Available online at www.sciencedirect.com
ScienceDirect
Computers, Environment and Urban Systems
31 (2007) 667–688

Computers, Environment and Urban Systems
www.elsevier.com/locate/compenvurbsys

Modeling urban growth in Atlanta using logistic regression

Zhiyong Hu [a,*], C.P. Lo [b]

[a] *Department of Environmental Studies, University of West Florida, Pensacola, FL 32514, USA*
[b] *Department of Geography, University of Georgia, Athens, GA 30602, USA*

Received 29 June 2005; accepted in revised form 7 November 2006

Abstract

This study applied logistic regression to model urban growth in the Atlanta Metropolitan Area of Georgia in a GIS environment and to discover the relationship between urban growth and the driving forces. Historical land use/cover data of Atlanta were extracted from the 1987 and 1997 Landsat TM images. Multi-resolution calibration of a series of logistic regression models was conducted from 50 m to 300 m at intervals of 25 m. A fractal analysis pointed to 225 m as the optimal resolution of modeling. The following two groups of factors were found to affect urban growth in different degrees as indicated by odd ratios: (1) population density, distances to nearest urban clusters, activity centers and roads, and high/low density urban uses (all with odds ratios < 1); and (2) distance to the CBD, number of urban cells within a 7×7 cell window, bare land, crop/grass land, forest, and UTM northing coordinate (all with odds ratios > 1). A map of urban growth probability was calculated and used to predict future urban patterns. Relative operating characteristic (ROC) value of 0.85 indicates that the probability map is valid. It was concluded that despite logistic regression's lack of temporal dynamics, it was spatially explicit and suitable for multi-scale analysis, and most importantly, allowed much deeper understanding of the forces driving the growth and the formation of the urban spatial pattern.

Keywords: Urban growth; Logistic regression; GIS; Scale; Fractal

* Corresponding author. Tel.: +1 850 474 3494; fax: +1 850 857 6036.
E-mail address: zhu@uwf.edu (Z. Hu).

0198-9715/$ - see front matter
doi:10.1016/j.compenvurbsys.2006.11.001

1. Introduction

An urban land use system is dominated by human activities with complex spatio-temporal dynamics. The main issues of great importance in land use modeling include spatial dynamics, temporal dynamics, incorporation of human drivers of land use changes, and scale dynamics (Veldkamp & Lambin, 2001). Dynamic simulation models and empirical estimation models have been used to model land use changes. Rule-based simulation models, such as Cellular Automata (CA), are most suitable for incorporating spatial interaction effects and handling temporal dynamics. However, CA models focus on simulation of spatial pattern rather than on interpretation or understanding of spatio-temporal processes of urban growth. Most dynamic simulation models cannot incorporate enough socioeconomic variables.

Empirical estimation models use statistical techniques to model the relationships between land use changes and the drivers based on historic data. As an empirical estimation method, logistical regression has been used in deforestation analysis (Geoghegan et al., 2001; Schneider & Pontius, 2001), agriculture (Serneels & Lambin, 2001; Walsh, Crawford, Welsh, & Crews-Meyer, 2001), and urban growth modeling (Allen & Lu, 2003; Landis & Zhang, 1998; Wu & Yeh, 1997). Statistical approaches can readily identify the influence of independent variables and also provide a degree of confidence regarding their contribution. In many cases, these models fit spatial processes and land use change outcome reasonably well (Irwin & Geoghegan, 2001). Urban growth modeling aims to understand the dynamic processes, and therefore interpretability of models is becoming crucial. Interpretation of statistical models is desirable for gaining knowledge of the processes driving the change of spatial patterns. Calibration of logistic regression is not so computation intensive as CA simulation, thus better suited to deal with scale dynamics by conducting scaling-up (from finer to coarser resolutions) modeling for a region covering a large spatial extent. Existing logistical regression models of urban growth are based on a single scale and the resolution is determined by the level of spatial detail in data or limited by the computation power. Spatial autocorrelation, which causes violation of the assumption of independent residuals, is often ignored in those models because the statistical methodology for considering autocorrelation is not well developed for logistic regression models as it is for least squares regression models.

In this paper, an approach to urban growth modeling using logistic regression is explained. The logistic regression model was applied to study the urban growth in Atlanta, Georgia. The modeling aims to discover the relationship between urban growth and social, econometric and biophysical factors and to predict the future urban pattern. A dynamic CA model has been previously applied to simulate the urban growth of Atlanta (Yang & Lo, 2003). This will allow comparison between these two approaches of modeling applied to the same city. The steps of the modeling are to: (1) conduct multi-resolution calibration of a series of logistic regression models and find the optimal resolution of modeling using a fractal analysis; (2) refine the model at the optimal resolution by correcting for spatial autocorrelation; (3) use the refined model to explain the driving forces of the urban growth; (4) validate the model by Relative operating characteristic (ROC) statistics; and (5) predict the future urban pattern.

2. Study area

The Atlanta, Georgia metropolitan region is defined here to include thirteen urban counties with a spatial extent of about 120 km × 140 km (Fig. 1). The first 10 counties form

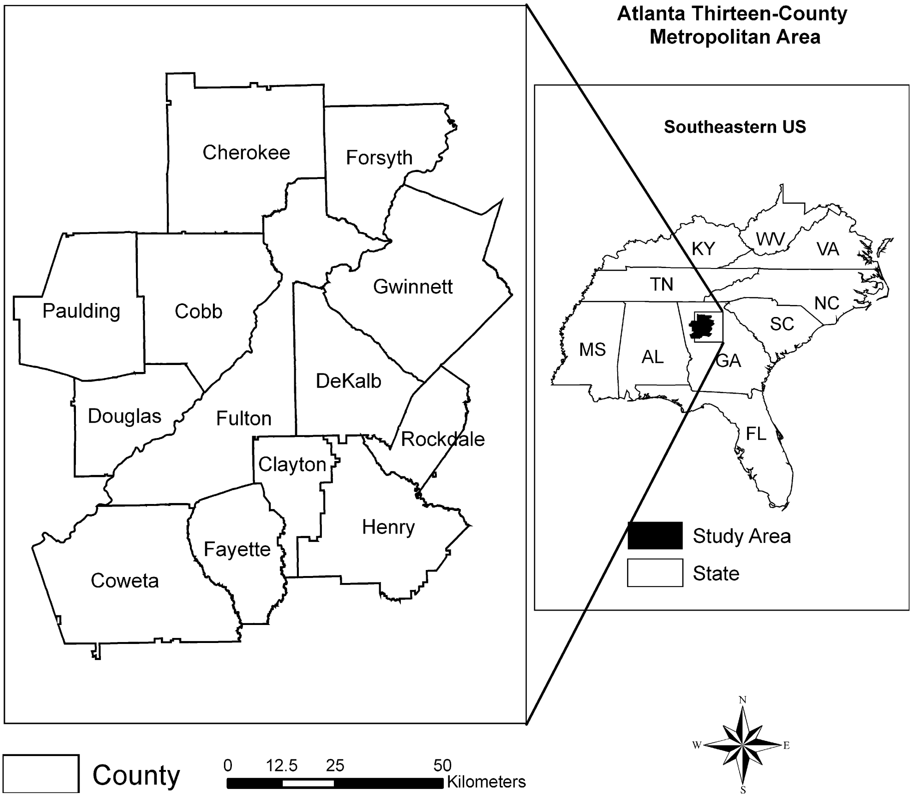

Fig. 1. Thirteen-county Atlanta metropolitan region.

the planning area of the Atlanta Regional Commission (ARC). In the last half of the 20th century, Atlanta, Georgia has risen to become the premier commercial, industrial, and transportation urban area of the southeastern United States and one of the fastest growing metropolitan areas in the Nation. Concomitant with the high rate of population growth has been an explosive growth of the urban extent. This has resulted in tremendous land cover changes within the metropolitan region, wherein urbanization has consumed vast acreages of forested and agricultural land adjacent to the city proper and has pushed the rural/urban fringe farther and farther away from the original Atlanta urban core.

