cities could be catastrophic especially in the present situation in Lagos that has constantly experienced remarkable urban expansion in a short period of time (Barredo et al., 2004). Predicting and understanding urban growth and change are critical to city planners and resource managers especially in rapidly changing environments (Knox, 1993; Turner et al., 1993).

The urban system is a very complex system and full of uncertainty; however, with the help of modelling and simulation, it is possible to reduce the uncertainty and increase the understanding of urban systems. Taking urban planning process as an example, planning is a future-oriented activity, strongly conditioned by the past and present; planners need to enhance their analytical, problem solving and decision making capabilities.

With the help of models, planners can facilitate scenario building and provide an important aid to future directed decision-making (Cheng, 2003). In this way, urban planning can be more scientific and reduce the subjectivity brought by decision makers. It is on this background that this research work is undertaken to implement a loosely-coupled Geographic Information Systems (GIS) Logistic Regression (LR) model in predicting future urban expansion in Lagos, Nigeria.

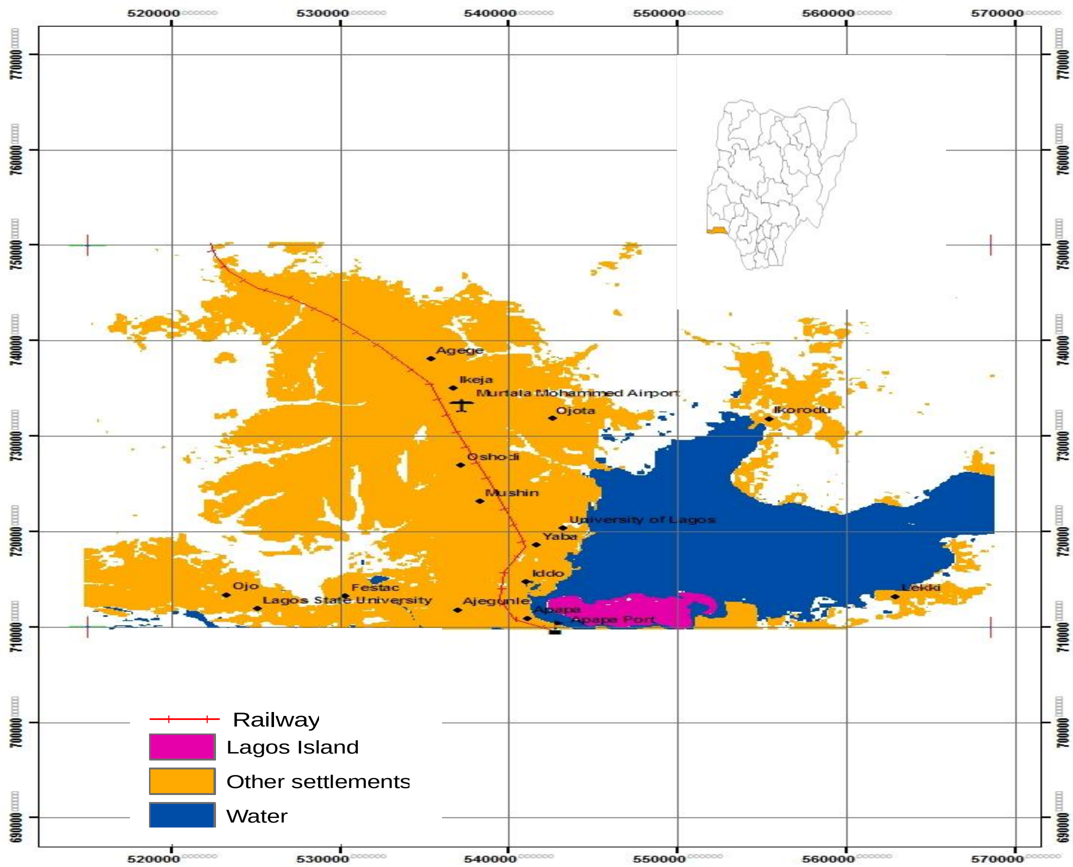


Figure 1 Lagos in relation to Nigeria

2. Data preparation

Historical Landsat TM images of Lagos acquired in 1984, 2000 and 2005 were used for the experiment (see Figures 2, 3, and 4). The Landsat images were classified in MATLAB using the k-means unsupervised algorithm. The result of the classification was visualised in ArcGIS. The classification results are shown in Figures 5, 6, and 7.

