I, Qiaojue Wang, hereby submit this original work as part of the requirements for the degree of Master of Arts in Geography.

It is entitled:
An Integrated Multilevel Approach to Urban Development Modeling at Grid, Census block and Municipality Levels

Student's name: Qiaojue Wang

This work and its defense approved by:

Committee chair: Changjoo Kim, Ph.D.

Committee member: Hongxing Liu, Ph.D.

Committee member: Michael Widener, Ph.D.
An Integrated Multilevel Approach to Urban Development Modeling
at Grid, Census block and Municipality Levels

A thesis submitted to the
Graduation School
of the University of Cincinnati
in partial fulfillment of the
requirements for the degree of

Master of Arts

in the Department of Geography
of the College of Arts and Sciences

2014

by

Qiaojue Wang

B.S. Wuhan University, China, 2011
Abstract

Urban development modeling is an essential part of land use and land cover studies. This article explores the geographic, socio-economic, demographic and commuting factors of urban development at grid, census block, and municipality levels. Specifically, multilevel multinomial logistic regression is used to model the urban development in Hamilton County, Ohio and Cincinnati metropolitan area during the periods of 1992 - 2010. In the first case study (Hamilton County, Ohio), the grid-level variables include slope, elevation, distance to downtown, distance to highway, and distance to railroad, all with 30-meter resolution. The municipality-level variables include elevation variance, road density, population density, and number of housing units. Parameter estimation results show that the municipality-level variables have significant influence on urban development. A prediction map of future urbanization is generated and validated against the real land use map of 2010 using the Relative Operating Characteristic (ROC) statistic. The ROC of 0.816 demonstrates good prediction power of the applied model. In the case study (Cincinnati metropolitan area), the same set of techniques employed by the first one is used, but the spatial unit of analysis is changed from grid level into census block and a municipality-level variable is added showing the percentage of residents who commute out of their home to work. The model is then validated by a ROC statistic of 0.859. This study shows that multilevel multinomial can be used for analyzing major factors contributing to urbanization, modeling development dynamics, and predicting future growth. The information provided by the model is meaningful for land use-related planning and decision-making.
Acknowledgements

I would like to express my sincere gratitude to Dr. Changjoo Kim, my advisor and committee chair, for his time, academic guidance, encouragement and inspiration throughout the entire period of the thesis. He is a wonderful teacher who has the largest impact on my education in University of Cincinnati. I would also like to thank my committee members, Dr. Widener and Dr. Liu, for their time and valuable input.

My thanks also go to other faculty members and my fellow graduate students for their kind help during the completion of my thesis.

Finally, and most importantly, I would like to thank my parents and my relatives, for their consistent love, support and encouragement.
Table of Contents

Abstract........................................................................................................................................ ii
Acknowledgements ......................................................................................................................... iv
Table of Contents ............................................................................................................................... v
List of Tables ....................................................................................................................................... viii
List of Figures .................................................................................................................................... ix

Chapter 1  Introduction ..................................................................................................................... 1
  1.1 Land use and land cover change ................................................................................................. 1
  1.2 Urban development .................................................................................................................... 1
  1.3 Research Objectives .................................................................................................................. 3

Chapter 2  Literature Review ............................................................................................................ 4
  2.1 Measure of urban development pattern and trend ..................................................................... 4
  2.2 Simulation-based model ............................................................................................................. 5
  2.3 Regression-based model ............................................................................................................ 7
  2.4 Model Validation ....................................................................................................................... 9