There has been an unbalanced and polarizing growth in Atlanta: a dividing line exists between the north and the south, strongly corresponding with the long-standing residential racial segregation patterns. This unbalanced growth has many dimensions which are shaping factors of the urban patterns: *population*, *race*, *income*, *employment*, *housing*, and *transportation patterns* (BICUMP, 2000). Explosive *population* growth is occurring in the northern and outer suburbs of the region. *Race* segregation is also observed. The white tend to live in the north of the Interstate Highway 20, and in the far southern suburban communities. Central Fulton and Dekalb counties are home to over 70% of the region's non-white population. Higher *income* families tend to live in the region's northern and far southern areas. The poor tend to live in the central city and southern parts. Unbalanced growth of *employment* also has fueled the urban sprawl. Most new jobs

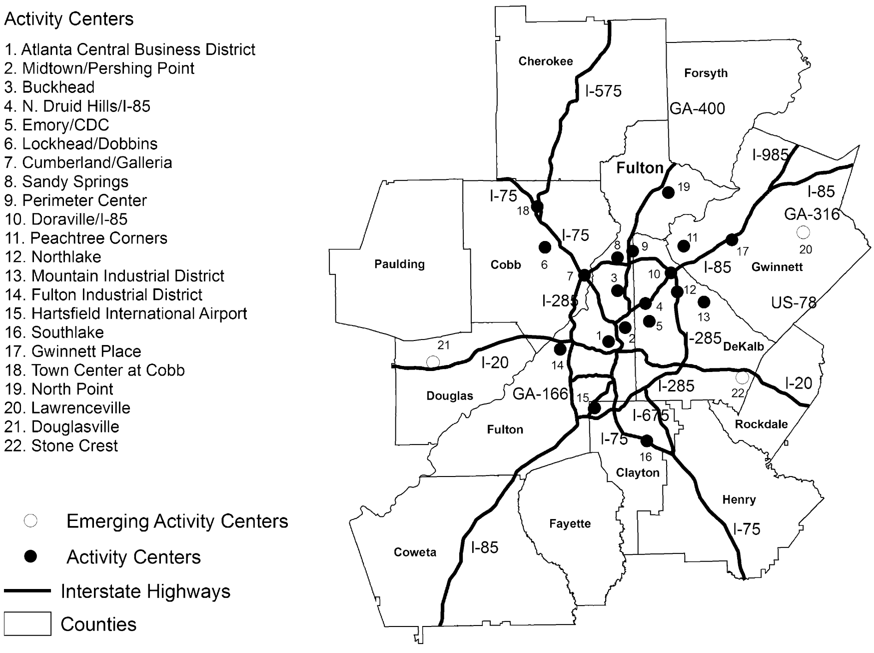

Fig. 2. Major activity Centers in Atlanta region in 1995 (Source: Atlanta Regional Commission).

and high-paying jobs are on the north side of the region (Atlanta Regional Commission, 1997). Many of the areas of greatest job increases are outside Atlanta's I-285 perimeter highway. The central city of Atlanta is slipping overall in its share of jobs. There is little or no job growth in the majority non-white neighborhoods. The spatial distribution of affordable *housing* is one of the important factors shaping metropolitan growth patterns. Many middle-class families cannot afford to live in the city of Atlanta's residential areas or in job-rich parts of the suburbs. Developments sprawl into the exurban fringe because many families cannot afford the near northside houses and avoid the southside (BICUMP, 2000). *Transportation* plays an important role in shaping urban development. The bulk of the Atlanta region's infrastructure funds have been spent on highways, particularly in the northern part of the region (Atlanta Regional Commission, 1997).

The massive suburbanization of economic activities since the 1960s contributed significantly to the spatial restructuring of the business landscape of the Metropolitan Atlanta, resulting in the clustering of high-order activities in new metropolitan-level urban centers – 'suburban downtowns' (Hartshorn & Muller, 1989). The emergence of these large multi-functional complexes in the outer suburban city has created a polycentric structure (Fig. 2). The decentralization continues with astonishing rapidity today.

3. Statement of model

A logistic regression model was used to associate the urban growth with demographic, econometric and biophysical driving forces and to generate an urban growth probability map. In a raster GIS modeling environment, the data layers are tessellated to form a grid of cells. The nature of the land use/cover change of a cell is dichotomous: either the presence of urban growth or absence of urban growth. If binary values 1 and 0 are used to represent urban growth and no urban growth respectively and if it is assumed that the probability of a cell changing to urban use follows the logistic curve as described by the logistic function (Kleinbaum, 1994):

$$f(z) = \frac{1}{1 + e^{-z}} \tag{1}$$

then the probability of a cell being urbanized can be estimated with the following logistic regression model:

$$P(Y = 1|X_1, X_2, \ldots, X_k) = \frac{1}{1 + e^{-\left(\alpha + \sum_{i=1}^{k} \beta_i X_i\right)}} \tag{2}$$

where $P(Y = 1|X_1, X_2, \ldots, X_k)$ is the probability of the dependent variable Y being 1 given $(X_1, X_2, \ldots, X_k)$, i.e. the probability of a cell being urbanized; X_i is an independent variable representing a driving force of urbanization, which can be of interval, ordinal or categorical nature; and β_i is the coefficient for variable X_i.

In this research, the land use/cover maps produced from Project ATLANTA for the years 1987 and 1997 are used (Yang & Lo, 2002), which show six categories of land use/cover: high-density urban, low-density urban, bare land, crop or grassland, forest, and water. Logistic regression modeling, as an empirical estimation approach, allows a data-driven rather than a knowledge-based approach to the choice of predictor variables. Nevertheless, we still made an informed selection of variables. Selection of social predictor variables was guided by a historical review of urban growth in Atlanta as reviewed in Section 2. The social variables correspond to the five dimensions shaping Atlanta urban patterns (*population*, *race*, *income*, *employment*, *housing*). Population density is often established as land use determinants to indicate labor availability, accessibility, or presence of local markets (Agarwal, Green, Grove, Evans, & Schweik, 2001; Allen & Lu, 2003). Correlations may exist between those demographic variables. Logistic regression calibration should check for multi-collinearity. Model calibration in this study had two stages including initial calibration and refining. Multicollinearity test was not examined in the initial stage in view of the large sample size and the possibility that part of the variables might be insignificant and excluded from the model in the refining stage. The choice of econometric and biophysical variables conforms to most dynamic simulation modeling practices, which usually consider the determining factors of ‘SLEUTH’ (slope, land use, exclusion, urban extent, transportation, hillshade) as in Clarke’s SLEUTH model (Clarke, Hoppen, & Gaydos, 1997; Dietzel & Clarke, 2006; Yang & Lo, 2003). These variables reflect the biophysical conditions, the spatial influences of major highways, economic activity centers, existing land use status, and institutional factors, such as land conservation. The 1990 census data were used for the social variables in model calibration. The 2000 census data were used for model prediction. The model should perform best if predictor data are collected at the year 1992, which lies halfway through the time period considered (1987–1997). There is a time lag of only two years between the calibration data collection year and the halfway year, the influence of which on model results should be minor. A 1995 map of major economic centers was used to calculate the distance to the centers for model calibration and a 2001 map for prediction (Atlanta Regional Commission, 1995 & 1997). A 2001 National Land Cover Data (NLCD) map was used for validation. An interaction term number of urban cells within a neighborhood was calculated as an independent variable to take spatial interaction effects into account.

The complete list of variables is shown in Table 1. Fig. 3 shows the map of urban growth from 1987 to 1997, which serves as the dependent variable Y. Although land use maps for both high-density and low-density urban were available for this study, the variable Y was defined based on the combination of the two urban types. This makes this study comparable with previous CA modeling. Most existing urban growth models do not differentiate between high-density and low-density urban growth. Fig. 4 shows the raster maps of the independent variables. Five design variables denoted as X_{14} through X_{18} representing five land use/cover classes respectively were generated to distinguish among the six categories of land use/cover by recoding the 1987 land use/cover map into a binary map for each land use/cover category. If all the five design variables take the value of zero, then a cell value in the "water" layer must be one; if any one of the five land use/cover classes takes the value of 1, a cell value in the "water" layer must be zero. Including a "water" variable in the model would be redundant and cause multi-collinearity. Initial model calibration used only the first 18 variables. The last two variables were incorporated into the model in the model refining stage to correct for spatial autocorrelation that might exist.

4. Multi-scale modeling and fractal analysis

Techniques of GIS have provided the potential to generate multi-resolution data sets for scale up modeling. The simple and uniform geometry of raster data is convenient

Table 1
List of variables included in the logistic regression model

Variable	Meaning	Nature of variable
Dependent		
Y	0 – no urban growth; 1 – urban growth	Dichotomous
Independent		
X_1	Population density (1000 person/km^2)	Continuous
X_2	Per capital income ($)	Continuous
X_3	Poverty rate	Continuous
X_4	Median housing rent ($)	Continuous
X_5	Percentage of white people	Continuous
X_6	Employment rate	Continuous
X_7	Slope (%)	Continuous
X_8	Distance to the nearest urban cluster (km)	Continuous
X_9	Distance to CBD (km)	Continuous
X_{10}	Distance to active economy centers (km)	Continuous
X_{11}	Distance to the nearest major road (km)	Continuous
X_{12}	Number of urban cells within a 7×7 cell window	Continuous
X_{13}	1 – Conservation area; 0 – not conservation area	Design
X_{14}	1 – High-density urban; 0 – not high-density urban	Design
X_{15}	1 – Low-density urban; 0 – not low-density urban	Design
X_{16}	1 – Bare land; 0 – not bare land	Design
X_{17}	1 – Cropland/grassland; 0 – not cropland/grassland	Design
X_{18}	1 – Forest; 0 – not forest	Design
E[a]	Easting coordinate (m)	Continuous
N[a]	Northing coordinate (m)	Continuous

[a] E and N are used to correct for spatial autocorrelation.

