Understanding the complexity of urban expansion requires an analysis of the factors influencing the spatial and temporal processes of rural-urban land conversion. This study aims at building a statistical land conversion model to assist in understanding land use change patterns. Specifically, GIS coupled with a logistic regression model and exponential smoothing techniques is used for exploring the effects of various factors on land use change. These factors include population density, slope, proximity to roads, and surrounding land use, and their influence on land use change is studied for generating a predictive model. Methods to reduce spatial autocorrelation in a logistic regression framework are also discussed. An optimal sampling scheme that can eliminate spatial autocorrelation whilst maintaining adequate samples to allow the model to achieve the comparable accuracy as the spatial autoregressive model is developed. Since much of the previous studies on modeling the spatial complexity of urban growth ignored temporal complexity, a modified exponential smoothing technique is employed to produce a smoothed model from a series of bi-temporal models obtained from different time periods. The proposed model is validated using the multi-temporal land use data in New Castle County, Delaware. It is demonstrated that our approach provides an effective option for multi-temporal land use change modeling and the modeling results help interpret the land use change patterns.
1. Introduction

Worldwide, the level of urbanization is rapidly increasing and metropolitan areas are growing fast, creating extensive land use changes and urban spatial expansion. In 2005, a total of 3.2 billion people were urban dwellers, accounting for 48.7% of the world population (UNPD, 2005). Population growth is particularly rapid in the urban areas of developing countries. One very notable offshoot of rapid urbanization is consumption of agricultural land by urban built-up areas. Though the rates of urbanization are not accelerating in industrialized countries, where 75 per cent of the population has been urbanized, the concentration of population in metropolitan areas is expected to continuously increase (UNPD, 2004). Furthermore, metropolitan population outside central cities has grown faster than downtown areas in many developed regions, indicating a strong tendency of the outward expansion of urban areas (Angel et al., 2005). In fact, many cities are rapidly growing at their fringes, engulfing villages and farmlands and transforming them into dense industrial and commercial areas, or less dense suburban developments.

The formulation of relevant policies and regulations depends primarily on detecting and projecting urban expansion irrespective of the decision to resist, or welcome urban expansion. Urban expansion, in this study, is seen as an increase in developed and residential areas transformed from rural areas. Quite frequently, the causality and consequences of urban expansion remain contentious areas. While urban expansion is inevitable in response to social and economic development and has often been viewed as a sign of the vitality of regional economy (Yang and Lo, 2003), some castigate it as “sprawl”, therefore unnecessarily costly. Anti-growth advocates have invoked their critics on the adverse effects of urban sprawl. Albeit almost everyone has witnessed urban expansion and anticipated its positive and negative impacts in their local environment, a research gap exists in a proper understanding of the spatial and temporal processes of rural-urban land use change. On one hand, with the on-going development of remote sensing technology, image processing, artificial intelligence, and machine learning, a wide variety of digital change detection algorithms have been developed for land use modeling over the last two decades (see reviews in Mas, 1999; Coppin et al., 2003). On the other hand, the need to develop geographic understanding of land use change in urban areas still exists. Especially, it is important to identify the factors that driving the dynamic processes of rural-urban land transformation and to explore their relative importance, and, on that basis, to simulate “what-if” decision making under alternative
growth scenarios. Furthermore, the validation of theoretical models continues to pose a challenge owing to the shortage of reasonable methods to represent the complicated interactions among the factors involved in the process of land use change.

This study attempts to fill in the aforementioned research gap between change detection and interpretation of change patterns. We measure land use changes and project the impact by developing a multi-temporal spatial regression model. To interpret and predict land use changes based on a series of bi-temporal models, our model is constructed using GIS-based predictor variables and exponential smoothing technique. A post-classification scheme is used to statistically analyze land use changes among multiple time stamps. The prognostic capacity of the model is validated based on a case study of New Castle County, Delaware, USA.

2. Land use change modeling

Urban land use change is the spatio-temporal reflection of urban growth. Such change is influenced by a set of social, economic, and political factors. Different driving forces have been identified in various studies, including the effects of natural environment, demographics, economy, transport system, preference (by people) for proximity, neighborhoods, and governmental policies (e.g., Cervero and Wu, 1997; Mayer and Somerville, 2000; Smersh et al., 2003; Angel et al., 2005). Those forces may produce a clustering and localized pattern of urbanization, where new development has tended to infill around existing development, as well as a dispersed trend, wherein urban land-uses increasingly spread out across a metropolitan region (Carrión-Flores and Irwin, 2004). The relative weights of specific forces depend critically on the context within which the actual development occurs. In the investigation on rural-urban land conversion in China, Cheng and Masser (2003) reported a wide range of factors, such as investment structure, industrial structure, commercialization of housing market, land leasing system, transportation network, and decentralization of decision-making process. They also highlighted the physical constraints imposed by water bodies and other spaces inappropriate for urban development. However, only a few of those factors were incorporated in their bi-temporal models. The land use change study by Landis and Zhang (2000) incorporated four categories of information: the transportation network, the urban structure (residential, commercial, public and industrial buildings), the locations available for change, and the locations where changes were impossible. Verburg et al. (2001)
related the physical growth of a city directly to its population growth, as an increase in population size that encouraged the agglomeration of businesses and new urban development. The existence and accessibility of transportation routes often dictate patterns of urban growth, as a modern society is heavily dependent on automobiles and urban growth often occurs along transportation corridors. As land use change modeling seeks to account for physical expansion of urban land and to project future urbanization under various scenarios, such models have tended to incorporate spatial parameters, such as present urban extent, major transportation routes, and protected lands, in the analysis process.

