Applied GIS

a free, international, refereed e-journal
(ISSN: 1832-5505)

URL:
http://www.appliedgis.net

MANAGING EDITOR:
Ray Wyatt – ray.wyatt@unimelb.edu.au

Volume 10, Number 1
March, 2014

CONTENTS:

All papers published during 2014 are part of *Volume* 10.
Each paper constitutes one *Number.*

Hence this paper should be cited as:
Okwuashi, O. & Ikediashi, D.I. (2013) – *GIS-based simulation of land use change*, *Applied GIS*, 10(1), 1-18

GIS-based simulation of land use change

Onuwa Okwuashi
Department of Geoinformatics & Surveying, University of Uyo, Nigeria
onuwaokwuashi@gmail.com

Dubem Isaac Ikediashi
School of the Built Environment, Heriot-Watt University, United Kingdom
dii1@hw.ac.uk

Abstract - This paper describes an application of *ArcGIS* to simulate land use changes in Lagos, Nigeria, using both Ordinary Least Squares (OLS) regression and Geographically Weighted Regression (GWR), over three epochs - 1963-1978, 1978-1984 and 1984-2000. Twelve salient, causal factors thought to be related to urban land use change in Lagos were used, such as distance to water, distance to residential structures, income potential and population potential. OLS was used to establish the regression coefficients, gauge the significance of each explanatory variable and estimate conformity with linear regression criteria. GWR was then used to simulate the urban form based on the results of the OLS model. For the three epochs the respective Kappa statistics for the simulated maps were 0.8858, 0.8366 and 0.8812, indicating an almost perfect agreement with the data of 1978, 1984 and 2000.

Keywords – GIS, linear regression, land use change, ordinary least squares, geographically weighted regression

1. Introduction

This research aimed to evaluate the use of the GIS for simulating land use changes in Lagos, Nigeria from 1963-1978, 1978-1984 and 1984-2000, based on the Ordinary Least Squares (OLS) regression and Geographically Weighted Regression (GWR) models. GWR (Fotheringham et al., 2002) is the local equivalent of global OLS.

The OLS model can be expressed mathematically as,

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

where:

y_i are the dependent variables,

β_0 denotes the intercept,

β_p is the slope coefficient for the p variable,

x_{ip} denotes the value of the p variable for i number of observations, and

ε_i represents the error parameter.

The GWR model can be expressed mathematically as,

$$y_i = \beta_0(u_i, v_i) + \sum_q \beta_q(u_i, v_i)x_{iq} + \varepsilon_i \tag{2}$$

where:

β_0 denotes the intercept,

β_q is the slope coefficient for the q variable,

x_{iq} denotes the value of the q variable for i number of observations,

ε_i represents the error parameter, and

(u_i, v_i) stands for the coordinates of the ith location in i observations.

Unlike GWR, OLS assumes that β_p is stationary or homoscedastic. This is a major difference between OLS and the GWR (Fotheringham et al., 2002; Mennis, 2006). That is, GWR takes the effect of spatial dependency into consideration by assuming that β_q is non-stationary or heteroscedastic. This means that the solutions of β_q vary across the globe for the same values of x_{iq}. Wherever x_{iq} is observed, the solution of β_q is affected by the location (u_i, v_i) (Fotheringham et al., 2002).

Nevertheless, in spite of the merits of the GWR model over the OLS model, it is still subject to the fundamental statistical assumptions that govern all linear regression models -

1. variables are normally distributed,
2. there is a linear relationship between the dependent and each independent variable,
3. there is no multicollinearity between the independent variables, and
4. there is no spatial autocorrelation

(Leung et al., 2000; Fotheringham et al., 2002; Wheeler & Tiefelsdorf, 2005; Farber & Páez, 2007).

In terms of choosing the independent variables, there are no hard-and-fast rules, or known global formula, for selecting land use drivers. Indeed, the list could be endless. Land use drivers are usually chosen on a case-to-case basis, because those used in one environment might not apply to another (Baker, 1989).

Hence our modelling used our own twelve, salient, causal factors which were thought to be related to urban change in Lagos, Nigeria:

1. distance to water,
2. " residential structures,
3. " industrial and commercial centres,
4. " major roads,
5. " railway,
6. " Lagos Island,
7. " international airport,
8. " international seaport,
9. " University of Lagos,
10. " Lagos State University,

11. income potential, and
12. population potential.

These variables overlap, albeit partially, those that have been chosen in several studies that were similar to our own, for example, the OLS and GWR models implemented by Thapa & Murayama (2009), Noresah & Ruslan (2009) and Shariff et al. (2010).

More specifically, Thapa & Murayama (2009) modelled Kathmandu's land use changes using:

1. population change,
2. distance to road,
3. " existing built-up areas,
4. " industrial estates,
5. " rivers,
6. " agricultural areas,
7. " forests, and
8. " shrubs lands,

Noresah & Ruslan (2009) modelled land use change in Malaysia based on the following variables:

1. road network distance to nearest interchange,
2. " centre in the study,
3. " nearest employment centre,
4. " nearest regional centre,
5. " nearest train station,
6. " nearest industrial area,
7. " nearest sea port,
8. " nearest airport,
9. proximity (Euclidean distance) to nearest urban built-up area,
10. " town of Sungai Petani,
11. " Federal Route 1,
12. " Federal Route 67,
13. neighbouring cells' proportion of residential land,
14. cells that fall on land zoned for residential use,
15. " land zoned for commercial use,
16. " land zoned for industrial use,
17. slope steepness, and
18. amount of land available for new development,

and Shariff et al. (2010) modelled land use change in Malaysia based on:

1. proximity to nearest expressway,
2. " Georgetown,
3. " nearest minor city centre,
4. " nearest highway,
5. " nearest airport,
6. " educational institutions,
7. " nearest major road,
8. " nearest forest reserve,
9. " population concentration centres,
10. " the 120-meter contour, and
11. " land available for development in 1990.

2. Data Preparation

Lagos has experienced rapid urban growth in a short period of time, due to unregulated planning (Abiodun, 1977). The city (Figure 1) is in a littoral environment, with relatively flat terrain and an area of about 2910km^2, lying between latitudes 6° 26´ and 6° 50´ N and longitudes 3° 09´ and 3° 46´ E (Braimoh & Onishi, 2007).