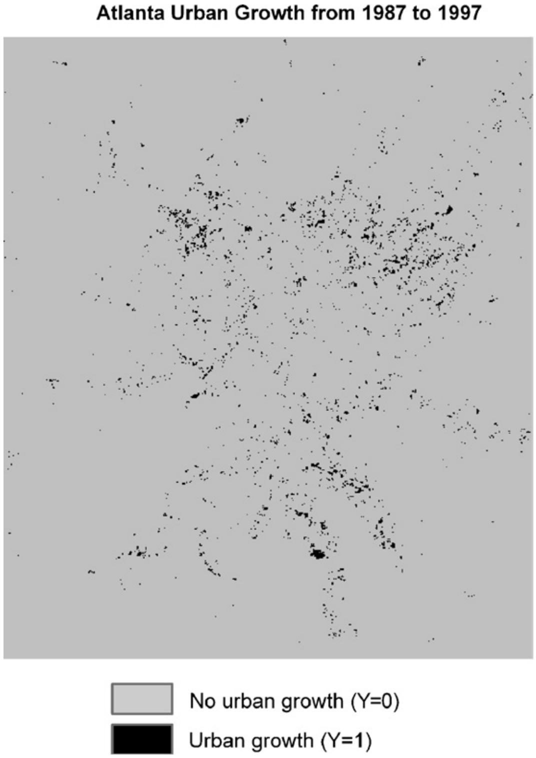

Fig. 3. Dependent variable Y – urban growth from 1987 to 1997.

for aggregation. A series of logistic regression models with the 18 independent variables (X_1–X_{18}) was calibrated in the support of Idrisi Kilimanjaro GIS software using full data points within the mask of the thirteen counties.

Model calibration was initially tried at a resolution of 25 m, but Idrisi Kilimanjaro running on a DELL desktop computer failed to accomplish the model calibration due to the intensive computation – noting that there are 16,699,756 data points for each layer and 19 layers in total. The raster layers at the resolution of 25 m were then aggregated to generate 50 m, 75 m, and so on up to 300 m data sets to accommodate the modeling at coarser resolutions.

Many existing scale studies rely heavily on aggregation methods to generate multi-scale raster data for analysis. GIS raster data aggregation generalizes an image by reducing the number of rows and columns while simultaneously decreasing the cell resolution. Methods for aggregating regular grids include the averaging method, sampling every *n*th cell, and dominant values (Bian, 1997). The averaging method applies to continuous ratio and interval data. Methods of sampling every *n*th cell and dominant values apply to nominal data. In this study, for the dependent variable urban growth (Y), the conservation area (X_{13}), and design variables for land use/cover types (X_{14}–X_{18}), the method of selecting dominant values was used to generate data with coarser resolutions. The DEMs of coarser resolutions were generated using the averaging method. Multi-resolution slope (X_7) data were created from DEM data of corresponding resolutions. Raster layers for the number

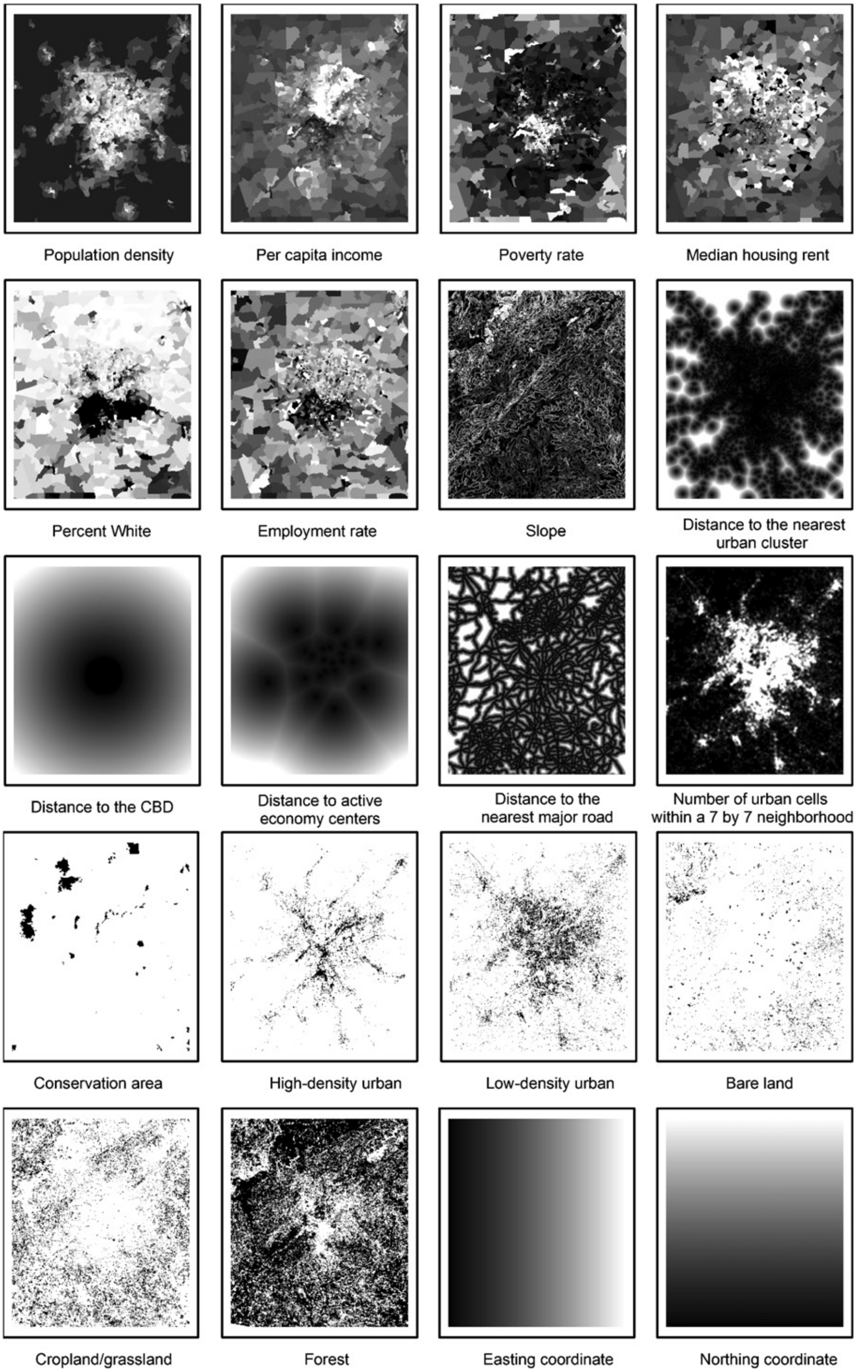

Fig. 4. Raster layers of independent variables. For ratio data layers, the whiter the tone, the larger quantity; for nominal data layers, black represents 1 and white represents 0.

of urban cells within a 7×7 cell neighborhood (X_{12}) were calculated from urban distribution maps. The selection of the size of the neighborhood window conforms to most practices in dynamic simulation models where sizes are often 3×3, 5×5, or 7×7. Existing scale studies show that variations of results from multi-scale analysis are not completely due to the 'real' scale effects, but rather they are artifacts attributable to the use of different resampling methods (Weigel, 1996). To minimize and account for the effects of data aggregation on modeling, an explicit aggregation operation was not applied to raster layers of demographic data (X_1–X_6), and distance variables (X_8–X_{11}), rather multi-resolution data for those variables were directly generated at each resolution.

The purpose of multi-resolution calibration of the logistic regression in this study was to find the optimal resolution for modeling. Moellering and Tobler (1972) argue that geographic processes operate at different scales and that one can determine the resolution level at which most processes operate. There are means to forecast at what resolutions new patterns may emerge and when the performance of models takes a significant turn. These turning points should be those at which the resolutions approach dominant operational scales. These resolutions are where modeling should be conducted (Meentemeyer, 1989; Moellering & Tobler, 1972). Previous studies in environmental modeling using simple linear regression found that *R*-square values are higher at coarser scales (Bian & Walsh, 1993; Kok & Veldcamp, 2001). Goodness-of-fit values might not be used to determine the optimal resolution due to a lack of a turning point. And furthermore, logistic regression does not have an equivalent to the *R*-square that is found in ordinary least square (OLS) regression. Although some come up with pseudo *R*-square statistics, this statistic does not mean what *R*-square means in OLS regression (the proportion of variance explained by the predictors). Therefore, pseudo *R*-square values are not suitable for determining the best resolution of modeling.