Correspondingly, a broad range of techniques have been attempted, including multiple regression (Theobald and Hobbs 1998), Markov chain analysis (Lopez et al., 2001; Weng, 2002), cellular automata (CA) (Clarke and Gaydos, 1998; Wu, 1998; Batty et al., 1999; Li and Yeh, 2001), and logistic regression (Wu and Yeh, 1997; Cheng and Masser, 2003). In spite of the varying levels of success demonstrated by these techniques, they are also limited in certain aspects. For example, multiple regressions require fundamental assumptions, such as the normal distribution, appropriated error structure of the variables, independence of variables, and model linearity (Olden and Jackson, 2001). Unfortunately, the data on land use change often violate most of these assumptions. As a result, such models hardly ensure high generalization performances for projecting future land use change. Markov chain analysis assumes that the future state of a system can be modeled purely on the basis of its immediately preceding state. This method can be used to describe land use change from one period to another. However, it lacks explanatory power as the causal relationships underlying the transition studies are left unexplored (Baker, 1989). Cellular automata are an effective bottom-up simulation tool for dynamic process modeling, benefiting from its simplicity, transparency, and strong capacities for scenario simulation (Clarke and Gaydos, 1998). Nevertheless, CA models focus on the simulation of spatial patterns rather than the interpretation of the spatio-temporal process of land use change. Besides, the calibration of a CA model is a time-consuming process. By contrast, logistic regression has been found to be a more effective tool to analyze and interpret land use change (Cheng and Masser, 2003; Munroe et al., 2004; Páez and Suzuki, 2001). This method is capable of establishing functional relationships between the probability of land use change and the drivers of change represented by a set of explanatory variables. The relative significance of each explanatory variable can be generated using the maximum likelihood estimation. Despite the strengths of
logistic regression modeling and other models, some crucial problems still remain in the context of land use change modeling. For instance, many of the post-change analytical studies have not considered multi-temporal changes as their models are constructed only based on two available snapshots of time. Spatial and temporal predictions in terms of the samples have also not been carefully considered.

With due consideration to the aforementioned problems, we have developed a multi-temporal spatial regression model using GIS-based variables to interpret and project the urbanization patterns caused by land use change. This model primarily considers the spatial predictors, such as population density, slope, proximity to residential and commercial areas, contiguity to major transportation routes, and the proportion of urban as well as rural cells in the vicinity. Agricultural land is considered as ‘available for change’, whilst forests, wetlands, or barren lands are considered ‘unsuitable for development’ from the perspective of environmental conservation. In the subsequent sections of the paper, we will elaborate our model using the case of New Castle County, Delaware. The general background of the study area is provided in the next section.

3. Study area

New Castle County, covering about 1,100 km², is one of the three counties of Delaware State (Fig.1). It is the urban and manufacturing center of Delaware and is home to approximately 60% of Delaware’s population. Industrial and urban developments are concentrated in the northern part of the county. Changes in land use show that both the northern and southern parts were becoming progressively urbanized and such tendency seemed to continue in the coming years.
4. Multi-temporal spatial regression modeling

To interpret the spatial patterns of urban expansion, this study employed both logistic regression that incorporated spatial sampling to deal with spatial autocorrelation and exponential smoothing techniques to integrate bi-temporal models. A post-classification assessment was used to analyze the land use change between two periods of time. ArcGIS was used to compile predictor variables. Spatial sampling was implemented to reduce spatial dependence. The effects of different sampling schemes were carefully evaluated to ensure that the sampling scheme can filter out much of the spatial autocorrelation whilst generating adequate samples for regression analysis. Exponential smoothing technique was employed to generate a smoothed model from a series of bi-temporal models. Binary logistic regression was carried out using a spatial change analysis plug-in that we developed using C++ and incorporated in ArcGIS.