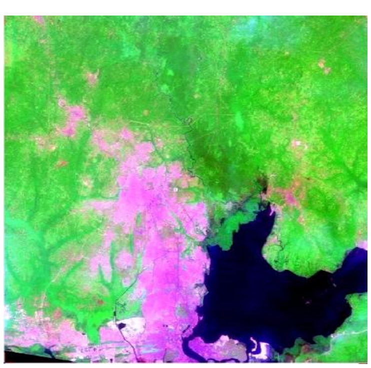


Figure 2 False colour 1984

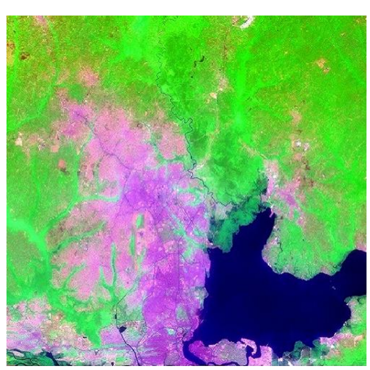


Figure 3 False colour 2000

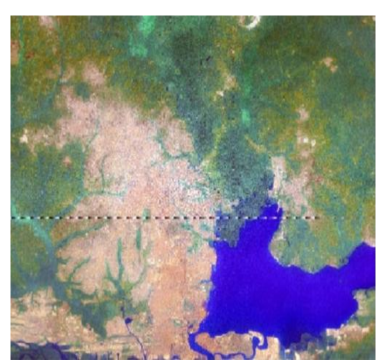


Figure 4 False colour 2005

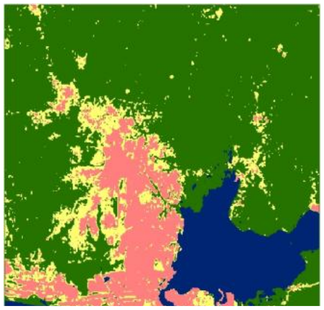


Figure 5 ***Classified image 1984***

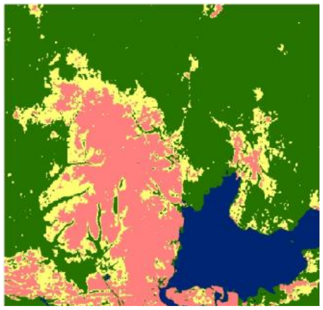


Figure 6 ***Classified image 2000***

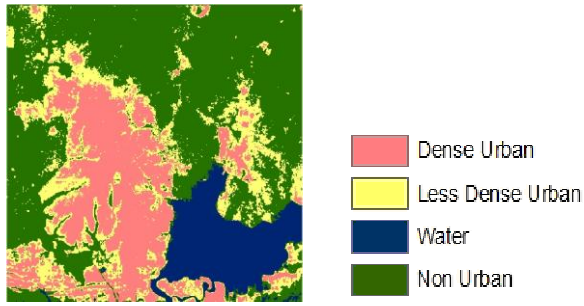
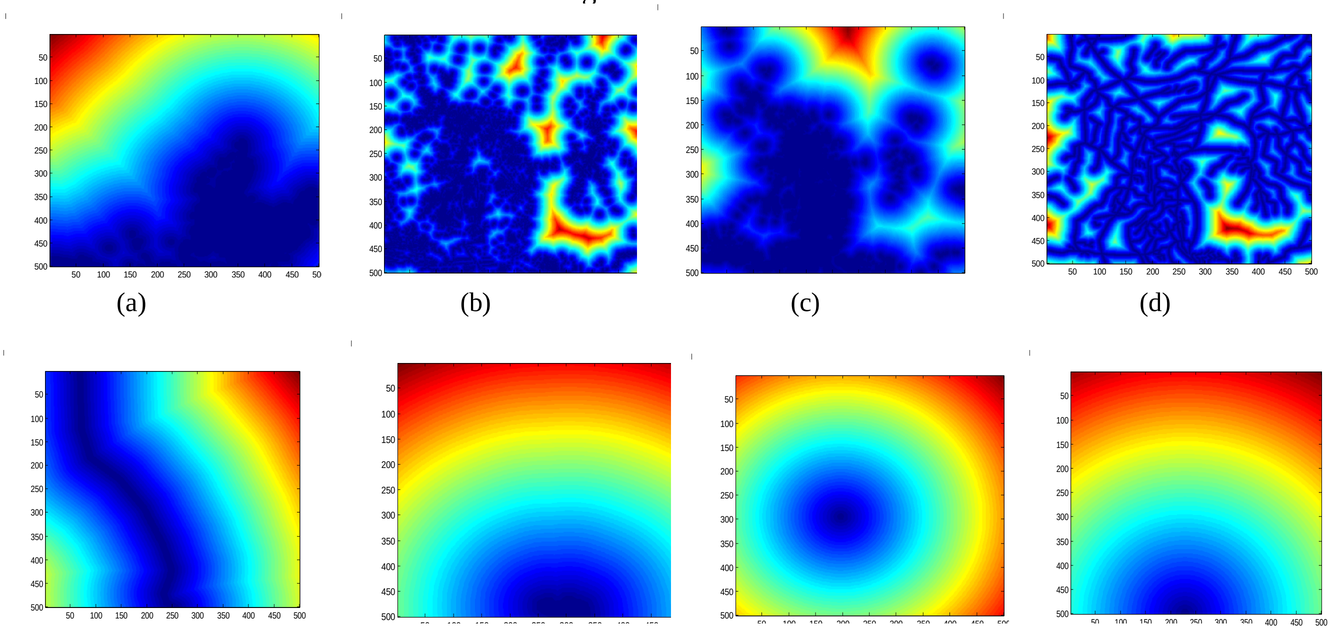


Figure 7 ***Classified image 2005***

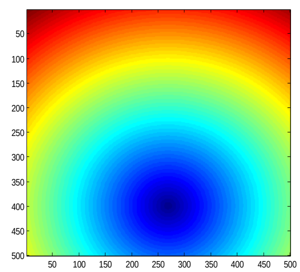
The selection of the forces driving land use is a crucial aspect of urban growth modelling, because land use drivers are the main characteristics that can help to understand the processes of land use transition (from nonurban to urban). There is no hard-and-fast rule or known global formula for selecting land use drivers. The list of land use drivers could be endless. Land use drivers are usually chosen on a case-to-case basis. Land use drivers in one environment might not apply to another (Baker, 1989). Ten salient explanatory variables extracted in ArcGIS and visualised in MATLAB were used for the modelling. The extracted variables are proximity variables from the 1984 image extracted in ArcGIS; they are: distance to water, distance to medium dense urban (residential structures), distance to heavily dense urban (industrial & commercial centres), distance to major Roads, distance to railway, distance to Lagos Island, distance to international airport, distance to international Seaport, distance to University of Lagos, and distance to Lagos State University (see Figure 8).

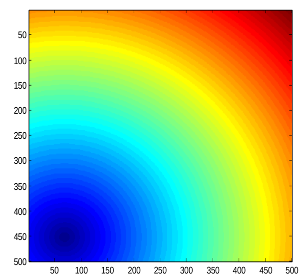
The variables were extracted as follows: distance to water was extracted by calculating the Euclidean distance of all locations to the nearest water body; distance to less dense urban centres was extracted by calculating the Euclidean distance of cells to the nearest less dense urban; distance to dense urban centres was extracted by calculating the Euclidean distance of cells to the nearest industrial and commercial centre; distance to major roads was extracted by calculating the Euclidean distance of all locations to the nearest major road; distance to railway was extracted by calculating the Euclidean distance of cells to the nearest railway; distance to Lagos Island was extracted by calculating the Euclidean distance of all cells to Lagos Island; distance to international airport was extracted by calculating the Euclidean distance of cells to the location of the international airport; distance to international seaport was extracted by calculating the Euclidean distance of all cells to the location of the international seaport; distance to University of Lagos was extracted by calculating the Euclidean distance of all cells to the location of University of Lagos; distance to Lagos State University was extracted by calculating the Euclidean distance of all cells to the location of Lagos State University.