This study used fractal analysis to determine the optimal scale of modeling. Fractal dimensions were calculated for the probability surface maps predicted using the logistic regression model calibrated from resolutions of 50 m to 300 m. The triangular prism surface area method (Clarke, 1986; Jaggie, Quattrochi, & Lam, 1993) was used to calculate the fractal dimensions. This method estimates lumped fractal dimension values from the predicted probability surface. The fractal dimensions were calculated using the software package Image Characterization and Modeling System (ICAMS) (Quattrochi, Lam, Qiu, & Zhao, 1997).

Fig. 5 shows the change of fractional dimension with the resolution of modeling. Fractal dimension increases almost linearly with the change of resolution from 50 m to 225 m, then decreases at the turning point of 225 m. This suggests that the urbanization probability surface does not demonstrate the property of self-similarity of real fractals since self-similar objects must have constant fractional dimension. Previous studies have demonstrated that true fractals with self-similarity at all scales are uncommon (Lam & Quattrochi, 1992) and most real-world curves and surfaces are not pure fractals with a constant fractal dimension at all scales. The change of fractal dimension across scale, though controversial to the strict sense of fractal dimension as defined by Mandelbrot (1983), can be interpreted positively and used to summarize the scale changes of the spatial phenomena. The scale at which the highest fractal dimension is measured may be the scale at which most of the processes operate (Cao & Lam, 1996; Goodchild & Mark, 1987; Lam & Quattrochi, 1992) and the model performs best. To test if the model in deed performs best at the turning point of 225 m, a series of urban growth probability maps generated

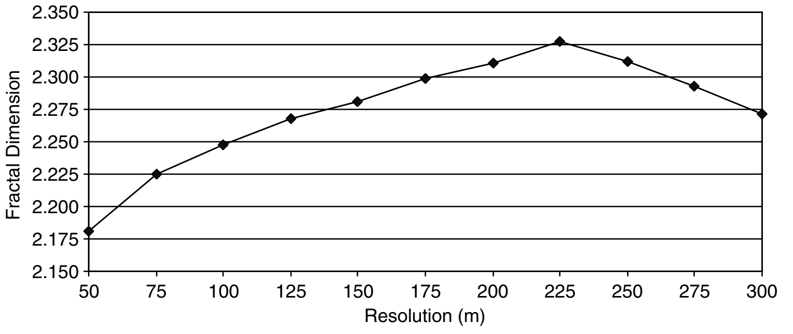

Fig. 5. Lumped fractal dimension of logistic regression model predicted probability surface plotted against resolution.

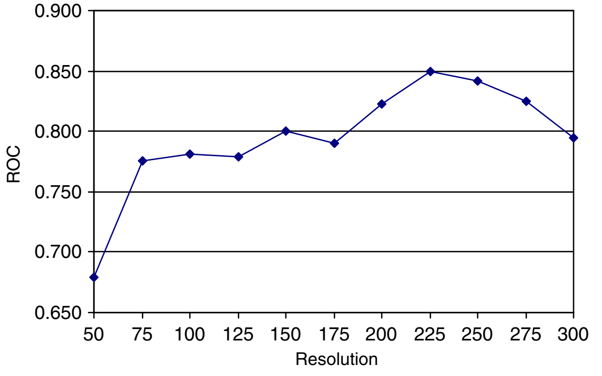

Fig. 6. Relative operating characteristic (ROC) against resolution.

from the logistic regression was compared against an actual urban growth map and ROC values, which validate the model performance, were calculated (for details of model validation, see Section 8). Fig. 6 shows the change of ROC statistics with resolution. The highest ROC value was achieved at the resolution of 225 m. Thus the resolution of 225 m was selected as the optimal scale at which the logistic model best represents the dynamics of urbanization and the underlying processes. The optimal resolution of 225 m should be a compromised resolution and avoids both individualistic fallacy and ecological fallacy.

5. Refining the model at the resolution of 225 m

The initial calibration of logistic regression models at various resolutions above used the full data set comprising all cell values within the 13 county study area. At each resolution, the model calibration resulted in a predicted urbanization probability surface map and a residual map indicating the difference between the predicted and the observed probability. The logistic regression model assumes that observations are independent of each other and the residuals are mutually independent. But this assumption may be violated due to the spatial autocorrelation. Spatial autocorrelation is the propensity for data values to be similar to neighboring data values.

To test the logistic regression residual for spatial autocorrelation, Moran's I for the King's case was calculated under a normality assumption that the cell values represent independent drawings from a single normally distributed population, hence a null hypothesis that there is no spatial autocorrelation. For the 13-county area within the residual image at the resolution of 225 m, the value of Moran's I is 0.283, indicating positive spatial autocorrelation. The Z-test statistic value is 252.29 with the p value of $2.22e-6$. The p value is much less than 5%, which leads to the conclusion that the null hypothesis of no spatial autocorrelation in the residuals can be rejected. In other words, spatial autocorrelation is present among the residual values.

Model fitting at the optimal resolution of 225 m was refined by correcting for spatial autocorrelation. Logistic regression models belong to the family of generalized linear models. Spatial forms of such models are not well developed. This study used three steps to correct for the effects of space.

The first step was applying raster GIS data aggregation and pixel thinning functions on the data layers. This was done when multi-resolution data sets were created for multi-scale modeling. The multi-resolution modeling process from 50 m cell size to 300 m cell size is a process of alleviating the spatial effect by considering a series of spatial lags from the first order of 50 m to the 11th order of 300 m. The effect of spatial dependence at the resolution of 225 m must be weaker than that at 50 m since the attribute similarity becomes weaker as spatial lags progress from the first order to a higher order.

The second step was including spatial coordinates of data points into the list of independent variables. Spatial autocorrelation can be alleviated to some extent by attempting to introduce location into the link function to remove any such effects present (Bailey & Gatrell, 1995). For example, spatial coordinates of observations might be introduced as additional covariates, or to classify regions in terms of their broad location and treat this classification as an extra categorical explanatory factor in the model. This assumes of course that one can "explain away" spatial dependence in terms of a first-order spatial trend, i.e., the first-lag autocorrelation.

The last strategy was sampling. A stratified random sample image was generated and used as the feature definition file to extract cell values of dependent and independent variables on which the refined logistic regression model was fitted. The spatial distances between sampling data points are larger than those between neighboring data points in the full data set, thus the spatial autocorrelation effects on modeling would be smaller than those by using the full data set. Since the urban growth map serves as the dependent variable in the logistic regression model, the small amount of area of urban growth also tends to be under-sampled if only a portion of the data points are sampled for modeling. Attention must be paid to selection of an appropriate sampling method. Stratified random sampling was applied to the area covered by the rectangle bounding the 13-county study area to generate a vector point file in a GIS environment. Stratified random sampling is thought to perform well when it is necessary to make sure that small, but important, areas are represented in the sample (Congalton, 1988). Since logistic regression model fitting would be performed within the 13 counties, only those points within the counties were extracted using point-in-polygon GIS operation. At the resolution of 225 m, the number of cells within the counties is 206,316, of which 20,389 cells have been sampled. Within the counties, the number of cells that have changed from non-urban to urban, i.e., the number of 1s for variable Y (1 = urban growth), is 20,631, accounting for 1.85% of the total number of cells. Of the 20,389 sample points, there are 370 points whose cell values are 1 in the Y variable layer. The percentage of 1s in the sample is 1.82%, matching very well with the percentage of 1.85% for the full data set, which demonstrates the representativeness of the stratified random sampling.

A maximum likelihood estimator (Hosmer & Lemeshow, 1989) was used to fit the model. The results of fitting the logistic regression model with the full 20 independent variables (M_{20}) are given in Table 2. At the $\alpha = 0.05$ level, population density (X_1), distance to the nearest urban cluster (X_8), distance to CBD (X_9), distance to active economy centers (X_{10}), distance to major roads (X_{11}), number of urban cells within a neighborhood defined by a window of 7×7 cells (X_{12}), design variables high-density urban area (X_{14}) and low-density urban area (X_{15}) are significant. At the $\alpha = 0.10$ level, besides the above variables, the variable UTM coordinate N is also significant. A probability map was derived using the refined model and a residual map calculated to evaluate the extent to which autocorrelation has been reduced. The value of Moran's I becomes 0.006 ($p = 0.074$), indicating very weak spatial autocorrelation. McFadden's pseudo R-square (McFadden, 1973) was used to test the goodness-of-fit of the model. Pseudo R square values between 0.2 and 0.4 are considered a good fit (Clark & Hosking, 1986; Domencich & McFadden, 1975). The pseudo R^2 value of the full model M_{20} is 0.147, indicating a weak fit.