4.1 Logistic regression

Logistic regression has been widely applied to model the outcomes of a categorical dependent variable while the independent variables can be mixture of both continuous and
categorical variables. It is a suitable approach to estimate the coefficients of casual factors from the observation of land use change as the determinants of land use change are usually a mixture of continuous and categorical variables. In this paper, a logistic regression model was used to associate the land use change with demographic, econometric, and physical driving forces and to generate an urban growth probability map. The nature of the land use/cover change of a cell was regarded as dichotomous: either the presence of rural-urban transition or no transition. We used binary values 1 and 0 to represent rural-urban transition and no transition, respectively. It was assumed that the probability of an agricultural cell changing to an urban cell would follow the logistic curve. The general form of logistic regression is as follows:

\[ y = a + b_1 x_1 + b_2 x_2 + \ldots + b_m x_m \]  
\[ y = \log_e \left( \frac{P}{1-P} \right) = \log \text{it}(P) \]  
\[ P(z = 1) = \frac{e^y}{1+e^y} \]

where \( x_1, x_2, \ldots, x_m \) are explanatory variables. The utility function \( y \) is a linear combination of the explanatory variables representing a linear relationship. The parameters \( b_1, b_2, \ldots, b_m \) are the regression coefficients to be estimated. If \( z \) is denoted as a binary response variable (0 or 1), value 1 \((z = 1)\) means the occurrence of a new unit (i.e. the transition from a rural unit to an urban unit), and value 0 \((z = 0)\) indicates no change. \( P \) refers to the probability of occurrence of a new unit, i.e. \( z = 1 \). Function \( y \) is represented as logit \((P)\), i.e. the log (to base e) of the odds or likelihood ratio that the dependent variable \( z \) is 1. As the \( y \) value increases, probability \( P \) inevitably increases. The role of each explanatory variable on probability value \( P \) is denoted by the regression coefficients \( b_1\) to \( b_m \). A positive sign indicates that the explanatory variable helps to increase the probability of change and a negative sign implies the reverse effect.

The heterogeneity of spatial data needs to be considered when employing logistic regression to model rural-urban land conversion. To avoid unreliable parameter estimation, spatial dependence should be tackled. This can generally be done using two approaches. A major approach is to build a model that incorporates an autoregressive structure (Anselin, 1988). For example, a spatial lag model can be used as an extension of Model (1):
\[ y = a + \sum_{i=1}^{n} x_i b_i + \rho Wy + \varepsilon \]  \hspace{1cm} (4)

where \( \rho \) is a coefficient on the spatially lagged dependent variable and \( W \) is a spatial weight matrix. Note that the maximum likelihood estimator (MLE) is usually employed to solve for the parameters of such a model that best fit the data:

\[
L = \left[ y \ln\left(\frac{\exp(a + X\beta + \rho Wy)}{1 + \exp(a + X\beta + \rho Wy)}\right) - (1 - y) \ln(1 + \exp(a + X\beta + \rho Wy)) \right]
\]  \hspace{1cm} (5)

The likelihood can be maximized using a simplex unvaried optimization routine (LeSage, 1998).

The logistic version of Model (4), which incorporates spatial autocorrelation, called Spatial Autologistic Regression model or Spatial AutoLogit (SAL) model has also been devised (Dubin 1995; Dubin 1997; LeSage, 1998). Such models have proven to be effective in regression with consideration of spatial autocorrelation (Páez and Suzuki, 2001). Similar to the spatial lag model, a spatial error model (Anselin, 1988) can also be utilized to deal with spatial autocorrelation in a regression model. Nonetheless, they are primarily used for diagnostic analysis rather than for extrapolation-like prediction (Jetz et al., 2005).

Another workable approach is to filter out spatial autocorrelation (Getis and Griffith, 2002). Designing a spatial sampling scheme to reduce spatial autocorrelation among the sampled sites can also serve the purpose (Munroe et al., 2004). Although spatial sampling could lead to a smaller sample size that would lose some information and generate conflicts with the large-sample of asymptotic normality using the maximum likelihood method, those effects can be minimized by a reasonable design of the spatial sampling scheme. In this paper, the later approach was adopted to both analyze and predict change patterns based on the sample data minimizing spatial autocorrelation. It was yet expected to achieve comparable modeling accuracy as the previous approach. The sampling scheme will be discussed in more detail in section 4.4.

4.2 GIS-based predictor variables

Table 1 lists the predictor variables used in the proposed model. Seven predictors were grouped into three categories: (1) Site specific characteristics; (2) Proximity extent; and (3) Neighborhood characteristics. Population was considered as a leading force propelling land
use change. Very often, population density is one of the important criteria for urban designation. Hence, population density was considered as a chief predictor. Slope, a physical characteristic of land cells, and the zoning plan were also included in this category. Proximity extent reflects the fact that urban expansion can be driven by job/business opportunities and transport infrastructure. The proximity variables measure the minimum Euclidean distances to the nearest commercial site, residential area, and transportation network, respectively. The distance decaying mechanism of various factors is reinforced by the type and size of the selected neighborhood. In this study, a circular neighborhood with a 200 m radius was selected after considering the effect of neighboring impacts in current land use distribution. The spatial distributions of some predictor variables, which were derived using the ArcGIS software, are shown in Fig. 2.