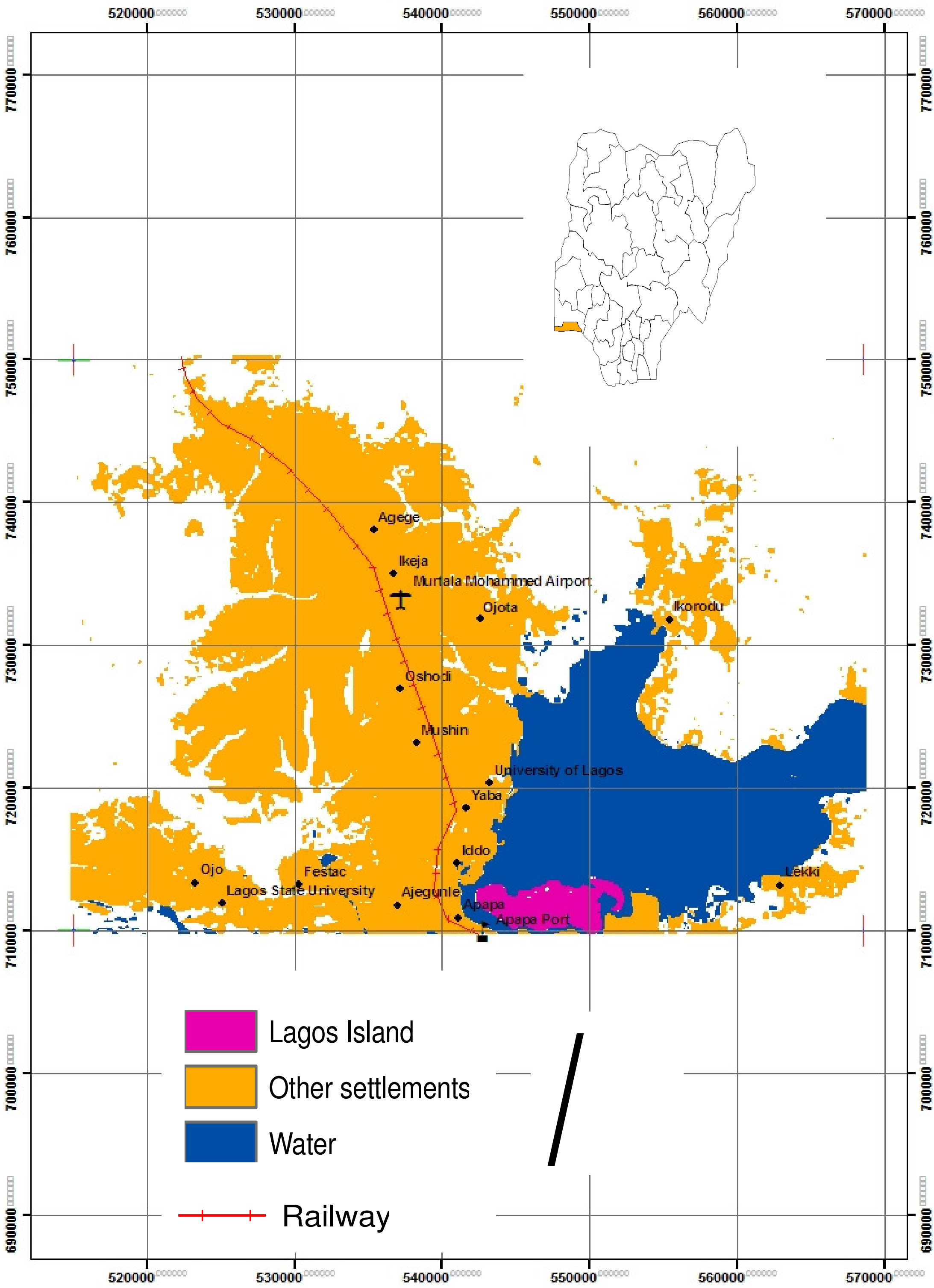

Figure 1 – Lagos, in Nigeria

The spatial data for our three land use epochs -1963-1978, 1978-1984 and 1984-2000 (Figure 2), were derived from historical land use information formulated from satellite remote sensing and base maps.