(a) (b) (c) (d)

(e) (f) (g) (h)





(i) (j)

Figure 8 (a) Distance to water (b) Distance to less dense urban (c) Distance to dense urban (d) Distance to major roads (e) Distance to railway (f) Distance to Lagos Island (g) Distance to Lagos international airport (h) Distance to Lagos Seaport (i) Distance to University of Lagos (j) Distance to Lagos State University

3. *Application*

3.1 Logistic regression

The LR model is the most commonly used parametric model for modelling land use change (Verburg et al., 2004). LR is the linear regression model usually used in cases where the dependent variable is dichotomous [0, 1]. Like all linear regression models, the logistic regression is a predictive model. Logistic regression is used to describe data and to explain the relationship between one dependent binary variable and one or more metric (interval or ratio scale) independent variables. A mathematical illustration of the LR model is given by this linear function,

$$q_i = \beta_0 + \sum_p \beta_p x_{ip} + \varepsilon_i \tag{1}$$

where q is a binary dependent variable, $\beta_0,...\beta_i$ are logistic regression coefficients to be estimated, while x_i are independent variables and ε_i is the residual error.

At the centre of the logistic regression analysis is the task of estimating the log odds of an event. Mathematically, logistic regression estimates a multiple linear regression function defined as logit(P). A logistic regression model can therefore be expressed as:

$$q = \ln\left(\frac{P(y=1/x)}{1-P(y=1/x)}\right), \tag{2}$$

where P is the probability that $q=1$, given x_i independent variables;

$\frac{P(y=1/x)}{1-P(y=1/x)}$ is called the odds, while $\ln\left(\frac{P(y=1/x)}{1-P(y=1/x)}\right)$ is called the logit.

Therefore,

$$P = \frac{e^q}{1+e^q} = \frac{e^{\beta_0+\sum_{i=1}^{n}\beta_i x_i}}{1+e^{\beta_0+\sum_{i=1}^{n}\beta_i x_i}} = \frac{1}{1+e^{-(\beta_0+\sum_{i=1}^{n}\beta_i x_i)}} \quad \text{(Pohlmann \& Dennis, 2003)} \tag{3}$$

3.2 Modelling

The model was calibrated by running a simulation for period 1984 to 2000. Only two maps are basically required for land use change calibration (see Wu, 2002; Li & Yeh, 2002; Braimoh & Onishi, 2007; Kocabas & Dragicevic 2007).

The calibrated simulation of sixteen years was tested by comparing it with the reference landuse data for 2000. Once the result of a calibration is satisfactory, the future prediction of urban growth can be carried out using the parameters of the already calibrated model, assuming that the calibrated factors will remain relatively stable during the period to be predicted. The 1984 image was calibrated and validated using 2000. The model was then used to predict 2005 and 2030 (see flowchart in Figure 9).

The urbanised areas were extracted to create an overlay map for calibration from 1984-2000 (by overlaying 1984 and 2000 maps) (see Figure 10). Three categories of land use were derived from the overlay: (i) undeveloped region; (ii) change region; and (iii) developed region.

A total of 10,000 points were selected randomly for training; 5000 from the developed cells and 5000 from undeveloped cells. No point was selected from the change region since it is not common to both 1984 and 2000 maps. In order to implement the training in MATLAB, the extracted 10 variables were imported into MATLAB from ArcGIS. The dependent variable comprises of ones (+1) (urban) and zeros (0) (non-urban), in compliance with the logistic regression model.