### Table 1. Summary of predictor variables for a rural-urban land conversion model

<table>
<thead>
<tr>
<th>Category</th>
<th>Variable</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Site-specific</td>
<td>Dens_Pop</td>
<td>Population density of the cell</td>
</tr>
<tr>
<td></td>
<td>Slope</td>
<td>Slope of the cell</td>
</tr>
<tr>
<td></td>
<td>Zoning</td>
<td>Zoning plan for development control</td>
</tr>
<tr>
<td>Proximity extent</td>
<td>Dist_Com</td>
<td>Distance from the cell to the nearest commercial site</td>
</tr>
<tr>
<td></td>
<td>Dist_Res</td>
<td>Distance from the cell to the nearest residential center</td>
</tr>
<tr>
<td></td>
<td>Dist_Road</td>
<td>Distance from the cell to the nearest road</td>
</tr>
<tr>
<td>Neighborhood characteristics</td>
<td>Per_Urb</td>
<td>Percentage of urban land use in the surrounding area</td>
</tr>
</tbody>
</table>
Figure 2. Spatial distribution of predicted factors: (a) population density; (b) slope; (c) zoning; (d) Dis_Com; (e) Dis_Res; (f) Dis_Road; (g) Per_Urb in 1984.
4.3 Data compilation

The data used in this study included land use and terrain data, demographic data, and transportation network data over three time-periods, 1984-1992, 1992-1997, and 1997-2002. The land use data were generated from digital orthophotos provided by the Delaware Office of State Planning Coordination (Fig. 3). All land use files in vector format were rasterized at a resolution of 50 m × 50 m. This resolution was fine enough for eliminating positional errors reported in the data.

Land-uses were classified into five types: residential, commercial, industrial, agricultural, and others. The first three land-uses (residential, commercial, and industrial) were categorized as urban uses. The others, including forest, water, and barren, were classified as unsuitable for the purpose of urban development. Agricultural land was considered as a potential land for urbanization. A post-classification comparison method was used to detect the land use change between two time stamps. Bi-temporal change maps were generated by overlaying individual classifications using ArcGIS.

Table 2 presents the areas of rural and urban land-uses in the form of cells in New Castle over different years. It is apparent that rural land-uses underwent a significant decrease, from 57.59% of the total area in 1984, to 50.21% in 1992, further to 47.47% in 1997, and continuously down to 46.47% in 2002. By contrast, urban land-uses increased persistently and substantially, from 25.13% in 1984, 34.78% in 1982, 37.63% in 1997, to 39.65% in 2002. The average annual change ratio of rural-urban land use decreased from 1.20% during the period 1984-1992, 0.57% during 1992-1997, to 0.40% during 1997-2002.
Figure 3. Rural-urban land use maps of New Castle County in different years
Table 2. Rural and urban land areas in New Castle in 1984, 1992, 1997, and 2002

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Cells</td>
<td>%</td>
<td>Cells</td>
<td>%</td>
<td>Cells</td>
</tr>
<tr>
<td>Rural</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cells</td>
<td>253018</td>
<td>57.59</td>
<td>220620</td>
<td>50.21</td>
</tr>
<tr>
<td>Urban</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cells</td>
<td>110407</td>
<td>25.13</td>
<td>152406</td>
<td>34.78</td>
</tr>
<tr>
<td>Others</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cells</td>
<td>75943</td>
<td>17.28</td>
<td>66342</td>
<td>15.10</td>
</tr>
</tbody>
</table>

As we could only obtain the population data for 1990 and 2000, the data for 1984, 1992, and 1997 were estimated using an interpolation method. First, the population densities of 1990 and 2000 were rasterized. For each raster cell, an exponential growth of the population was assumed to be \( \alpha \cdot \exp(\beta \cdot \text{year}) \) (Lopez et al., 2001). The values of \( \alpha \) (the population in the base year) and \( \beta \) (the time interval between the base year and the year estimated) were derived from the 1990 and 2000 data. Then the population densities for 1984, 1992, and 1997 were calculated based on the derived \( \alpha \) and \( \beta \).

The transportation network data were based on the shapefile of the 2001 road network, the only available network data. Since large-scale construction of the US transportation network occurred around the 1960s, the geographic difference of transportation network among 1984, 1992 and 1997 could be assumed negligible. However, using the 2001 road network as the proxy of the 1984, 1992 and 1997 road networks might affect the accuracy of our model.