Following the significance test, it is logical to construct a reduced model which excludes those variables thought to be insignificant. Of the five design variables for land use/cover, only two (X_{14} and X_{15}) are significant. The other three (X_{16}, X_{17} and X_{18}) are insignificant. Thus confusion arises since we are not sure about the contribution of land use/cover as a single variable to the model when only a part of the design variables are significant. Statisticians suggest that we must be careful in our use of the Wald statistics to assess the significance of the coefficients and that whenever a categorically scaled independent variable is included (or excluded) from a model, all of its design variables should be included (or excluded) (Hosmer & Lemeshow, 1989). Strict adherence to the $\alpha = 0.10$ level of significance would justify excluding the three land use/cover types from the model. However, the probability of urbanization of a land lot should be influenced by its initial land use/cover status and initial land use/cover should be considered important in land use/cover change dynamics in a biophysical and cultural sense. Thus all the five design variables for land use/cover were kept in the reduced model. The results of fitting the reduced logistic regression model (M_{12}) is shown in Table 3. The pseudo R^2 value of 0.278 indicates a good fit of the model.

6. Model interpretation

Urban development tends to occur in an area of lower population density (X_1). The estimated odds ratio is 0.570068, or 1/1.754177, which is less than one, indicating that the probability of urban growth in an area of higher population density is less than the probability of urban growth in an area of lower population density. Specifically, the odds of urban development would decrease by 0.754177 if population density increases by 1000 person/km^2. Like most other American cities, urban sprawl and suburbanization in the Atlanta metropolitan region are characteristic of low-density urban development, which replaces farmland, forest and open space with single-family homes on large lots.

Table 2
Estimated coefficients and odds ratios for the logistic regression model containing the 20 independent variables (M_{20})

Variable	Coefficient	Standard error	Odds ratio	Z	$P > \|Z\|$
X_1	−0.611000	0.215141	0.542808	−2.84	**0.005*
X_2	−0.000013	8.87E−06	0.999987	−1.49	0.135
X_3	0.003925	0.008132	1.003933	0.48	0.629
X_4	0.000189	0.000263	1.000188	0.72	0.473
X_5	−0.000340	0.002980	0.999656	−0.12	0.908
X_6	0.010102	0.015917	1.010153	0.63	0.526
X_7	−0.017820	0.031064	0.982337	−0.57	0.566
X_8	−0.960000	0.129032	0.382893	−7.44	**0.000*
X_9	0.017700	0.008939	1.017858	1.98	**0.047*
X_{10}	−0.083000	0.011277	0.920351	−7.36	**0.000*
X_{11}	−0.730000	0.119672	0.481909	−6.10	**0.000*
X_{12}	0.017299	0.007073	1.017450	2.45	**0.014*
X_{13}	−0.183890	0.522982	0.832031	−0.35	0.725
X_{14}	−2.529920	0.812327	0.079665	−3.11	**0.002*
X_{15}	−1.513310	0.747177	0.220180	−2.03	**0.043*
X_{16}	0.922617	0.768840	2.515866	1.20	0.230
X_{17}	0.731231	0.733594	2.077636	1.00	0.319
X_{18}	0.374261	0.722887	1.453916	0.52	0.605
E	2.58E−07	2.65E−06	1.000000	0.10	0.922
N	4.99E−06	2.72E−06	1.000005	1.84	**0.095*
Constant	13.52805	10.68137	N/A	1.27	0.205

*These variables are significant at $\alpha = 0.10$ level.

Y	Urban growth
*X_1	Population density (1000 person/km^2)
X_2	Per capital income ($)
X_3	Poverty rate (%)
X_4	Median housing rent ($)
X_5	Percentage of white people (%)
X_6	Employment rate (%)
X_7	Slope (%)
*X_8	Distance to the nearest urban cluster (km)
*X_9	Distance to CBD (km)
*X_{10}	Distance to active economy centers (km)
*X_{11}	Distance to the nearest major road (km)
*X_{12}	Number of urban cells within a 7×7 cell window
X_{13}	Conservation area
*X_{14}	High-density urban
*X_{15}	Low-density urban
X_{16}	Bare land
X_{17}	Cropland/grassland
X_{18}	Forest
E	Easing coordinate (m)
*N	Northing coordinate (m).

This may not be the case in developing countries where in the process of rapid urbanization new commercial and industrial facilities and residential subdivisions housing middle-class people often replace slums and villages populated with a large number of the poor and lower-class workers living in peripheral areas, old towns, or city centers.

Table 3
Estimated coefficients and odds ratios for the logistic regression model containing the 12 independent variables (M_{12})

Variable	Coefficient	Standard error	Odds ratio	Z	$P > \|Z\|$
X_1	−0.562004	0.211278	0.570068	−2.66	0.008
X_8	−0.963000	0.128916	0.381746	−7.47	0.000
X_9	0.020996	0.006731	1.021222	3.12	0.002
X_{10}	−0.084002	0.010461	0.919431	−8.03	0.000
X_{11}	−0.732010	0.118831	0.480946	−6.16	0.000
X_{12}	0.018747	0.006998	1.018924	2.68	0.007
X_{14}	−2.044372	0.809861	0.086837	−3.02	0.003
X_{15}	−1.468988	0.745736	0.230158	−1.97	0.049
X_{16}	0.959266	0.767965	2.609780	1.25	0.212
X_{17}	0.790909	0.732248	2.205400	1.08	0.280
X_{18}	0.421508	0.721633	1.524258	0.58	0.559
N	0.000057	0.000030	1.000057	1.90	0.027
Constant	18.804220	9.677099	N/A	1.94	0.052

Y	Urban growth
X_1	Population density (1000 person/km^2)
X_8	Distance to the nearest urban cluster (km)
X_9	Distance to CBD (km)
X_{10}	Distance to active economy centers (km)
X_{11}	Distance to the nearest major road (km)
X_{12}	Number of urban cells within a 7×7 cell window
X_{14}	High-density urban
X_{15}	Low-density urban
X_{16}	Bare land
X_{17}	Cropland/grassland
X_{18}	Forest
N	Northing coordinate (m).

Urban areas tend to grow close to the nearest urban cluster. Distance to the nearest urban cluster (X_8) has a coefficient of −0.963. The odds ratio is equal to 0.381746, or 1/2.619543. The probability of urban development in an area is estimated as 2.619543 times as large as the probability of urban development in an area 1 km further away from the nearest urban area. This demonstrates that pulling force has taken effect in the scale economy where commercial facilities tend to cluster together in a localized area.

The decentralized, polycentric suburbanizing trend in the metropolitan Atlanta area is evidenced by the interpretation of the odds ratios for the two predictors: distance to the CBD (X_9) and distance to active economy centers (X_{10}). The odds of urban development in an area 1 km further away from the CBD is estimated as 1.021222 as large as that in area closer to the CBD. The odds ratio for distance to active economic centers is 0.919431, or 1/1.087629, which means that the odds of urban development in area close to active economy centers is estimated as 1.087629 times as large as that in area 1 km further away from active economic centers. The closer it is to major activity centers, rather than to the CBD, the more probable a land lot will be developed for urban use.

The model also demonstrates that urban development has been controlled by road accessibility. The odds ratio for distance to major roads (X_{11}) is 0.480946, or 1/2.079235. The odds of urban development in an area closer to major roads is estimated as 2.079235 times as large as the odds of urban development in an area 1 km further away from major roads. The road influence contributes to the spatial patterns of ribbon and strip development.

A land lot with more neighboring areas that are urban is more likely to be developed for urban use. The variable number of urban cells within a neighborhood of 7×7 cell size (X_{12}) has an odds ratio equal to 1.018924. With an increase of 1 urban cell within the neighborhood, the odds of development will increase 0.018924. The use of a land lot is often influenced by the land use/cover status of the adjacent area. Land managers and real estate developers have some propensity of imitating the land use/cover behaviors in the neighborhood.