Chapter 3  Study Area and Data ...................................................................................................... 11
  3.1 Study Area ............................................................................................................................... 11
<table>
<thead>
<tr>
<th>Chapter</th>
<th>Title</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chapter 4</td>
<td>Methodology</td>
</tr>
<tr>
<td>4.1</td>
<td>Multilevel multinomial logistic regression model</td>
</tr>
<tr>
<td>4.2</td>
<td>Intra-class correlation (ICC)</td>
</tr>
<tr>
<td>4.3</td>
<td>Model validation</td>
</tr>
<tr>
<td>Chapter 5</td>
<td>Results and Analysis</td>
</tr>
<tr>
<td>5.1</td>
<td>Case Study 1: Hamilton County, Ohio</td>
</tr>
<tr>
<td>5.1.1</td>
<td>Modeling result</td>
</tr>
<tr>
<td>5.1.2</td>
<td>ICC</td>
</tr>
<tr>
<td>5.1.3</td>
<td>Prediction map</td>
</tr>
<tr>
<td>5.1.4</td>
<td>ROC statistic</td>
</tr>
<tr>
<td>5.2</td>
<td>Case Study 2: Cincinnati metropolitan area</td>
</tr>
<tr>
<td>5.2.1</td>
<td>Modeling result</td>
</tr>
<tr>
<td>5.2.2</td>
<td>ICC</td>
</tr>
<tr>
<td>5.2.3</td>
<td>Prediction map</td>
</tr>
<tr>
<td>5.3.4</td>
<td>ROC statistic</td>
</tr>
<tr>
<td>Chapter 6</td>
<td>Discussion &amp; Conclusion</td>
</tr>
</tbody>
</table>
List of Tables

Table 1: Typical ROC range in other research areas .......................................................... 10
Table 2: Summary of Data Management ............................................................................ 16
Table 3: Contingency table used in ROC statistic ............................................................... 24
Table 4: Fixed effects of Hamilton County, Ohio ............................................................... 28
Table 5: ICCs for Hamilton County, Ohio ......................................................................... 31
Table 6: Mutable cells in buffer area of downtown ............................................................. 33
Table 7: Mutable cells in buffer area of highway and railroad ........................................... 34
Table 8: ROC statistic for three categories of urban development in Hamilton County, Ohio ........................................................................................................................................ 35
Table 9: Fixed effects of Cincinnati Metropolitan Area ..................................................... 37
Table 10: ICCs for Cincinnati metropolitan area ............................................................... 40
Table 11: ROC statistic for three categories of urban development in Cincinnati metropolitan area .................................................................................................................................. 43
List of Figures

Figure 1: Hamilton County, Ohio ................................................................. 12
Figure 2: Cincinnati metropolitan area ....................................................... 13
Figure 3: Procedure for Hamilton County, Ohio ......................................... 18
Figure 4: Procedure for Cincinnati metropolitan area .................................. 19
Figure 5: ROC Curve example ................................................................... 25
Figure 6: Prediction map of urban development of Hamilton County, Ohio (2010) ....... 32
Figure 7: ROC curve of prediction map of Hamilton County, Ohio ............. 35
Figure 8: Change map of Open water to Urban ......................................... 38
Figure 9: Prediction map of Cincinnati metropolitan area ............................. 41
Figure 10: ROC curve of prediction map of Cincinnati metropolitan area ........ 43
Chapter 1

Introduction

1.1 Land use and land cover change

Land use and land cover change studies are becoming increasingly popular in recent years, and they involve complicated physical and socio-economic factors. Land use is different from land cover despite the fact that the two terms are often used interchangeably. Land use is the description of how people utilize the land, which involves the management and modification of natural environment into built environment, such as field and settlement. Land cover refers to the biophysical attribute of the earth’s surface, such as open water, grass, bare ground, etc (Lambin et al., 2001; Ramadan et al., 2004).

1.2 Urban development

Urban development is one of the most prevalent topics in the study of land use and land cover change since it is directly linked with the economic growth, indicated by the fact that urban population forms an overwhelming majority in all developed countries (B.K. Chakrabarty, 2007; Kaya and Curran, 2006; Seto and Fragkias, 2005; W.F. Smith,
1975). As human activities grow and demand for natural resources increases, the research on urban development gains more and more attention. A better understanding of the patterns, mechanisms, and effects of urban development is the key to better planning and management.

Modeling of urban development helps policy-makers identify the important factors of land use change and balance the relationship between environmental protection, economic growth and social welfare (Chakir et al., 2009). Urban development is generally considered as the result of the interplay between socio-economic, demographic and environmental factors (Overmars et al., 2005). Geist and Lambin (2002) identified the proximate and underlying causes as the driving forces of urban development. Proximate causes are the behaviors and activities that directly affect urban development, while underlying causes are the demographic, socio-economic, and ecological variables governing the proximate causes. Many regression-based techniques have been used to explore the relationships between urban development and underlying causes (Nelson et al., 2001).