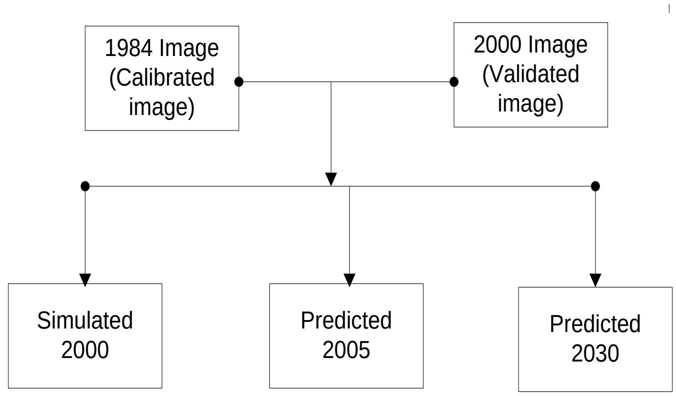


Figure 9 Modelling and prediction flow chart

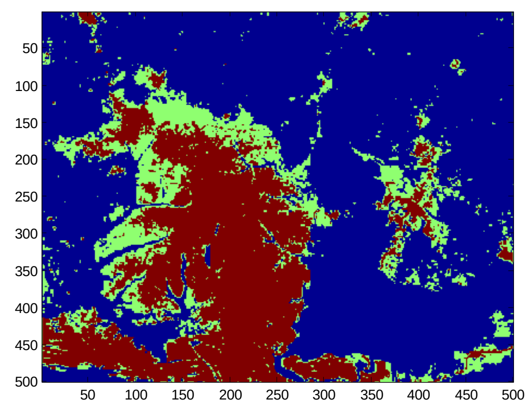


Figure 10 Overlay of 1984 with 2000 (blue=non-urban; green=change; red=urban)

The statistical results of each explanatory variable are given in table 1. All the 10 variables are significant in the logistic regression model based on a two-tailed test at 95% Confidence Level (CL). From Table 1, p-values less than 5% are considered significant in the model. An explanatory variable with the highest coefficient (in absolute value terms) is the most significant while an explanatory variable with the lowest coefficient is the least significant. Therefore distance to less dense urban is the most significant in the model while distance to University of Lagos is the least significant in the model.

Table 1 Logistic regression results for 1984-2000 simulation (* significant at <0.05)

Variable	Coefficient	Standard error	P-value
Intercept	7.039	0.3602	4.63E-85
Distance to water	1.1985	0.3871	0.00195972378911621*
Distance to less dense urban	-34.0887	2.129	1.06233080997733E-57*
Distance to dense urban	-14.2074	1.0864	4.42790863979757E-39*
Distance to major roads	-0.0529	0.0114	0.0000037168120729672*
Distance to railway	0.0106	0.0031	0.000698520607956676*
Distance to Lagos Island	0.0225	0.0058	0.000110473142347996*
Distance to international airport	-0.0207	0.0029	1.93156725438357E-12*
Distance to seaport	-0.0445	0.0084	1.05932642838911E-07*
Distance to University of Lagos	0.0071	0.005	0.0150303924514527*
Distance to Lagos State University	0.0112	0.0028	0.0000572382143319916*

The simulation of 2000 from the 1984 was executed at a transition probability cut of 0.303 (for the 16 years period). The simulated 2000 map was validated with the actual map of 2000, by comparing the True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN) matches using both maps. The overlay map of the simulated and actual 2000 maps is presented in Figure 11. A performance matrix derived from the result shown in Figure 11 is given in Table 2.

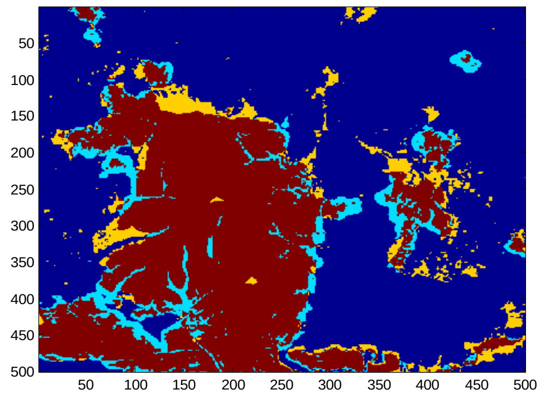


Figure 11 *Simulation result of 2000*

TP (True Positive) TN (True Negative)

FP (False Positive) FN (False Negative)

Table 2 Performance matrix of the simulation result for 2000

	Reference data 2000	
	Developed	Undeveloped
Simulated data 2000		
Developed	75,593 (TP)	15,213 (FP)
Undeveloped	11,863 (FN)	147,331 (TN)

The Kappa statistic was used to evaluate the performance matrix in table 2. Kappa statistic can be expressed mathematically as,

$$K = \frac{P_o - P_c}{1 - P_c} \tag{4}$$

where,

$$P_o = \sum_{i=1}^{m} P_{ii} = \frac{1}{N}\sum_{i=1}^{m} n_{ii} \tag{5}$$

and,

$$P_c = \sum_{i=1}^{m} P_{i+}P_{+i} = \frac{1}{N^2}\sum_{1=1}^{m} n_{i+}n_{+i} \quad , \tag{6}$$

(Ma & Redmond, 1995; Lo & Yeung, 2007).