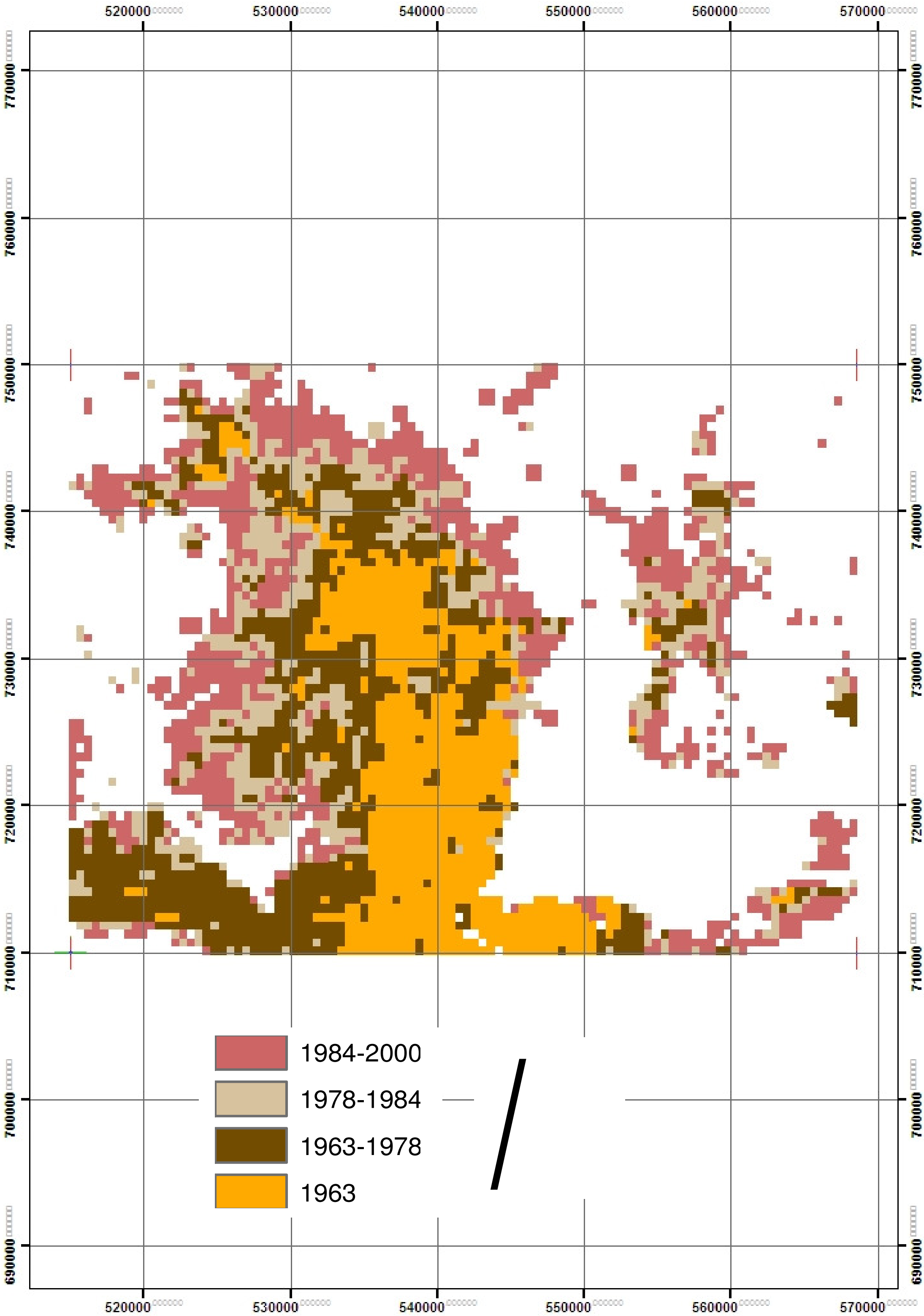

Figure 2 – The growth of Lagos

Our twelve land use explanatory variables were grouped into two categories:

1. proximity variables - distance to water, residential structures, industrial/commercial centres, major roads, railway, Lagos Island, international airport (1984-2000 only), international seaport, University of Lagos, Lagos State University (1984-2000 only), and
2. weighted variables - income potential and population potential.

The proximity variables were extracted using *ArcGIS* while the weighted variables were obtained using *MATLAB*.

Land use modelling from 1984-2000 was based upon all the twelve explanatory variables, whereas 10 explanatory variables were used for the periods of 1963-1978 and 1978-1984. This was because two of the selected land use drivers (international airport and Lagos State University) came into being after 1978. Dependent variables were represented as discrete variables.

All maps were gridded (as shown in Figure 2) before being modelled with the OLS and GWR models (Erener et al., 2010; Okwuashi, 2011). The present year and target year maps were overlaid to determine the changed regions between the target years, and so by the target year three categories of land use had been identified:

1. undeveloped;
2. changed; and
3. developed,

and these constituted the data for the dependent variable - undeveloped = 1; changed = 2; and developed = 3.

The original values of the independent variables were scaled to [0, 1] using a standard transformation formula (Gong, 1996; Li & Yeh, 2002):

$$x_i' = (x_i - \min)/(\max - \min) \qquad (3)$$

where:

x_i' is the land use variable,
$\min$ is the lowest value in the land use vector,
$\max$ is the highest value in the land use vector, and
x_i represents the land use variables.

This scaling technique is effective for ensuring that all the independent variables are equally weighted.

Equations 4-6 and equations 7-9 are the OLS and GWR equations respectively - for the periods of 1963-1978, 1978-1984 and 1984-2000 respectively;

$$LUC\_1963\text{–}1978_i = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \beta_5 x_5 + \beta_6 x_6 + \beta_7 x_7 + \beta_8 x_8 + \beta_9 x_9 + \beta_{10} x_{10} + e_i \qquad (4)$$

$$LUC\_1978\text{–}1984_i = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \beta_5 x_5 + \beta_6 x_6 + \beta_7 x_7 + \beta_8 x_8 + \beta_9 x_9 + \beta_{10} x_{10} + e_i \qquad (5)$$

$$LUC\_1984\text{–}2000_i = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \beta_5 x_5 + \beta_6 x_6 + \beta_7 x_7 + \beta_8 x_8 + \beta_9 x_9 + \beta_{10} x_{10} + \beta_{11} x_{11} + \beta_{12} x_{12} + e_i \qquad (6)$$

$$\begin{aligned} LUC\_1963\text{–}1978_i = {} & \beta_0 + \beta_1 x_1(u_1,v_1) + \beta_2 x_2(u_2,v_2) + \beta_3 x_3(u_3,v_3) + \beta_4 x_4(u_4,v_4) + \beta_5 x_5(u_5,v_5) + \beta_6 x_6(u_6,v_6) \\ & + \beta_7 x_7(u_7,v_7) + \beta_8 x_8(u_8,v_8) + \beta_9 x_9(u_9,v_9) + \beta_{10} x_{10}(u_{10},v_{10}) + e_i \end{aligned} \qquad (7)$$

$$LUC\_1978\text{-}1984 = \beta_0 + \beta_1 x_1(u_1,v_1) + \beta_2 x_2(u_2,v_2) + \beta_3 x_3(u_3,v_3) + \beta_4 x_4(u_4,v_4) + \beta_5 x_5(u_5,v_5) + \beta_6 x_6(u_6,v_6)$$
$$+ \beta_7 x_7(u_7,v_7) + \beta_8 x_8(u_8,v_8) + \beta_9 x_9(u_9,v_9) + \beta_{10} x_{10}(u_{10},v_{10}) + e_i \qquad (8)$$

$$LUC\_1984\text{-}2000 = \beta_0 + \beta_1 x_1(u_1,v_1) + \beta_2 x_2(u_2,v_2) + \beta_3 x_3(u_3,v_3) + \beta_4 x_4(u_4,v_4) + \beta_5 x_5(u_5,v_5) + \beta_6 x_6(u_6,v_6)$$
$$+ \beta_7 x_7(u_7,v_7) + \beta_8 x_8(u_8,v_8) + \beta_9 x_9(u_9,v_9) + \beta_{10} x_{10}(u_{10},v_{10}) + \beta_{11} x_{11}(u_{11},v_{11}) + \beta_{12} x_{12}(u_{12},v_{12}) + e_i \qquad (9)$$

where:

*LUC_* = Land Use Change

x_1, x_2, x_3, x_4 x_5, x_6, x_7, x_8, x_9 x_{10}, x_{11} and x_{12} represent the independent variables,

β_0 is the intercept

$\beta_1,...,\beta_{12}$ are the coefficients of the independent variables

u, v are the horizontal coordinates, and

e_i is the error term.

The data were prepared in *MATLAB* and thereafter imported into *ArcGIS* for modelling.

The dependent variables needed to be shape files consisting of discrete variables, and the attribute tables of the dependent variables had tot contain:

1. unique or primary keys,
2. all the attributes of the independent variables (explanatory variables), and
3. discrete values of the dependent variable.

The dependent variables could not be represented by using continuous variables because the resulting map-overlay generated only three land use types - undeveloped, change and developed.