Of the five land use/cover types, only high-density urban (X_{14}) and low-density urban (X_{15}) areas have negative coefficients, resulting in odds ratio of less than 1. The odds of urban development in the existing urban area is estimated only as 0.08684 times for high-density urban use and 0.23016 times for low-density urban use respectively as large as the odds of urban development in non-developed area. Certainly the cost of redeveloping commercial and industrial areas is much higher than that in redeveloping residential areas. New urban development has occurred mainly in undeveloped peripheral urban–rural fringe areas or open space within established urban areas (infill development). It should be noted that the dependent variable Y has binary nominal values with 1 representing the change from non-urban to urban (urban growth) and 0 no such change. If X_{14} and X_{15} have values of 1, logically speaking, there will be no urban growth ($Y = 0$), thus the odds of urban development on existing urban area should be zero. However, in a logistical regression, the odds of Y being 1 is calculated using the equation $\hat{\psi} = e^{\hat{\alpha}+\sum_{i=1}^{k}\hat{\beta}_i X_i}$, and the odds ratio for the dichotomous variable X_i is calculated based on the equation $\frac{\psi_1}{\psi_0} = e^{\beta_i}$. The parameter α can be interpreted as the logarithm of the background odds that would result for the logistic model without any X's at all. The odds and odds ratio are never equal to zero.

The probability of urban development in bare land (X_{16}) is larger than the probability of urban development in areas covered with cropland or grassland (X_{17}). The probability of transition from cropland or grassland to urban use is larger than that of deforestation (X_{18}) for urban use. This can be seen from the odds ratio values of 2.61, 2.21 and 1.52 in a decreasing order for bare land, cropland or grassland, and forest, respectively. All values are greater than one, indicating a higher probability of urban development in those areas. It should be noted that in the study area much bare land is forest clear-cut area. So it can be said that urban development has taken place mainly at the expense of the depletion of green space.

It is interesting to notice that the UTM coordinate northing (N) is a significant predictor and has an odds ratio value slightly greater than 1 whereas the UTM coordinate easting (E) is not significant and has an odds ratio value of 1. The variables N and E were originally intended to correct for spatial autocorrelation. The interpretation of the odds ratio for N has shown that it not only acts as a spatial autocorrelation corrector, but also indicates an unbalanced growth along the north–south direction, since its odds ratio is a little higher than 1, which means urban growth tends to occur in the northern part of the region (with higher N coordinate values). The result conforms to the conclusion in the study report by BICUMP (2000) as reviewed in Section 2.

7. Prediction of urbanization probability

The probability of urbanization was predicted by plugging the coefficients of the logistic regression model containing the 12 significant predictors (M_{12}) into Eq. (2). To take

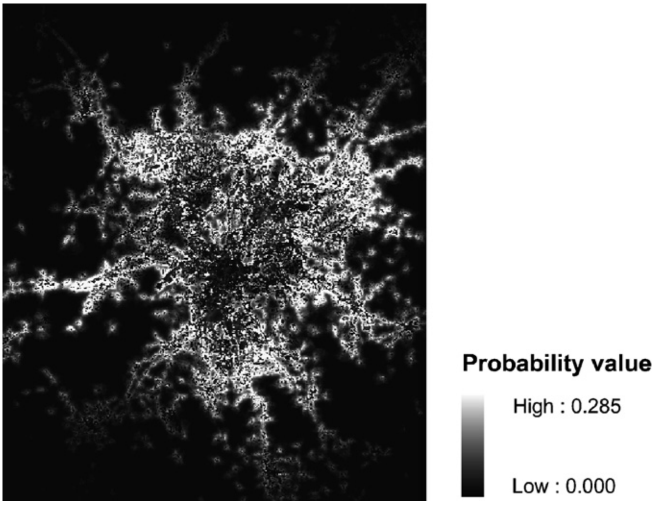

Fig. 7. Urbanization probability maps of Atlanta, Georgia. Lighter tones indicate higher probabilities of urban growth.

temporal dynamics into as much consideration as possible, raster layers were updated with newer cell values. The population density (X_1) surface was regenerated using the 2000 census data. Distance to active economy centers (X_{10}) was derived based on the active economy centers in 2001 (Atlanta Regional Commission, 2001). Distance to the nearest urban cluster (X_8) and number of urban cells within a 7×7 cell window (X_{12}) were calculated based on the 2001 NLCD data. Design variables high-density urban (X_{14}), low-density urban (X_{15}), bare land (X_{16}), cropland/grassland (X_{17}), and forest (X_{18}) were also extracted from the 2001 NLCD land use map. The map of predicted probability of urbanization is shown in Fig. 7, which is a 10-class quantile classification of the predicted probability values. The lighter tones indicate higher probabilities of urban growth. The future urban distribution pattern is easily discernable from this map. Some new emerging clusters far from existing urban areas can be seen. Most probable areas for urban development are closer to major highways and existing urban clusters.

8. Model validation using ROC method

Relative operating characteristic (ROC) was used to validate the logistic regression model. Recently the ROC method was brought to the field of land use/cover change modeling to measure the relationship between simulated change and real change (Pontius, 2000; Schneider & Pontius, 2001). ROC method is an excellent method to evaluate the validity of a model that predicts the occurrence of an event by comparing a probability image depicting the probability of that event occurring and a binary image showing where that class actually exists. In this study, the ROC method offers a statistical analysis that answers one important question: “How well is urban growth concentrated at the locations of relatively high suitability for urban growth?” Basically, ROC assesses how well the pair of maps agrees in terms of the location of cells being urbanized. Model validation using ROC reported a summary ROC value, a ROC curve as well as the coordinates of the points on the curve that were used to calculate the ROC value. A ROC value of 1 indicates that there is a perfect spatial agreement between the actual urban growth map and the predicted probability map. A ROC value of 0.5 is the agreement that would be expected due to chance, i.e., the cells values on the predicted probability image were assigned to random locations.

To conduct model validation, the image map of urban growth probability predicted from the logistic regression model was compared against that of actual urban growth (reference image) obtained by comparison of the 1987 land cover map with the NLCD 2001 land cover map. First the ranked image of probability of urbanization was sliced at a series of threshold levels. A threshold refers to the percentage of cells in the probability image to be reclassed as 1 in preparation for comparison with the reference image. The series of thresholds was specified at an equal interval of 5%. The threshold values are cumulative, therefore setting the equal interval thresholds 5, 10, 15, ..., 95 would yield 20 threshold intervals 0–5%, 0–10%, 0–15%, ... and 95–100%. ROC began with the cell ranked the highest for probability, reclassified it as 1 and continued down through the ranked cells until 5% of the cells had been reclassified as 1. The remaining 95% was classified as 0. This slice image was then compared with the reference image. Then ROC continued for the successive threshold. For each slice generated from each threshold, a two-by-two contingency table was created based on the comparison of the slice image with the reference image (Table 4). In the table, A represents the number of true positive cells which are predicted as urban growth and are actually urban growth in the reference image. B is the number of false positive cells. C is the number of false negative cells. D is the number of true negative cells. From each contingency table for each threshold, one data point (x, y) was generated where x is the rate of false positives (false positive %) and y is the rate of true positives (true positive %):

Table 4
Contingency table showing the comparison of the slice image of predicted urban growth probability with the reference image

		Reference image	
		Urban growth (1)	No urban growth (0)
Slice image of predicted probability	Urban growth (1)	A (true positive)	B (false positive)
	No urban growth (0)	C (false negative)	D (true negative)

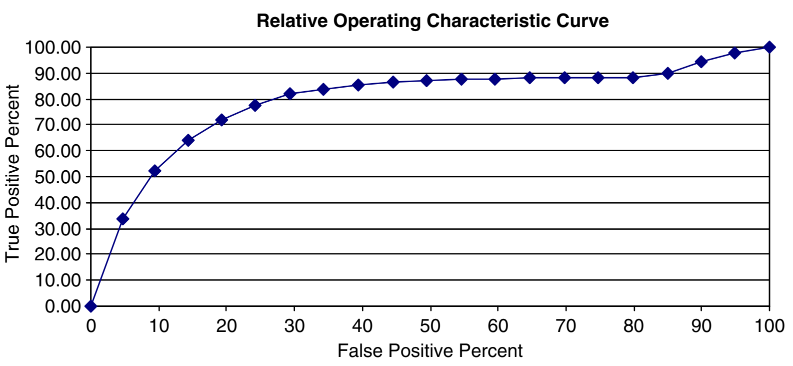

Fig. 8. ROC curve.

$$\text{true positive } \% = \frac{\mathrm{A}}{\mathrm{A}+\mathrm{C}} \tag{3}$$

$$\text{false positive } \% = \frac{\mathrm{B}}{\mathrm{B}+\mathrm{D}} \tag{4}$$

These data points were connected to create a ROC curve from which the ROC value was calculated. The ROC statistic is the area under the curve that connects the plotted points. The ROC curve is shown in Fig. 8. The ROC value is 0.85.