4.4 Spatial sampling

In an econometric model, it is assumed that the error terms for each individual observation are uncorrelated. For this reason, a special approach is needed to sample the dataset in a way that filters any residual spatial autocorrelation efficiently. Theoretically, spatial autocorrelation should be subject to distance decay. This study used the regular sampling technique of a non-overlapping moving window and the central cell was retained for each window. To verify the spatial autocorrelation of observations proximate in space, a check of joins was determined each time by comparing land use change types of “adjacent” cells. A join is defined by sequential occurrences of like land use changes in adjacent cells, which are those central cells of the current sampling window and the sampling windows to the east, west, north, and south of this window. Fig. 4 presents the plot showing changes in the number
of joins against the increases in window size. On one hand, small sampling window is insufficient for removing spatial autocorrelation, whilst on the other hand, large sampling window leads to a smaller sample size that result in the loss of certain information and conflict with the large-sample of asymptotic normality of the maximum likelihood method, upon which logistic regression is based. We finally chose to sample the data using a 9×9 window, or 450 m (× 450 m) sampling distance after comparing different sample window sizes. At such a distance, much of the spatial autocorrelation for all the bi-temporal models (i.e., 84-92, 92-97, and 97-02) were filtered out, yet enough samples for regression analysis were still retained. The model proposed in this paper with the above sampling approach will be discussed later in section 5.1 in comparison with a spatial autoregressive model, where spatial autocorrelation is considered explicitly (LeSage, 1998).

![Figure 4. Number of joins in terms of window size](image)

**4.5 Exponential smoothing**

Previous studies done by others laid greater emphasis on spatial complexity of urban growth, whilst overlooking temporal complexity. The proposed rural-urban land conversion model was built using multi-periodic data, which allows the modeling of spatial complexity across different periods. In particular, temporal complexity was modeled using a time-series analysis
Exponential smoothing, a popular scheme to produce a smoothed time series, was used to combine these models into a prognostic model. This method allots exponentially decreasing weights as the observation becomes older. In other words, recent observations are given relatively more weight in forecasting than older observations.

Single exponential smoothing is a well-known and widely used smoothing technique. This smoothing scheme begins by setting \( S_2 \) to \( y_1 \), where \( y \) stands for the original observation, and \( S_i \) stands for smoothed observation or EWMA (exponential weighted moving average), which is calculated from the observations \( y_{1~i-1} \) and used to estimate a future observation \( y_i \). The subscripts refer to the time periods, 1, 2, ..., and \( n \). For the third period, \( S_3 = \alpha y_2 + (1-\alpha)S_2 \), and so on. There is no \( S_1 \) since there is no previous observation available to be used to estimate the observation at period 1; the smoothed series starts with the smoothed version of the second observation, which establishes estimations from previous observations. For any time period \( t \), the smoothed value \( S_t \) can be found by computing

\[
S_t = \alpha y_{t-1} + (1-\alpha)S_{t-1}, \quad 0 < \alpha \leq 1, \quad t \geq 3
\]

The constant or parameter \( \alpha \) is called the smoothing constant. The speed at which the older responses are dampened (smoothed) is a function of the value of \( \alpha \). When \( \alpha \) is close to 1, dampening is fast and when \( \alpha \) is close to 0, dampening is slow. The optimal value for \( \alpha \) is the value which results in the smallest mean square error (MSE).

Although exponential smoothing is effective in generating a smoothed curve for ordinary time-series data, this method cannot be straightforwardly applied to land use change modeling as it requires the dataset to show a clear temporal trend. A modified method was, therefore, developed by reordering the time-series dataset in terms of the change ratio of a period instead of the recentness of a period. In this sense, a period witnessed large change was given a relatively higher weight in the model than a period that experienced small change.

Although exponential smoothing is used, in our case, to integrate the bi-temporal models, it generalizes the details of land use change over different phases within the long period. Exponential smoothing can be implemented at the coefficient level or the utility function level. At the coefficient level, smoothing is implemented on each coefficient of the rural-urban land conversion regression models. At the utility function level smoothing is implemented on the utilities of the rural-urban land conversion regression models. Smoothing
at the coefficient level requires a linear relationship. Since the utility function of the land conversion regression model is non-linear, performing smoothing on the utilities was employed in this study.

5. Results and discussion

5.1 Logistic regression results

As can be observed from Table 2, there exist a large number of cells without land use change for the three bi-temporal periods. The percentage of cells with land use changes is 9.65% in 1984-1992, 2.85% in 1992-1997, and 2.02% in 1997-2002. The cells without land use change significantly outnumber the cells with land use change. To maintain the model accuracy, it must ensure that small, but important, areas be represented in the samples (Congalton, 1988). Therefore, a ratio between change and unchanged sample cells should be specified. In our case, the unchanged/changed cells ratio was empirically found to be 2/3 to maximize the modeling accuracy and the sample sizes of the three bi-temporal models were accordingly set as 800, 700 and 600, respectively.