3. Modelling

3.1 Using the ordinary least squares (OLS) tool

Firstly, we assessed whether the explanatory variables could effectively estimate the regression coefficients. All experiments were based on a two-tailed test at the 95% Confidence Level (CL). In *ArcGIS*, the OLS model furnishes the Joint Wald Statistic to rate the overall significance of the model, as shown in Table 1. The hypothesis for the Joint Wald Statistic can be stated as:

H_0: the explanatory variables in the model are not effective at the 95% CL

H_1: the explanatory variables in the model are effective (reject H_0 if p-value <0.05)

Periods	Jarque-Bera Statistic	Degrees of freedom	*P*-value
1963-1978	2420.8369	10	0.000000*
1978-1984	3228.5063	10	0.000000*
1984-2000	3627.0334	12	0.000000*

Table 1 - Joint Wald Statistic (*significant at p<0.05)

Since the Joint Wald Statistic indicated that the explanatory variables for all the three tested periods effectively generate viable regression coefficients, we assessed the significance of each one using the OLS model, as shown in Tables 2, 3 and 4. The latter show that even though not all explanatory variables were significant, their calculated t-statistics show that they all have an impact. Hypotheses for assessing the significance of each explanatory variable can be stated as follows:

H_0: the coefficients are zero at the 95% CL

H_1: the coefficients are not zero (reject H_0 if p-value <5%)

Variable	Coefficient	Std Error	t-statistic	P-value	VIF [1]
Intercept	0.847101	0.018673	45.363946	0.000000*	- - - - -
Distance to water	-0.117348	0.055944	2.097593	0.035964*	1.509747
Distance to residential areas	-1.124156	0.050378	22.314590	0.000000*	1.159984
Distance to industrial and commercial centres	-0.806965	0.045379	17.782810	0.000000*	1.342798
Distance to major roads	-0.421015	0.047971	8.776518	0.000000*	1.271059
Distance to railway	-0.716223	0.042507	16.849342	0.000000*	1.194308
Distance to Lagos Island	-0.397699	0.042050	9.457734	0.000000*	1.325239
Distance to international seaport	-0.174235	0.039471	4.414274	0.000013*	1.254065
Distance to University of Lagos	-0.062781	0.041584	1.509759	0.131166	1.274922
Income potential	0.520350	0.043490	11.964778	0.000000*	1.141347
Population potential	0.230385	0.035664	6.459927	0.000000*	1.141712

Table 2 - Statistical results for assessing the significance of each independent variable in the model of 1963-1978 land use change (*significant at p<0.05 or t>1.96)

Variable	Coefficient	Std Error	t-statistic	P-value	VIF [1]
Intercept	1.092893	0.022225	49.173422	0.000000*	- - - - - - -
Distance to water	-0.208187	0.065870	3.160574	0.001597*	1.440754
Distance to residential	-1.919315	0.062447	30.735343	0.000000*	1.244771
Distance to industrial and commercial	-1.010004	0.059138	17.078679	0.000000*	1.410719
Distance to major roads	-0.927202	0.074672	12.417024	0.000000*	1.086095
Distance to railway	-0.655001	0.052848	12.393995	0.000000*	1.270764
Distance to Lagos	-0.308435	0.050105	6.155783	0.000000*	1.321559

Island					
Distance to international seaport	-0.140088	0.047237	2.965640	0.003041*	1.236370
Distance to University of Lagos	0.137918	0.049860	2.766108	0.005689*	1.261713
Income potential	0.746732	0.054971	13.584038	0.000000*	1.178389
Population potential	0.339526	0.057042	5.952223	0.000000*	1.182373

Table 3 - Statistical results for assessing the significance of each independent variable in the model of 1978-1984 land use change (*significant at $p<0.05$ or $t>1.96$)

Variable	Coefficient	Std Error	t-statistic	P-value	VIF [1]
Intercept	-2.290807	0.023219	98.658942	0.000000*	- - - - - - -
Distance to water	-0.222466	0.075164	2.959756	0.003099*	1.176389
Distance to residential	-2.606189	0.077746	33.522009	0.000000*	1.104094
Distance to industrial and commercial	-2.113546	0.092167	22.931596	0.000000*	1.137527
Distance to major roads	-1.486506	0.106371	13.974726	0.000000*	1.063784
Distance to railway	-0.910023	0.058224	15.629605	0.000000*	1.050858
Distance to Lagos Island	-0.339675	0.056222	6.041727	0.000000*	1.093061
Distance to international airport	-0.035236	0.054862	0.642261	0.520725	1.102546
Distance to international seaport	-0.189215	0.062306	3.036886	0.002412*	1.093976
Distance to University of Lagos	0.147534	0.060228	2.449619	0.014312*	1.104240
Distance to Lagos State University	-0.719737	0.058858	12.228258	0.000000*	1.052826
Income potential	0.721885	0.063363	11.392814	0.000000*	1.154016
Population potential	0.414852	0.048891	8.485296	0.000000*	1.153289

Table 4 - Statistical results for assessing the significance of each independent variable in the model of 1984-2000 land use change (*significant at $p<0.05$ or $t>1.96$)

The multiple R^2 values for periods, for 1963-1978, 1978-1984, and 1984-2000, were:

0.32, 0.36 and 0.38 respectively, and the calculated the calculated adjusted R^2 values for the same periods were roughly the same.

We also tested for conformity to the assumptions on which OLS are based - multicollinearity, linearity, normality, spatial autocorrelation, and homoscedasticity/stationarity, and the Variance Inflation Factor (VIF), as shown in the right-hand columns of Tables 2, 3 and 4, was used to assess the effect of multicollinearity. Explanatory variables with a VIF of over 7.5 are considered redundant, and should be excluded from the model, but since our calculated VIF values were all less than 7.5, it was concluded that they are all important and so should be retained in the model.

To test linearity test we used scatter plots, an ideal form of which is shown in Figure 3a. Figure 3b plots a dependent variable (1963-1978) against an independent variable (distance to water 1963-1978) and it clearly shows that land use change did not conform to a typical linear regression relationship between dependent and an independent variables. It shows that the relationship between the land use dependent variable (y-axis) and the independent variable (x-axis) is nonlinear.