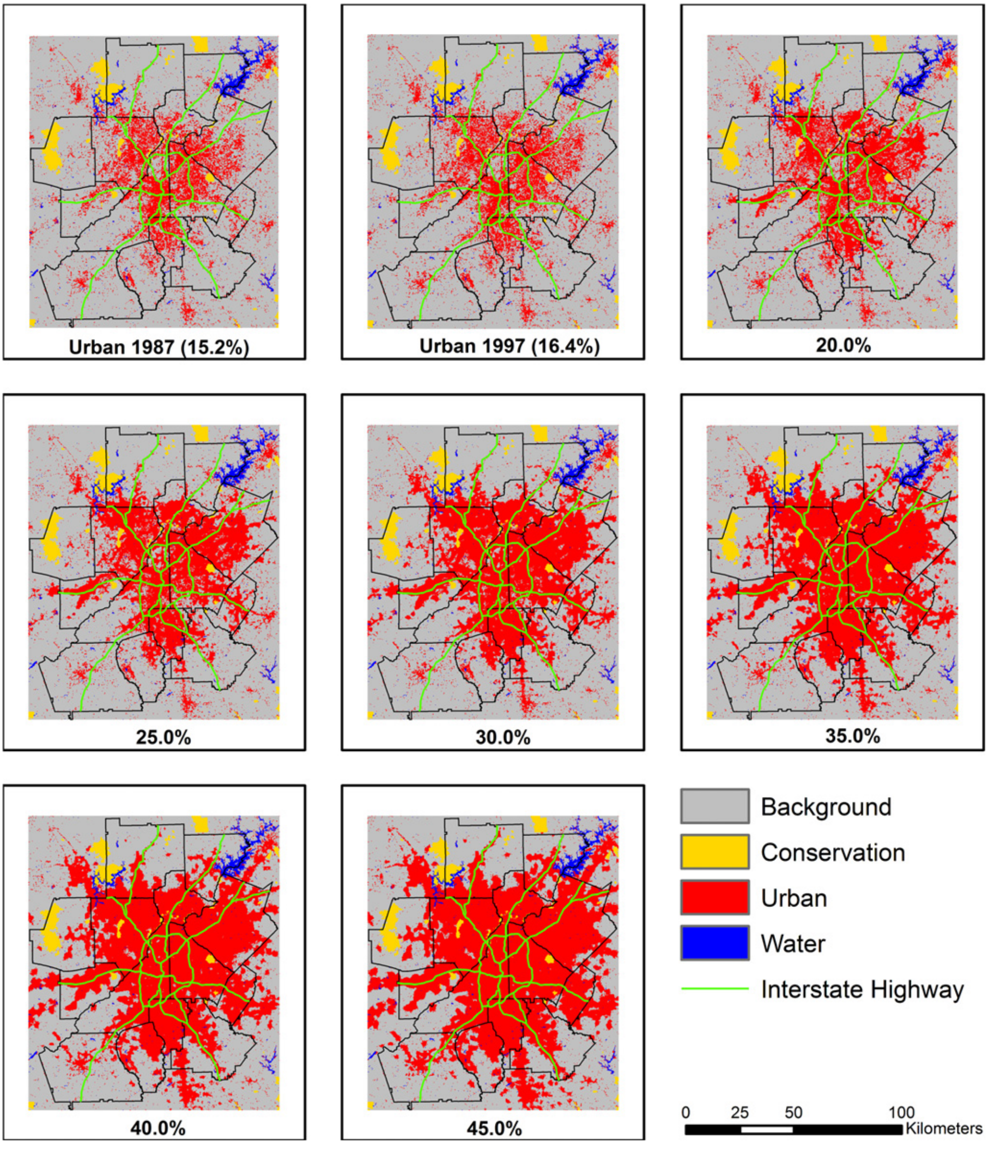

Fig. 9. Historical urban patterns and predicted urban patterns given percentage of urban area.

9. Prediction of spatial patterns of urban distribution

The probability map can be used for producing maps of urban distribution if any quantitative data on the future total areas of urban distribution, for example, urban planning data, are given. Based on the probability map, we can answer the question: "Where would urban growth occur if we know the amount of growth?" In the study area, urban area accounted for 15.2% in 1987 and 16.4% in 1997. What would the urban distribution patterns look like if urban area increases to 20%, 25%, 30%, 35% and 40%? To produce the spatial pattern of urban distribution given a certain amount of urban area, the increase of the number of urban cells compared to the 1987 base urban map was calculated. Then the number of urbanized cells was allocated to the probability map in the order of high probability value to low probability value. This generated a growth map. Next the growth map was combined with the 1987 base map to produce the urban distribution map. Fig. 9 shows a series of maps with increasing proportions of urban area. The series of maps clearly demonstrates the future trend of urban growth. It can be seen that urban growth will occur around existing or newly formed urban clusters or along the major roads.

10. Discussion and conclusion

Logistic regression modeling was used to identify and improve our understanding of the demographic, econometric and biophysical forces that have driven the urban growth and to find the most probable sites of urban growth in Atlanta. The following two groups of factors were found to affect urban growth in different degrees as indicated by odd ratios: (1) population density, distances to the nearest urban clusters, activity centers and roads, and high/low density urban uses (all with odds ratios < 1); and (2) distance to the CBD, number of urban cells within a 7×7 cell window, bare land, crop/grass land, forest, and UTM northing coordinate (all with odds ratios > 1). The predicted spatial patterns of the future urban areas are the compromised outcomes of the above driving forces.

The previous research by Yang and Lo (2003) using Clarke's CA model (Clarke & Gaydos, 1998) for the same study area predicted the future patterns under three scenarios: unchecked spontaneous growth, consideration of road planning and environmental protection, and controlled growth. The first two scenarios generated very similar patterns characterized by huge compact urban agglomerations. The last scenario led to a pattern very similar to what is predicted from this study: urban growth areas will mainly be around existing urban areas and close to major roads, while some new clusters located at a distance from the existing urban areas can also form. Like the CA model, the logistic regression has the ability to incorporate organic, spontaneous, and diffusive growth mechanisms, road influence, as well as ecological preservation and environmental protection practices. The influence of distance to the nearest urban cluster corresponds to Clarke's 'edge growth' rule. This study included distance to roads or "road-gravity" as exogenous variables. The road influence is one of the growth rules in Clarke's CA model. The interaction term – number of urban cells in a window – has successfully captured the neighborhood effect which is also one of the growth rules in the dynamic CA model. Like the CA model, the logistic regression model is spatially explicit. The probability map generated from the model can be used to predict where urban growth will occur.

In the broader context of land use/cover change modeling, the criterion for evaluating a land use/cover change model is how effectively the model deals with the dimensions of space, time, human and scale dynamics. This study has shown that a logistic regression model has strengths relative to a CA model in two aspects. First, the logistic regression model can not only includes such biophysical variables as SLEUTH (slope, land use, exclusion, urban extent, transportation, hillshade) in the CA model, it is better for incorporation of human drivers. The model's ability to include as many demographic and econometric variables as necessary allows us to better understand human forces in shaping urban patterns. Second, logistic regression allows multi-scale calibration due to less demand of computation resource. Thus the model is better for capture of scale dynamics. This study used the fractal method to determine the optimal scale at which most processes operate to drive the urban growth and the model has the highest prediction ability. Previous computation intensive CA models were calibrated at a single resolution determined by data measurement levels or constrained by the computation power, which may not be appropriate for best understanding the land use change processes.

Despite the logistic regression model's strengths, this study has shown the limitations of the model. First, although the logistic regression model can incorporate demographic data, it suffers the same limitation as CA models in considering other factors which may have effects on the urban growth. These factors include, for example, personal or household preferences for locations, urban or regional development policies, and globalization of economy. Second, unlike the CA model, the logistic regression model is not temporally explicit. Its output probability map can only indicate *where* urban development will occur, but not *when* this will take place. Although the model prediction used updated raster layers of explanatory variables to generate an urban growth probability map, it does not have the ability of self-modification of the system status, hence a lack of temporal dynamics. Third, while the optimum resolution can be determined by multi-scale modeling and the fractal analysis, the modeling based on the single resolution could not capture all the processes behind urban growth. Modeling at the optimum scale ignores some processes that might be important and operate at lower-end or higher-end scales away from the optimal scale determined.