Gibson et al. (2000) suggested using the concepts of scale and level in the regression-based approaches of land use modeling. Scale refers to the spatial, temporal, quantitative, or analytical dimension used to measure study objects. Level is the context in which the data are nested. To study the urbanization factors at different levels and scales, multilevel logistic modeling is an appropriate approach. Multilevel modeling has seen applications in public health, economics, and education (Diez-Roux, 2000; Paterson et al., 2008; Rabe-Hesketh et al., 1991). Recently, it has become more popular in
geographic research such as commute time modeling, land evaluation and land use (Hurni, 2000; Schwanen et al., 2004; Pan et al., 2005). Most of previous literature uses multilevel models to study the effects of social context on individual behaviors. In this study, multilevel analysis is applied to study urban development because land use policies exhibit clear hierarchical structure, nested within administrative regions.

1.3 Research Objectives

This paper aims to model and predict the location and probability of urban development with multilevel multinomial logistic regression. To show the diversity of land use change, four categories of urban development are considered: open water to urban, forest to urban, pasture to urban, and no change (i.e., remaining developed). Geographic, demographic, transportation and commuting factors are investigated in the regression modeling. The urbanization process is assumed to be irreversible (i.e., urban areas cannot change back to non-urban land use types). The assumption is justified by comparing the land use maps over the years. Predictions are validated by ROC (Relative Operation Characteristic) statistics (Pontius, 2001).
Chapter 2

Literature Review

2.1 Measure of urban development pattern and trend

With the advance of computer technology, GIS has been widely used to analyze land use change, such as deforestation and urbanization. It has advantages of storing, retrieving, analyzing and visualizing large dataset, thereby facilitating the decision-making process. In addition, remote sensing is applied in land use and land cover change studies. Land use maps are considered as the important tool to represent, characterize, and model land use changes. Land use maps of different time periods can be generated by appropriately classifying remote sensing images (Tan et al., 2005; Weng, 2001).

Li et al. (2004) used land use maps of two different years to locate and quantify land use changes during the period. Deng et al. (2009) detected and monitored land use change in the city of Hangzhou by using multi-sensor satellite images from Spot-5 multispectral and LRRandsat-7 ETM panchromatic data, via supervised classification and principal component analysis. Masser (2001) used remote sensing images from different time periods to detect and monitor urban development, and derived strategies in response to such development. Milesi et al. (2003) utilized GIS to analyze land use thematic maps and explored how urban expansion affects the ecosystem in Southeastern United States. They applied computer technology and spatial statistical analysis to detect urban
development, which provide an important measure of urban development pattern and trend. However, this is not adequate to explain the complex process of land use change by simply comparing and detecting the land use change. Land use change modeling is necessary in order to get a better understanding of the change process.

2.2 Simulation-based model

Simulation-based models analytically describe the spatio-temporal dynamics of urban development (Dietzel et al., 2005; Sims et al., 2004). Landis (1995) developed a California Urban Feature (CUF) model to simulate the urban development pattern affected by different urban planning policies in the San Francisco Bay and Sacramento area. Muller et al. (1994) used a Markov chain model to simulate urban development in the Niagara region in Ontario, Canada. Hathout (2002) also used the Markov chain analysis to explain how the urban development affects the loss of agricultural land by comparing West and East St. Paul in Manitoba, Canada. The major limitation of the Markov chain analysis is the lack of geographic information. The spatial distribution of urban development or different land use categories cannot be represented.

This issue is improved with the incorporation of Cellular Automata (CA), which is capable of incorporating spatial interaction effects and handling temporal dynamics. The CA model simulates urban development based on the assumption that the past urban development affects future urban development through local interactions among land uses. CA models operate on an array of identically programmed automata, or cells that
exist in one of a finite number of states. Macro-scale behaviors can be generated through repeated application of behavior rules according to the interaction between individual cells and their neighbors (Park and Wagner, 1997).

Arsanjania (2011) used a hybrid model of logistic regression, Markov chain, and cellular automata to model the urbanization of Tehran, Iran during 1986 – 2026. The independent variables considered in the regression are distance to CBD, Census, distance to nearby cities, northing coordinates, population density, distance to residential areas, distance to single buildings, easting coordinates, farming, distance to building blocks, elevation, distance to interchange, open lands, distance to parks, distance to roads, slope, and distance to streams. The 89% ROC value was reported by comparing model prediction and actual map of Year 2006.