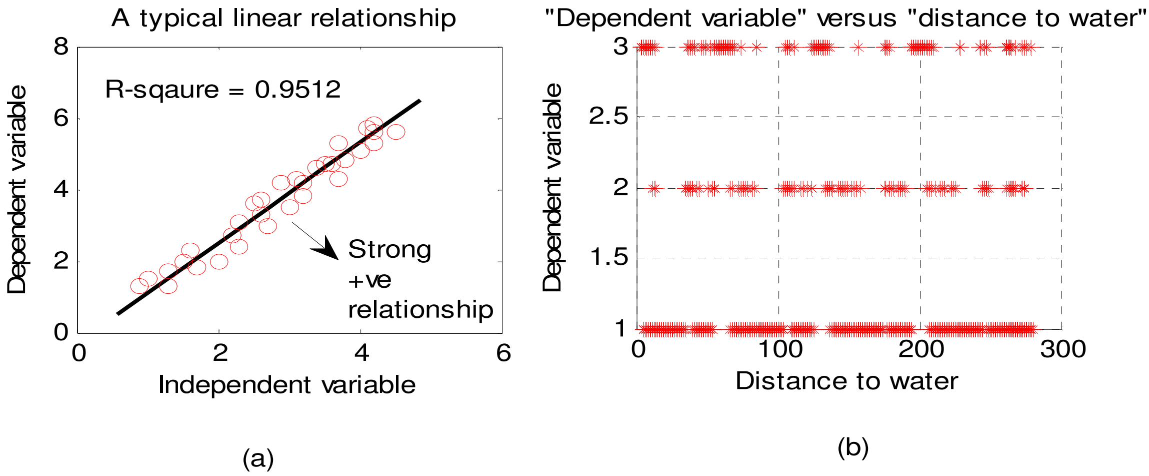

Figure 3 – Testing for linearity

The Jarque-Bera Statistic test (Table 5) was used to test whether the model residuals are normal:

H_0: the residuals are normally distributed at the 95% CL

H_1: the residuals are not normally distributed (reject H_0 if p-value <0.05)

and since all the calculated p-values for 1963-1978, 1978-1984, and 1984-2000 were <0.05, this indicates that the residuals deviated from the normal distribution.

Periods	Jarque-Bera Statistic	Degrees of freedom	*P*-value
1963-1978	706.1168	10	0.000000*
1978-1984	392.9615	10	0.000000*
1984-2000	318.7622	12	0.000000*

Table 5 - The Jarque-Bera Statistic test (*significant at p<0.05)

Hence plots for the periods of 1963-1978, 1978-1984 and 1984-2000, as shown in Figure 4 indicate that the model residuals deviate from a normal distribution. That was the reason why the *ArcGIS* Jarque-Bera test was statistically significant for all three periods.

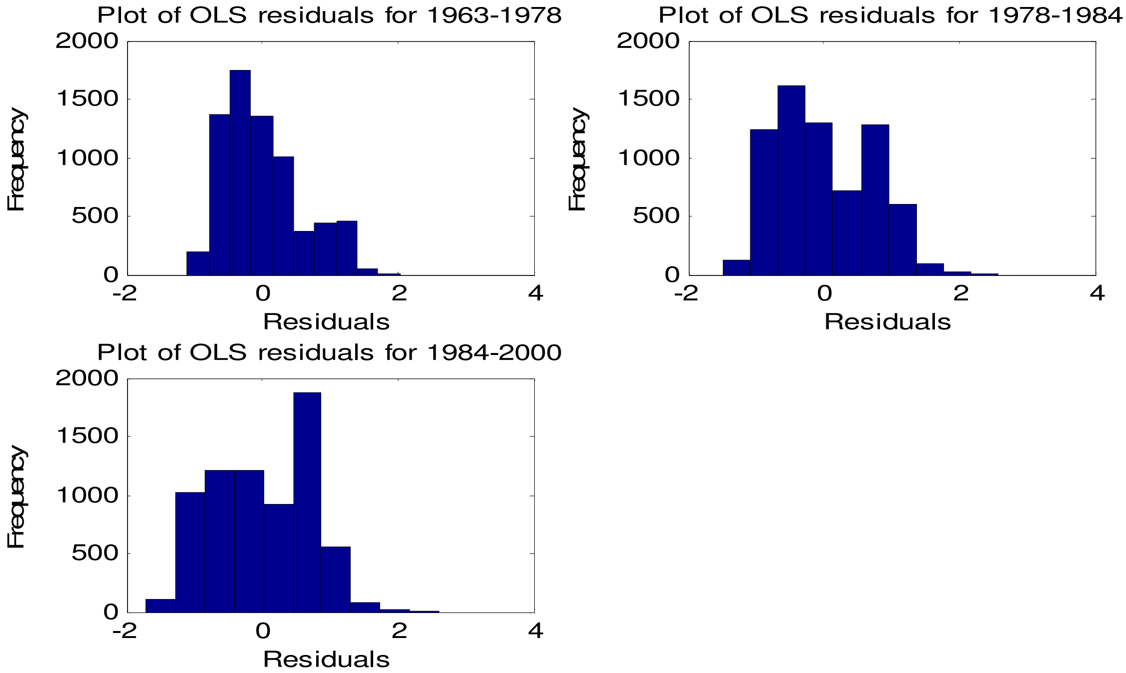

Figure 4 - Residual plots

To test for autocorrelation we used the Moran's I Tool. Moran I values close to +1 indicate positive spatial autocorrelation, values close to -1 indicate negative spatial autocorrelation and values close to zero indicate that the model residuals are random. Results are shown in Figure 6, which indicates that all three periods' model residuals were not spatially autocorrelated.

Periods	Moran's I index
1963-1978	0.0811
1978-1984	0.0740
1984-2000	0.0693

Table 6 – Results of the test for spatial autocorrelation

To address homoscedasticity or non-stationarity, the Koenker (BP) Statistic test was used, and results shown in Table 7.

H_0: the model is stationary at the 95% CL

H_1: the model is non-stationary (reject H_0 if p-value <0.05)

Periods	Koener (BP) Statistic	Degrees of freedom	*P*-value
1963-1978	1237.7354	10	0.000000*
1978-1984	532.0031	10	0.000000*
1984-2000	74.8344	12	0.000000*

Table 7 - Koenker (BP) Statistic (*significant at p<0.05)

All three results, for the periods of 1963-1978, 1978-1984 and 1984-2000, indicated significant non-stationarity, which implied that the model was not homoscedastic. Hence because the OLS model had significant non-stationarity, we needed to use the GWR model.

3.2 Using the geographically weighted regression (GWR) tool

It is desirable that the standard residuals (under and over prediction) are randomly distributed, and Figure 5 shows that the standard residuals for the three periods were indeed random. This implies that no key independent variable was omitted from the model (Thapa & Murayama, 2009). High local R^2 values indicate areas on the map where the GWR model predicted well, whereas low local R^2 values indicate areas on the map that were less well predicted by the GWR model. Cond denotes condition number.