Where,

P_o = proportion agreement observed

P_c = proportion agreement expected by chance

n_{ii} = the total number of correctly classified points by class along the diagonal of the error matrix

N = the total number of points checked (sampled)

P_{ii} = the proportion of correctly classified sample points by class at the diagonal of the error matrix (i.e. n_{ii} / N)

P_{i+} = the marginal distribution of the sample data (n_{i+} / N where n_{i+} is the row sum by class)

P_{+i} = the marginal distribution of the reference data (n_{+i} / N where n_{+i} is the column sum of class)

m = the total number of classes

The Kappa statistic is more reliable than other validation techniques because it has the ability to evaluate the actual agreement and chance agreement (Fung & LeDrew, 1988). Kappa statistic is computed from a performance matrix resulting from the comparison of the reference with the predicted data. The computed Kappa statistic using the performance matrix in Table 2 was 0.7640. According to Landis and Koch (1977) Kappa result can be appraised using the interpretation given in Table 3. Using Table 3 the computed Kappa statistic implied that the predicted data had a substantial agreement with the reference data.

Since the result of the calibration was satisfactory, the future prediction of urban expansion was carried out using the parameters of the already calibrated model. First to test the predictive ability of the calibration (from 1984-2000) on forecasting future urban expansion, known urbanized area of 2005 was first predicted using transition probability cut (less than that of 1984-2000) for 21years period (1984 – 2005) and validated using the reference data of 2005. The overlay map for the predicted and reference maps and the performance matrix for 1984-2005 are presented in Figure 12 and Table 4 respectively.

Table 3 Interpretation of Kappa statistic

KAPPA	INTERPRETATION
< 0	No agreement
0.0 – 0.20	Slight agreement
0.21 – 0.40	Fair agreement
0.41 – 0.60	Moderate agreement
0.61 – 0.80	Substantial agreement
0.81 – 1.00	Almost perfect agreement

Table 4 Performance matrix of the prediction result for 2005

	Reference data 2005	
	Developed	Undeveloped
Predicted data 2005		
Developed	78,405 (TP)	13797 (FP)
Undeveloped	21,777 (FN)	136,021 (TN)

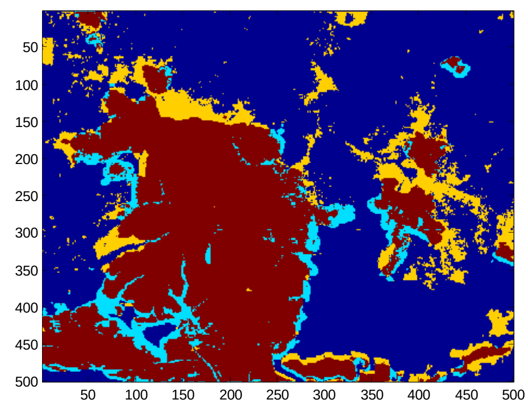


Figure 12 *Prediction result of 2005*

TP (True Positive) TN (True Negative)
FP (False Positive) FN (False Negative)

The computed Kappa statistic was 0.6998. Based on the interpretation on Table 3, the prediction for 1984-2005 had a substantial agreement with the reference data. Since the calibrated has been proven to be reliable based on the testing of the calibration with the 2005, a future predicted was implemented by lowering the transition probability cut below that used for the 1984-2005 prediction. A prediction for 46 years was obtained for the period 1984-2030 (see Figure 13). From Table 5, using the actual maps (between 1984-2005) Lagos expanded by 56.90% from 1984-2000 and by 64.04% from 1984-2005. An urban expansion of 129.49% was predicted between 1984 and the forecasted 2030.