Several lessons have been learned from the study and suggest further research. First, when using a logistical regression model to study urban growth, we must be cautious about spatial autocorrelation that often exists in spatially referenced data which violates the assumption of the model. This study has demonstrated the use of GIS data aggregation, inclusion of spatial coordinates as variables and stratified random sampling to account for the effect of spatial autocorrelation. Second, the lack of ability to consider personal behaviors necessitates combination of the emerging land use modeling technique – agent based modeling (ABM) (Rand et al., 2003) with statistical models. Third, to overcome the weakness of logistical regression modeling in dealing with temporal dynamics, further research will have to seek a self-modifying approach so that the model variables can update themselves automatically. A possible solution would be to combine a logistic regression model with a CA model such that the predicted probability values can participate in the formulation of the rules in the dynamic simulation. Last, future research will have to address the multi-scale characteristics of land use/cover systems by using multilevel statistics or a hierarchical modeling structure so that the scalar dynamics of the land use/cover change driving forces operating from both bottom-up (micro-behavior) and top-down (such as regional planning policies) can be handled.

References

Agarwal, C., Green, G. M., Grove, J. M., Evans, T. P., & Schweik, C. M. (2001). A review and assessment of land-use change models: dynamics of space, time, and human choice. *USDA Forest Service, Northeastern Research Station, General Technical Report NE-297*.

Allen, J., & Lu, K. (2003). Modeling and prediction of future urban growth in the Charleston Region of South Carolina: a GIS-based Integrated Approach. *Conservation Ecology, 8*(2), 2 [online] URL:http://www.consecol.org/vol8/iss2/art2.

Atlanta Regional Commission (1997 & 2001). *Atlanta Region Outlook*. Atlanta, GA: Atlanta Regional Commission.

Bailey, T. C., & Gatrell, A. C. (1995). *Interactive spatial data analysis*. Harlow Essex/New York, NY: Longman Scientific & Technical/Wiley.

Bian, L. (1997). Multiscale nature of spatial data in scaling up environmental models. In D. A. Quattrochi & M. F. Goodchild (Eds.), *Scaling in remote sensing and GIS* (pp. 13–25). Boca Raton, FL: CRC/Lewis Publishers.

Bian, L., & Walsh, S. J. (1993). Scale dependencies of vegetation and topography in a mountainous environment of Montana. *Professional Geographers, 45*, 1–11.

BICUMP (2000). *Moving beyond sprawl: The challenge for Metropolitan Atlanta*. Washington, DC: Brookings Institution Center on Urban & Metropolitan Policy.

Cao, C., & Lam, N. S. (1996). Understanding the scale and resolution effects in remote sensing and GIS. In D. A. Quattrochi & M. F. Goodchild (Eds.), *Scaling in remote sensing and GIS* (pp. 57–71). Boca Raton, FL: CRC/Lewis Publishers, Inc.

Clark, W. A., & Hosking, P. L. (1986). *Statistical methods for geographers*. New York: Wiley.

Clarke, K. C. (1986). Computation of the fractal dimension of topographic surfaces using the triangular prism surface area method. *Computers and Geosciences, 12*(5), 713–722.

Clarke, K. C., & Gaydos, L. J. (1998). Loose-coupling a cellular automata model & GIS: long-term urban growth prediction for San Francisco and Washington/Baltimore. *International Journal of Geographical Information Science, 12*, 699–714.

Clarke, K. C., Hoppen, S., & Gaydos, L. (1997). A self-modifying cellular automaton model of historical urbanization in the San Francisco Bay area. *Environment and Planning B: Planning and Design, 24*, 247–261.

Congalton, R. G. (1988). A comparison of sampling schemes used in generating error matrices for assessing the accuracy of maps generated from remotely sensed data. *Photogrammetric Engineering and Remote Sensing, 54*(5), 593–600.

Dietzel, C., & Clarke, K. (2006). The effect of disaggregating land use categories in cellular automata during model calibration and forecasting. *Computers, Environment and Urban Systems, 30*, 78–101.

Domencich, T. A., & McFadden, D. (1975). *Urban travel demand: Behavioural analysis*. Amsterdam: North-Holland.

Geoghegan, J., Villar, S. C., Klepeis, P., Mendoza, P. M., Himmelberger, Y. O., & Chowdhury, R. R. (2001). Modeling tropical deforestation in the southern Yucatan peninsular region: comparing survey & satellite data. *Agriculture, Ecosystems and Environment, 85*, 25–46.

Goodchild, M. F., & Mark, D. M. (1987). The fractal nature of geographical phenomena. *Annals of the Association of American Geographers, 77*(2), 265–278.

Hartshorn, T. A., & Muller, P. O. (1989). Suburban downtowns and the transformation of metropolitan Atlanta's business landscape. *Urban Geography, 10*(4), 375–395.

Hosmer, D. W., & Lemeshow, S. (1989). *Applied logistic regression*. New York: Wiley.

Irwin, E. G., & Geoghegan, J. (2001). Theory, data, methods: developing spatially explicit economic models of land use change. *Agriculture, Ecosystems and Environment, 85*, 7–23.

Jaggie, S., Quattrochi, A., & Lam, S. (1993). Implementation and operation of three fractal measurement algorithms for analysis of remote-sensing data. *Computers and Geosciences, 19*(6), 745–767.

Kleinbaum, D. G. (1994). *Logistic regression: A self-learning text*. New York: Springer.

Kok, K., & Veldcamp, A. (2001). Evaluating impact of spatial scales on land use pattern analysis in Central America. *Agriculture, Ecosystems and Environment, 85*, 205–221.

Lam, N. S., & Quattrochi, D. A. (1992). On the issues of scale, resolution, and fractal analysis in the mapping sciences. *Professional Geographers, 44*, 88–98.

Landis, J., & Zhang, M. (1998). The second generation of the California urban futures model. *Environment and Planning B: Planning and Design, 25*, 795–824.

Mandelbrot, B. B. (1983). *Fractal geometry of nature*. San Francisco: Freeman.

McFadden, D. S. (1973). Conditional logit analysis of qualitative choice behavior. In P. Zarembka (Ed.), *Frontiers in econometrics*. New York: Academic Press.

Meentemeyer, V. (1989). Geographical perspectives of space, time, and scale. *Landscape Ecology, 3*, 163–173.

Moellering, H., & Tobler, W. (1972). Geographic variances. *Geographical Analysis, 4*, 34–50.

Pontius, R. G. (2000). Quantification error versus location error in the comparison of categorical maps. *Photogrammetric Engineering and Remote Sensing, 66*(8), 1011–1016.

Quattrochi, D. A., Lam, N. S., Qiu, H. L., & Zhao, W. (1997). Image characterization and modeling system (ICAMS): a geographic information system for the characterization and modeling of multiscale remote sensing data. In D. A. Quattrochi & M. F. Goodchild (Eds.), *Scaling in remote sensing and GIS* (pp. 295–307). Boca Raton, FL: CRC/Lewis Publishers, Inc.

Rand, W., Brown, D. G., Page, S. E., Riolo, R., Fernandez, L. E. & Zellner, M. (2003). Statistical validation of spatial patterns in agent-based models. URL: http://cscs.umich.edu/research/projects/sluce/.

Schneider, L., & Pontius, R. G. (2001). Modeling land use change in the Ipswich watershed, Massachusetts, USA. *Agriculture, Ecosystems and Environment, 85*, 83–94.

Serneels, S., & Lambin, E. F. (2001). Proximate causes of land-use change in Narok District, Kenya: a spatial statistical model. *Agriculture, Ecosystems and Environment, 85*, 65–81.

Veldkamp, A., & Lambin, E. F. (2001). Predicting land-use change. *Agriculture, Ecosystems and Environment, 85*, 1–6.

Walsh, S. J., Crawford, T. W., Welsh, W. F., & Crews-Meyer, K. A. (2001). A multiscale analysis of LULC and NDVI variation in Nanrong district, northeast Thailand. *Agriculture, Ecosystems and Environment, 85*, 47–64.

Weigel, S. J. (1996). Scale, resolution and resampling: representation and analysis of remotely sensed landscapes across scale in geographic information systems. Ph.D. dissertation, Louisiana State University.

Wu, F., & Yeh, A. G. (1997). Changing spatial distribution and determinants of land development in Chinese cities in the transition from a centrally planned economy to a socialist market economy: a case study of Guangzhou. *Urban Studies, 34*(11), 1851–1880.

Yang, X., & Lo, C. P. (2002). Using a time series of satellite imagery to detect land use and land cover changes in the Atlanta, Georgia metropolitan area. *International Journal of Remote Sensing, 23*(9), 1775–1798.

Yang, X., & Lo, C. P. (2003). Modeling urban growth and landscape changes in the Atlanta metropolitan area. *International Journal of Geographical Information Science, 17*(5), 463–488.