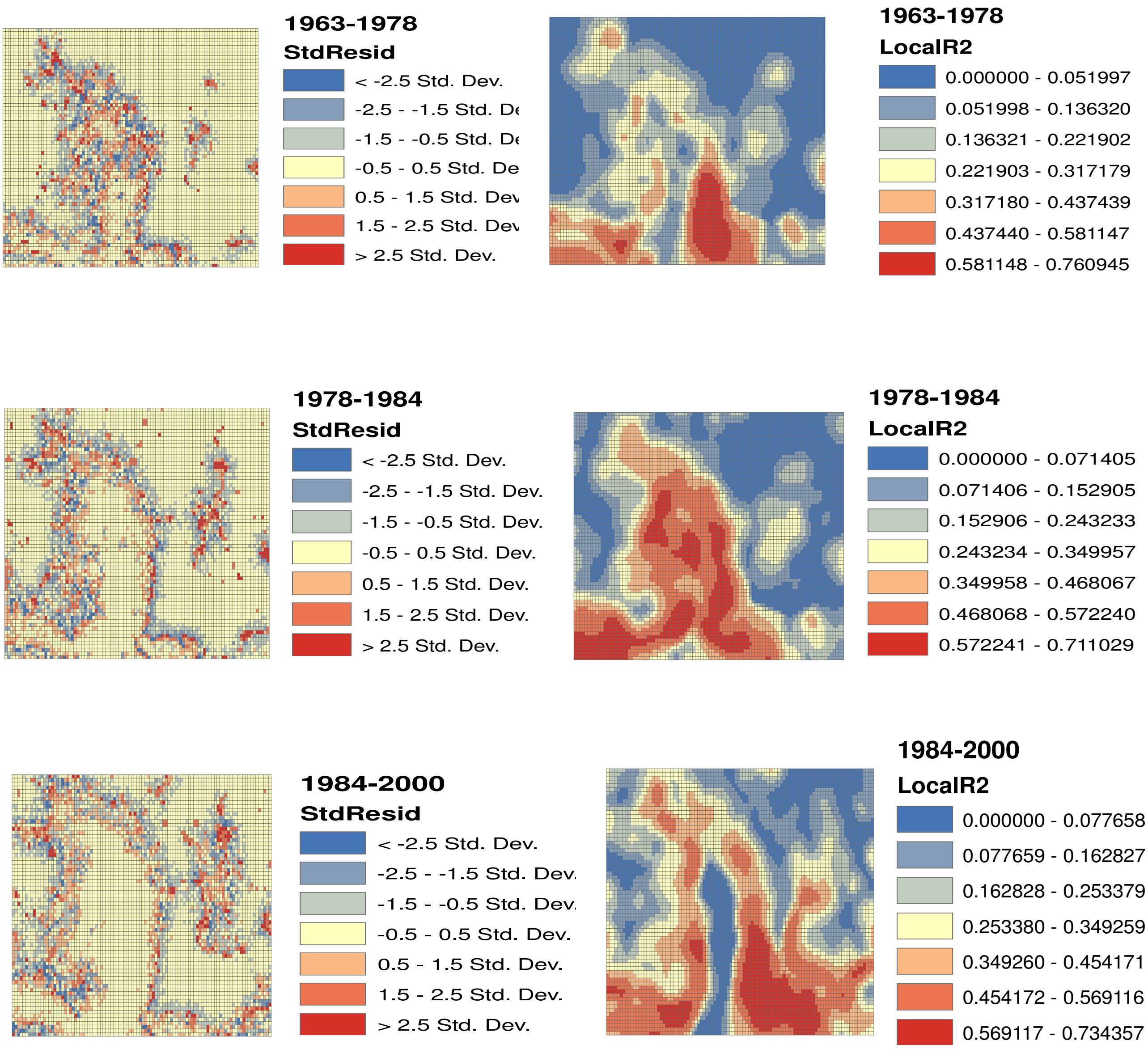

Figure 5 - Estimated standard residuals and local R^2 coefficients

In Figure 6, the Cond measurement assesses local collinearity in the GWR model. Regions with Cond values greater than 30 may not be reliable, but since the predicted Cond maps for the periods of 1963-1978, 1978-1984 and 1984-2000 all yielded Cond values below 30 it was concluded that the results from the model can be trusted.

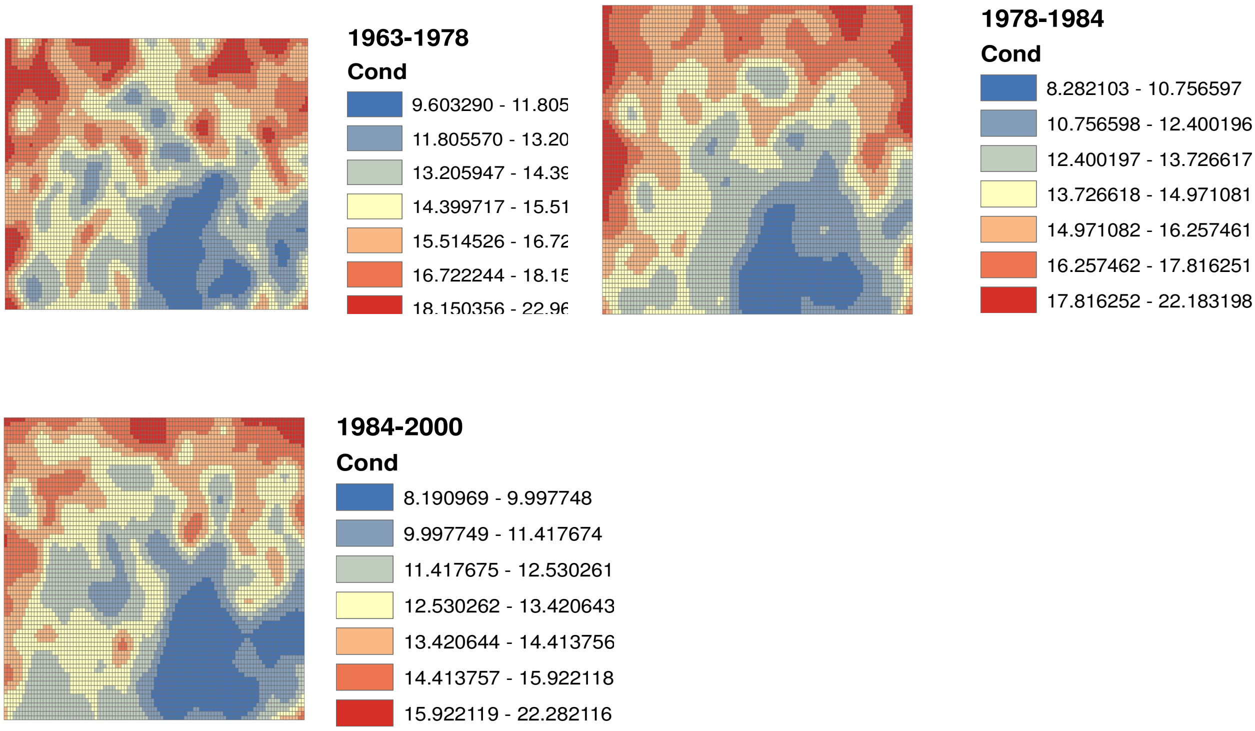

Figure 6 - Estimated condition numbers

Unlike the OLS model, the GWR model does not generate the Jarque-Bera Statistic, and so the only way to investigate the normality of the GWR model's residuals is to examine them graphically by plotting them. Plot (Figure 7) show that the residuals deviated slightly from a normal distribution but the GWR values were more normally distributed than those of the OLS model displayed in Figure 4. That is, the GWR model enhanced the normality of the model when compared to the OLS model.