Table 3 presents the regression results generated by our logistic regression model estimated using the maximum likelihood algorithm. All the three models are statistically significant with the p-value as 0.000. The overall percentages of correctness are 75.62% for the 1984-1992 period, 71.13% for the 1992-1997 period, and 68.52% for the 1997-2002 period, respectively.
Table 3. Regression results

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Coefficient</td>
<td>t- probability</td>
<td>Coefficient</td>
</tr>
<tr>
<td>Pop_Dens</td>
<td>0.2323</td>
<td>0.0241</td>
<td>0.2716</td>
</tr>
<tr>
<td>Zoning</td>
<td>-0.2702</td>
<td>0.0244</td>
<td>-0.6721</td>
</tr>
<tr>
<td>Dist_Road</td>
<td>-1.5481</td>
<td>0.0000</td>
<td>-0.0441</td>
</tr>
<tr>
<td>Dist_Com</td>
<td>0.2762</td>
<td>0.0066</td>
<td>0.0084</td>
</tr>
<tr>
<td>Dist_Res</td>
<td>-0.4429</td>
<td>0.0019</td>
<td>0.2342</td>
</tr>
<tr>
<td>Per_Urb</td>
<td>1.1289</td>
<td>0.0000</td>
<td>1.9035</td>
</tr>
<tr>
<td>Constant</td>
<td>-0.8155</td>
<td>0.0000</td>
<td>-0.8278</td>
</tr>
</tbody>
</table>

Models Description

-2Log Likelihood: -412.89, -438.28, -458.79
P-Value: 0.000, 0.000, 0.000
PCP: 75.62%, 71.13%, 68.52%
Sample size: 800, 700, 600

We employed the same samples used above in a Spatial AutoLogit (SAL) model (LeSage, 1998) to examine whether the proposed sampling scheme could remove spatial autocorrelation and achieve similar modeling accuracy as the SAL model. The basic principle is that if $\rho$ in front of the spatially lagged variable $W\gamma$ is close to zero in the SAL model, spatial dependence in the samples is not significant. In addition, the Lagrange Multiplier (LM) test for the Chi-Square distribution with only one freedom can be used to detect whether the samples present spatial autocorrelation in the maximum likelihood residuals:

$$
LM = (1/T) [e'W e/\sigma^2] ~ \chi^2(1)
$$
$$
T = \text{tr}(W + W^T).*W
$$

where $e$ denotes residuals, and $.*$ denotes the element by element matrix multiplication.

Table 4 shows the modeling and statistic test results using the SAL model. The $\rho$ for the spatially lagged variable $W\gamma$ is 0.051, 0.053, and 0.064 for the three periods, respectively. They are all close to zero and are statistically significant at the 95% level (t-probability < 0.05), indicating that spatial autocorrelation has almost been removed using the proposed
sampling scheme. Furthermore, the LM test value is 3.29, 3.22, and 3.61 for the three periods, respectively, all of them being less than 6.64. This suggests that spatial autocorrelation in the regressive residuals is not significant at the 95% level, i.e., spatial autocorrelation almost does not exist in the residuals.

Table 4. Regression using the SAL model

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>ρ</td>
<td>Standard Error</td>
<td>t-probability</td>
</tr>
<tr>
<td>0.0511</td>
<td>0.0564</td>
<td>0.048</td>
</tr>
<tr>
<td>LM=3.29&lt;6.64</td>
<td>LM=3.22&lt;6.64</td>
<td>LM=3.61&lt;6.64</td>
</tr>
</tbody>
</table>

Table 5 compares the accuracies generated by our proposed model and the SAL model. In general, the PCPs with respect to all types of area (urban area, agricultural area, and overall) are very similar, which shows that our model with spatial autocorrelation filtered out in advance can achieve comparable accuracy as the SAL model.

Table 5. Comparison of PCPs between the proposed model and the SAL model in each period

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Proposed Model</td>
<td>SAL Model</td>
<td>Proposed Model</td>
<td>SAL Model</td>
</tr>
<tr>
<td>Urban</td>
<td>65.17%</td>
<td>65.17%</td>
<td>60.69%</td>
</tr>
<tr>
<td>Rural</td>
<td>80.56%</td>
<td>81.67%</td>
<td>75.16%</td>
</tr>
<tr>
<td>Overall</td>
<td>75.47%</td>
<td>76.20%</td>
<td>71.13%</td>
</tr>
</tbody>
</table>

It should be noted that, although SAL model can generate slightly better results than the proposed method in terms of accuracy, its modeling process is far more complex and time consuming due to the incorporation of the autoregressive structure. This is especially true for land use change analysis when there is a large dataset. For instance, the dataset used in our case consists of over $3.5 \times 10^5$ cells. Therefore, designing a spatial sampling scheme to remove spatial autocorrelation as in our model was preferred in land use modeling.
5.2 Validation of the smoothed model

To construct a unified model for the period of 1984-2002 from the three bi-temporal models, we used the exponential smoothing technique discussed in section 4.5 to smooth the utilities of those models. It was observed that urban expansion of the study area tended to slow down after 1992, although it was rapid from 1984 to 1992. This pattern met the requirement of exponential smoothing where the parameter could correspond to the change ratio of a period.