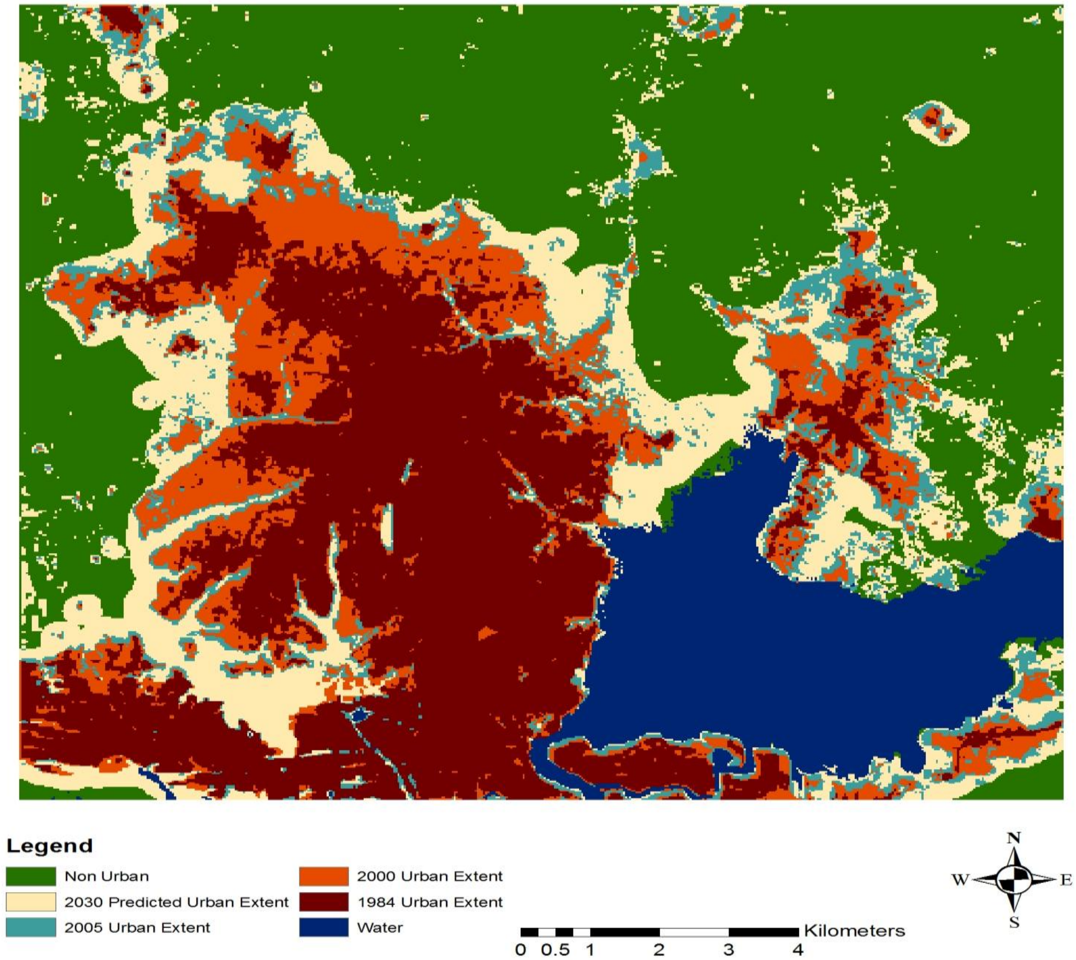


Figure 13 Pedicted result for 2030 in relation to 1984, 2000 & 2005

Table 5 Urban expansion from 1984 to the predicted 2030

Year	No of urban cells	Urban expansion
1984	57,882 cells	-
2000	90,814 cells	32,932 cells (56.90%) (1984-200)
2005	94,948 cells	37,066cells (64.04%) (1984-2005)
2030	132,836 cells	74,954cells (129.49%) (1984-2030)