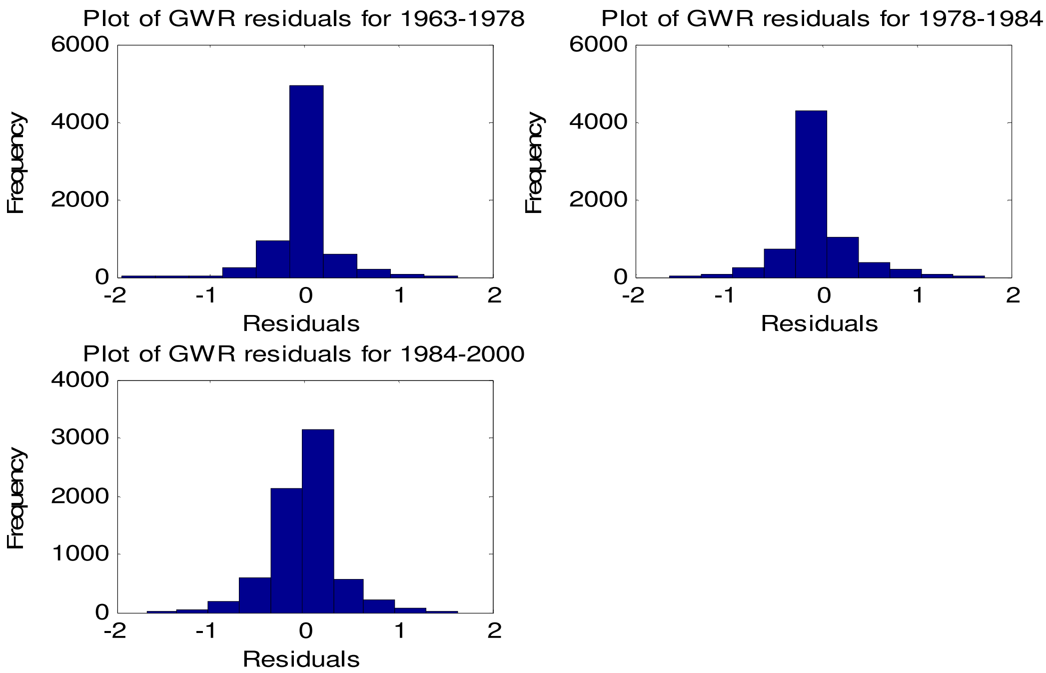

Figure 7 - Normality plot

Finally, the presence of spatial autocorrelation in the data of the GWR residuals was tested -even though the standard residual maps shown in Figure 7 suggested that residuals were random. *ArcGIS*'s Moran's I indices for 1963-1978, 1978-1984and 1984-2000 are shown in Table 8 and they indicate that the model residuals were random, or autocorrelation free, because values are close to zero. In other words, the GWR model reduced the autocorrelation effect in the model when compared with the OLS model (Table 6) - its results were much closer to zero.

Periods	Moran's I index for GWR
1963-1978	0.004824
1978-1984	0.005082
1984-2000	0.006313

Table 8 - Spatial autocorrelation test

Unlike logistic regression models, linear regression models do not yield probability values (Menard, 1995; Pohlmann & Dennis, 2003) of the target variable. In this work, therefore, a cut-off value is used to distinguish between the developed and undeveloped land use categories, thereby generating the GWR simulated maps for 1978, 1984 and 200 that are presented in Figure 8, where:

TN (True Negative) = *Undeveloped cells that were correctly predicted as undeveloped cells*

FP (False Positive) = *Undeveloped cells that were wrongly predicted as developed cells*

FN (False Negative) = *Developed cells that were wrongly predicted as undeveloped cells*

TP (True Positive) = *Developed cells that were correctly predicted as developed cells*

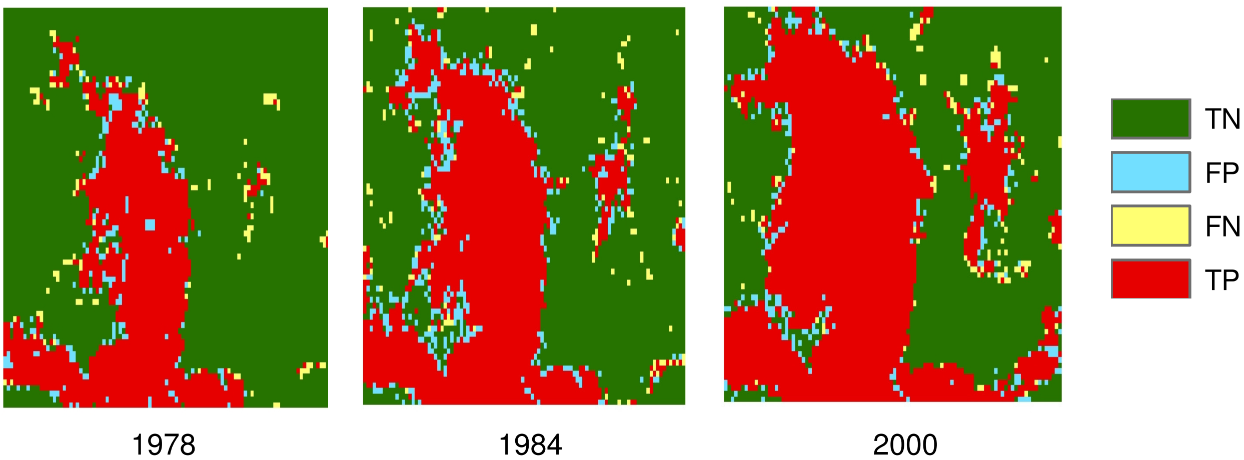

Figure 8 - Simulated maps

Cell-by-cell comparison of the simulated and the reference maps in 1978, 1984 and 2000 resulted in the performance matrices of Tables 9, 10 and 11. Conventionally, the rows of the matrix represent the values of the simulated data while the columns of the matrix represent the real values (Lo & Yeung, 2007).