Fig. 5 shows that the accuracies vibrate with different $\alpha$ values. Since the cells with unchanged land-uses dominated the whole area, its accuracy largely influences the overall accuracy. However, the PCP of the cells with land use changes is a more important indicator to land use change analysis as it reflects the capability of the model in predicting the change. After careful comparisons of both the overall PCP and the PCP of cells with land use changes, a value of $\alpha = 0.6$ was selected. This $\alpha$ value was pragmatic because the pace of urban expanding reached to the peak in the period of 1984-1992. For the entire study period (1984-2002), the model associated with more robust urban expansion was more important than other models that were associated with slower paces of urban growth, and should be allocated a dominant weight. Thus the prognostic model used to predict future land use patterns is:

$$logit(P) = 0.6logit_{84-92}(P) + 0.6*(1-0.6)logit_{92-97}(P) + (1-0.6)^2logit_{97-02}(P)$$ (8)
Our model is validated in the following ways. At the outset, the candidate cells and their respective development status of 1984-2002 are first determined. For each candidate cell, the probability of change is computed with the fitted model (8). Subsequently, the probability of conversion is compared with a critical probability. If the probability of conversion at a cell is greater than the critical probability, the land represented by that cell is treated as ‘developed’; otherwise the land use remains unchanged. For binary logistic regression, the critical probability is normally set as 0.5.

Table 6 provides the comparisons of accuracy achieved by our smoothing (spatio-temporal) model and the direct (spatial) model. Our smoothing model (see Equation 8) is a generalization of three bi-temporal models, whereas the direct model is just a bi-temporal model which does not consider the changes occurred inside the period of 1984-2002. Our smoothing model achieves a higher overall PCP (70.36% vs. 68.87%). The accuracies for both the cells with land use change and the cells with land use change are also better (62.5% vs. 60.31%; 72.67% vs. 71.39%). These results indicate that the exponential smoothing method has achieved a better estimation than the direct method. All the PCPs are satisfactory as compared with other research outputs in this field (Li and Yeh, 2001). Figure 6 presents a visual comparison between the predicted land use change generated by our model and the actual land use change.

Table 6. Comparison of prediction with observation for the period of 1984-2002 using direct and smoothing models

<table>
<thead>
<tr>
<th>Predicted</th>
<th>Direct model</th>
<th>Smoothing model</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Observed</td>
<td>Predicted</td>
</tr>
<tr>
<td></td>
<td>Cells with</td>
<td>Cells with</td>
</tr>
<tr>
<td></td>
<td>land use</td>
<td>land use</td>
</tr>
<tr>
<td></td>
<td>change</td>
<td>change</td>
</tr>
<tr>
<td>Cells with land use</td>
<td>48147</td>
<td>49895</td>
</tr>
<tr>
<td>change</td>
<td>77679</td>
<td>74180</td>
</tr>
<tr>
<td>Overall</td>
<td>125826</td>
<td>124075</td>
</tr>
<tr>
<td>Cells without land use</td>
<td>31685</td>
<td>29937</td>
</tr>
<tr>
<td>change</td>
<td>193785</td>
<td>197284</td>
</tr>
<tr>
<td></td>
<td>225470</td>
<td>227221</td>
</tr>
<tr>
<td>Overall</td>
<td>79832</td>
<td>79832</td>
</tr>
<tr>
<td></td>
<td>271464</td>
<td>271464</td>
</tr>
<tr>
<td></td>
<td>351296</td>
<td>351296</td>
</tr>
<tr>
<td>PCP</td>
<td>60.31%</td>
<td>62.50%</td>
</tr>
<tr>
<td></td>
<td>71.39%</td>
<td>72.67%</td>
</tr>
<tr>
<td></td>
<td>68.87%</td>
<td>70.36%</td>
</tr>
</tbody>
</table>
Figure 6. (a) Predicted land use change in 2002 and (b) actual land use change in 2002

5.3 Discussion

Compared with simple equilibrium models, our model, incorporating simple rules of the effects of spatial adjacency that govern system dynamics and produce emergent patterns, more successfully reflects the demographic and physical forces that have driven the urban growth. It is also more powerful in forecasting future urban expansion more accurately. Furthermore, the model effectively addresses some of the crucial problems (i.e., model validation and generalization of multiple bi-temporal logistic regression models) that have not yet been well studied in other related works.

Our model, statistically and visually, explicates certain trends of the reality in land-use conversion over time. The model provides some interesting insights into whether factors that have been hypothesized to generate rural-urban land-use change in the literature are significant in influencing rural-urban land conversion in our study area. In our case, as revealed by our results (Table 3), rural-urban land-use change experiences both agglomerated and dispersed patterns. First, urbanized neighboring area, population density and distance to the commercial site positively affect rural-urban land conversion. However, the significance varies among these three predictor variables and with time. Population density and urbanized neighborhoods are consistently significant over time. By contrast, the positive impact of
distance to the commercial site reduces with time. These regression results show the distinctive agglomerative effect of urbanization, as much of rural-urban land conversion tends to occur near high-density and already urbanized areas, but not necessarily closer to the existing commercial site. It should be noted that slope is not statistically significant during all the periods, which is hence removed in the regression model.