4. Conclusion

The ten land use explanatory variables that were used for the modelling were all significant at 95% CL which implied that all the explanatory variables contributed to the urban expansion of Lagos. Distance to urban centre (Residential and commercial/industrial) had the highest impact in the model indicating that urban growth tends to occur close to the nearest urban area. The predicted result obtained for 2030 produces a valuable insight for planners. The spatial growth of Lagos in 2030 is clearly a direct consequence of urban sprawl as urban areas tend to expand close to the nearest urban cluster. So the question is what urban development and planning measures can be applied to manage the future urban expansion of Lagos? Lagos is the largest and the most congested city in Nigeria; and as a result, poses the greatest planning burden to Lagos State and the Nigerian government. Therefore, incorporating predictive models into planning will transform the present prevailing spontaneous planning into proactive and sustainable planning. The result of this research will not only assist the urban planners in foreseeing the future urban expansion in Lagos, but will also serve as a signal to planners in Lagos as well as other sprawling cities in Nigeria that there is an urgent need to control the incessant urban sprawl ravaging most cities in the county; which has put profound pressure on housing, infrastructure, and the environment, thus causing environmental degradation and the decaying of infrastructure.

References

Abiodun, J. O. (1974). Urban growth and problems in metropolitan Lagos. *Urban Studies, 11* (33), 341-347.

Angotti, T. (1993). *Metropolis 2000: Planning, poverty and politic*s. London, England: Routledge.

Baker, W. L. (1989). A review of models of landscape change. *Landscape Ecology 2*, 111–133.

Barredo, J.I., Demicheli, L., Lavelle, C., Kasanko, M., & McCormick, N. (2004). Modelling future urban scenarios in developing countries: An application case study in Lagos, Nigeria. *Environment and Planning B: Planning and Design, 32,* 65-84.

Braimoh, A. K., & Onishi, T. (2007). Spatial determinants of urban land use change in Lagos,

Cheng, J., 2003. *Modelling spatial and temporal urban growth*. Enschede. NL ITC: 203.

Fung, T. & LeDrew, E. (1988). The Determination of optimal threshold levels for changes detection using various accuracy indices. *Photogrammetric Engineering and Remote Sensing, 54*(10), 1449-1454.

Knox, P. L. (1993). *The restless urban landscape.* Englewood Cliffs, NJ: Prentice-Hall.

Kocabas, V., & Dragicevic, S., (2007). Enhancing a GIS cellular automata model of land use change: Bayesian networks, influence diagrams and causality. *Transactions in GIS 11*(5), 681-702.

Landis, J. R. & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics, 33*,159 – 174.

Li, X., & Yeh, A. G. (2002). Neural-network-based cellular automata for simulating multiple land use changes using GIS. *International Journal of Geographical Information Science*, *16*(4), 323-343.

Lo, C. P. & Yeung, A. K. W. (Eds.) (2007). *Concepts And Techniques Of Geographic Information Systems* (2^{nd} ed.). Upper Saddle River, NJ: Pearson Prentice Hall.

Ma, Z. & Redmond, R. L. (1995). Tau coefficients for accuracy assessment of classification of remote sensing data. *Photogrammetric Engineering and Remote Sensing 61*, 435-439.

Pickett, S.T.A., Cadenasso, M.L., Grove, J.M., Nilon, C.H., Pouyat, R.V., Zipperer, W.C., Constanza, R., 2001. Urban ecological systems: linking terrestrial ecological, physical, and socioeconomic components of metropolitan areas. Annual Review of Ecology and Systematics 32,127–157.

Pohlmann, J. T., & Dennis, W. L. (2003). A comparison of ordinary least squares and logistic regression. *Ohio Journal of Science, 103* (5), 118-125.

Turner, B. L., Moss, R. H., & Skole, D. L. (1993). *Relating land use and global land cover change: A proposal for an IGBP-HDP core project*. International-Geosphere- Biosphere Program, IGBP Report No. 24, HDP Report No:5. Stockholm: Royal Swedish Academy of Sciences.

Verburg, P. H., Schot, P., Dijst, M., & Veldkamp, A. (2004). Land use change modelling: current practice and research priorities. *Geojournal, 61*(4), 309–324.

Wu, F. (2002). Calibration of stochastic cellular automata: The application to rural urban land conversions. *International Journal of Geographical Information Science, 16*(8), 795-818.