	Reference data 1978	
	Developed	Undeveloped
Simulated data 1978		
Developed	1669 (TP)	185 (FP)
Undeveloped	123 (FN)	5023 (TN)

Table 9 - Performance matrix for period 1963-1978

	Reference data 1984	
	Developed	Undeveloped
Simulated data 1984		
Developed	2369 (TP)	399 (FP)
Undeveloped	139 (FN)	4093 (TN)

Table 10 - Performance matrix for period 1978-1984

	Reference data 2000	
	Developed	Undeveloped
Simulated data 2000		
Developed	3282 (TP)	266 (FP)
Undeveloped	150 (FN)	3302 (TN)

Table 11 - Performance matrix for period 1984-2000

The Kappa statistic was then calculated for Tables 9, 10 and 11. The Kappa statistic can be expressed mathematically as,

$$K = \frac{P_o - P_c}{1 - P_c} \tag{10}$$

where,

$$P_o = \sum_{i=1}^{m} P_{ii} = \frac{1}{N}\sum_{i=1}^{m} n_{ii} \tag{11}$$

and,

$$P_c = \sum_{i=1}^{m} P_{i+}P_{+i} = \frac{1}{N^2}\sum_{1=1}^{m} n_{i+}n_{+i} \quad , \tag{12}$$

(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

According to Landis and Koch (1977) the Kappa result, shown in Table 12, can be appraised using the interpretation given in Table 13, and Table 13 indicates that the computed Kappa statistic implies the simulated data almost perfectly agreed with the reference data.

Periods	Kappa statistic
1963-1978	0.8858
1978-1984	0.8366
1984-2000	0.8812

Table 12 - Calculated 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 13 - Interpretation of Kappa statistic

4. Conclusion

There was almost perfect agreement between the simulated/predicted and the reference land use data despite linear regression's statistical assumptions were not always met. The unavoidable representation of the dependent variables of land use data as discrete variables invariably makes the land use data non-linear.

So spatial statisticians on one hand find linear regression models very attractive and informative when applied to land use change modelling - the individual impact of explanatory variables can be carefully explored, including the benefit of exploring other statistical outputs for helping to explain the data in greater detail. On the other hand, traditional statisticians consider any modelling that violates fundamental statistical assumptions to be scientifically untenable. That is, the trade-off for deriving rich statistical outputs is the inevitable violation of the fundamental statistical assumptions underlying linear regression models.

It is important to state that the main use of the OLS and GWR models is to obtain statistical inferences such as the coefficient, the standard error, the t-statistic, the p-value and the VIF by relating a continuous independent variable with a set of independent variables (Leung et al., 2000; Fotheringham et al., 2002; Wheeler & Tiefelsdorf, 2005; Noresah & Ruslan, 2009; Thapa & Murayama, 2009; Shariff et al., 2010).

Nevertheless, our work here extended the GIS-based OLS and GWR models to the simulation of urban forms based upon the postulates of Menard (1995) and Pohlmann & Dennis (2003). The latter states that OLS, just like logistic regression models, can model binary variables using linear probability models. Although OLS might give predicted values beyond the range (0,1), the analysis may still be useful for classification and hypothesis testing.

It is perhaps because the normal distribution and homogeneous error variance assumptions of OLS will probably be violated whenever a binary dependent variable is used, especially when the probability of the dependent event varies widely, no researcher has so far extended the use of OLS and GWR results to the simulation and prediction of urban forms. In this paper, however, we have taken a first, preliminary step along this path.

References

Abiodun, J.O. (1997) - The challenges of growth and development in metropolitan Lagos. In C. Rakodi (Ed.), *The urban challenge in Africa: Growth and management of its large cities,* Tokyo, United Nations University Press, 192-222.

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

Braimoh, A.K., & Onishi, T. (2007) - Spatial determinants of urban land use change in Lagos, Nigeria, *Land Use Policy, 24*(2), 502-515.

Erener, A., Sebnem, H., & Düzgün, B. (2010) - Improvement of statistical landslide susceptibility mapping by using spatial and global regression methods in the case of More and Romsdal (Norway), *Landslides, 7,* 55–68.

Farber, S. & Páez, A. (2007) - A systematic investigation of cross-validation in GWR model estimation: Empirical Analysis and Monte Carlo simulations, *Journal of Geographical Systems, 9*(4), 371-396.

Fotheringham, A.S., Brunsdon, C. & Charlton, M. (2002) - *Geographically weighted regression: The analysis of spatially varying relationships*, Chichester, John Wiley & Sons.

Gong, P. (1996) - Integrated analysis of spatial data from multiple sources: Using evidential reasoning and artificial neural network techniques for geological mapping, *Photogrammetric Engineering & Remote Sensing, 62*(5), 513-523.

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

Leung, Y., Mei, C., & Zhang, W. (2000) - Testing for spatial autocorrelation among the residuals of the geographically weighted regression, *Environment and Planning A, 32*(5), 871-890.

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* (2nd ed.), Upper Saddle River, 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.

Menard S. (1995) - *Applied logistic regression analysis,* Thousand Oaks, Sage.

Mennis, J. (2006) - Mapping the results of geographically weighted regression, *The Cartographic Journal, 43*(2), 171-179.

Noresah M.S. & Ruslan, R. (2009) - Modelling urban spatial structure using geographically weighted regression, presented to *18th World IMACS / MODSIM Congress*, Cairns, Australia.

Okwuashi, O. (2011) - *Application of Geographic Information Systems and Cellular Automata Based Models to Land Use Change Modelling of Lagos, Nigeria,* unpublished doctoral dissertation, Victoria University of Wellington, New Zealand.

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

Shariff, N.M., Gairola, S. & Talib, A. (2010) - Modelling Urban Land Use Change Using Geographically Weighted Regression and the Implications for Sustainable Environmental Planning, in Swayne, D.A. et al (eds.) - *Modelling for environment's sake : Proceedings of the Fifth Biennial Conference of the International Environmental Modelling and Software Society,* Ottawa, Canada.

Thapa, R. & Murayama, Y. (2009) - Land use change factors in Kathmandu Valley: A GWR approach, presented to *10th International Conference on Geocomputation*, UNSW, Sydney, 255-260.

Wheeler, D. & Tiefelsdorf, M. (2005) - Multicollinearity and correlation among local regression coefficients in geographically weighted regression, *Journal of Geographical Systems, 7,* 161–187.