Secondly, negative congestion externalities from higher agglomeration may push development away from established urban areas. Variables that have certain effect to control or to reduce development density, including zoning policy, distance to the road, and distance to the residential area, are negatively related to rural-urban land change. Surely, zoning policy is powerful to control urban development. An increase in development density at the neighborhood level may generate certain adverse effects that further on-side development is controlled and therefore urban expansion is encouraged. Urban development tends not to remain very close to the existing residential area in order to avoid the deterioration of living environment. The negative effect of Dist_Road variable reflects the fact that transportation infrastructure acts as a centrifugal force for the physical expansion of urbanization towards those sub-urban and outlying areas. With the development of transportation network and cheaper means of transportation, especially affordable private cars, residential and commercial activities are increasingly dispersed to the localities where living environment is more attractive.

Spatially, the effect of all distance-related variables is insignificant after 1992, indicating that urban development can be increasingly independent of the effect of space-friction. Transportation development can produce a dispersed trend of rural-urban land conversion and increase the spatiality of urban development. At the county level, our results indicate that rural-urban land conversion is close to existing urban development but gradually occurs in a dispersed pattern away from areas that are already substantially developed. Nonetheless, it is difficult and beyond the scope of this paper to judge the economic and social efficiency of such development pattern.

Overall, our regression findings are consistent with the pattern of urban expansion depicted by Fig. 3. That series of maps show that in 1984 the northern part of the New Castle County was more urbanized, as more urban-typed land-uses (commercial, residential, and industrial) were identified. By contrast, the land in the southern part was mainly used for rural functions.
During the period that those maps referred to, rural-urban land conversion primarily seemed to occur in the northern part, proximate to the previously urbanized lands. It can be also observed that, along the major transport route, residential land use gradually expanded southward over time. Urbanized areas of New Castle County sprawled from north to south, resulting in an elongated city.

Our model is slightly subject to the data constraint, a universal problem of urban land use modeling. Owing to data scantiness, it is impossible to incorporate all variables for measuring urban expansion. We have simplified the model construction and emphasized the role of spatial factors on urbanization. The effect of socioeconomic variables is assumed to be reflected by the selected spatial factors. This would lead to approximations at different levels of analyses and distort the effect of some factors, as well as overlook the influence of other possible factors. The compilation of the data also suffers a few problems of spatial inconsistency and estimation (e.g. positional errors in the 1984 land use data and differences in resolution and classification; time extrapolation of population densities). Moreover, the selection of predictor variables would affect the explanatory and prognostic power of the model. For example, proximity predictor variables (e.g. distance to the nearest residential area) can be sensitive to the distribution of spatial objects such as commercial sites. Arguably, only the distance to an urban center may not sufficiently explain land use change. In some cases, the size of the urban center matters in affecting the decision making of urban planners and investors on urban development.

6. Conclusion

Concerns pertaining to ungainly urban expansion have recaptured the attention of policy makers, academics, and voters during the last decade (Angel et al., 2005). Formulation of effective policy for urban development requires a better understanding of the factors affecting land use change. This study develops a logistic regression to effectively model and analyze the role of spatial factors on urban expansion and to predict future land use patterns across space and time. By introducing an appropriate sampling scheme, spatial autocorrelation is successfully minimized yet with sufficient samples to enable the logistic model to achieve the accuracy comparable to other spatial logistic regression models. Hence, our method is not limited by spatial autocorrelation and can make both space and time-based predictions. By
using exponential smoothing, multiple bi-temporal models are generalized to create a comprehensive model to support change analysis and prediction over a long period of time.

Using the empirical data from New Castle County, Delaware, the prognostic capacity of our model is assessed. Though some limitations exist because of several data constraints noted earlier, the regression results reveal that our model has achieved a reasonable goodness-of-fit of the actual land use development. Hence, it is clear that our model can facilitate the prediction of the sites prone to urbanization. However, it should be noted that rural-urban land conversion modeling based on a single case would have other limitations, as different regions have been urbanized under substantially different socioeconomic conditions. While the exponential smoothing method can perform certain generalization of the bi-temporal models, capturing the land use dynamics occurred during a defined period is still a challenging task. Also, deriving a smoothing model efficiently rather than enumerating the alpha value still merits further investigation. Nonetheless, it is suggested that, using the techniques developed in this study, the dynamics of land use transformation should be modeled with integration of time dependency and the key dimensions of urban spatial expansion.

**Acknowledgements**

This research is funded by the Hong Kong Research Grants Council (RGC) under CERG project no. CUHK 444107, and their support is gratefully acknowledged. We would also like to thank the anonymous reviewers for their valuable comments and suggestions.

**References**