Syphard (2005) calibrated a cellular automaton (CA) model using historical growth pattern in the region and used it to forecast the urban growth in the Santa Monica Mountains National Recreation Area (SMMNRA) from 2000 to 2050. Herold (2003) used Remote Sensing data, the SLEUTH urban growth model, and FRAGSTATS statistical software to study the urban growth of Santa Barbara, CA. The RS data including images ranging from 1929 to 1998, with which urban areas are identified. The SLEUTH model utilizes a rule-based cellular automata algorithm to process Remote Sensing data and forecasts urban growth between 2000 and 2030. The combined one-century urbanization maps, including both observed and projected, are fed into the FRAGSTATS software with which key spatial metrics, such as edge density, contagion, and fractal parameters are calculated. Results suggest the major alternation of calculated
metrics does match the real changes of urbanization pattern. Therefore, the spatial metrics are useful tools for analysis relating urban development.

The advantages of CA-based model are the simplicity, flexibility, intuitiveness and transparency of CA. It can be integrated with Geographical Information System and modeled at high spatial resolution with computational efficiency (Sante et al., 2010). However, the character of simplicity limits the ability of the model to represent real-world phenomena. Moreover, due to the flexibility of the model, modelers are free to design the most suitable model for each case, which results in the lack of a standard method for the definition of transition rules.

2.3 Regression-based model

To discover the underlying factors of land use change and explain how they interact with each other, regression-based models are employed (Seto and Kaufmann, 2003). Lambin (1997) indicated that the regression-based models are capable of explaining (1) which geographic and socio-economic variables cause land-cover changes, (2) the locations of the changes affected by the variables, and (3) the rate of the advancement of land-cover changes. Hu (2007) adopted a single-level, logistic regression model to simulate urban growth dynamics of Atlanta, GA. The land use images of Atlanta in 1987 and 1997 are employed for determining urbanization. A grid size of 225 meter is selected as the optimal spatial resolution according to the fractal dimension analysis. Results of parameter estimation show that population density, distances to
nearest urban clusters, activity centers and roads, and high/low density urban uses are negatively correlated with urban growth and distance to the CBD, number of urban cells within a 7*7 cell window, bare land, crop/grass land, forest, and UTM northing coordinate are positively correlated with urban growth. The generated urbanization probability map is validated by the ROC measure of 0.85.

Pontius (2001) used a regular logistic regression to build a model of deforestation in Massachusetts and predicted the land cover changes. However, the aforementioned studies only focus on the pixel-level variables, while ignoring the variables at ecological levels. Overmars et al. (2005) developed three regressive models to predict the land use change, namely geographic approach model, household model, and enhanced spatial model. These models explore the land use change drivers at the watershed scale and household level. The geographic approach model uses common geographic variables collected from maps, census report and remote sensing images. The household model identifies the relationships between land use changes and supposed driving factors (e.g., human-environmental factors) at the household level. The enhanced spatial model combines the strengths of geographic and household models, generating better prediction results. Though the data utilized in this study are nested at different levels, the structure for all three models is still single-level.

Overmars et al. (2006) improved the enhanced spatial model into a multilevel model while modeling the land use in the Philippines from field to village level. Five binary multilevel models are constructed to predict the changes of yellow corn and banana fields. Variances of prediction results are compared. Results show that the
explanatory variables can account for group level variability. Though the application of multilevel model is limited by data availability, the study shows that the hierarchical modeling structure is useful for land use change. Moreover, the binary logistic regression model is used to calculate the probability of land use change. However, due to the fact that binary model cannot differentiate the degrees of influence by urbanization driving forces on different types of non-urban land use, multinomial logistic regression is more appropriate (Luo et al., 2008; Yu et al., 2008). For example, Yu et al. (2008) modeled the dynamics of urban development of Jiayu County, Hubei Province of China with four non-urban types: cultivated, orchard, forest, and bare land. The predictive map was created by interpolating the odds of change ratios.

2.4 Model Validation

Land use models need to be validated prior to real applications. Natarajan et al. (2000) recommended a multinomial probit model using Monte Carlo Expectation-Maximization (MCEM) method, which is widely used in biometrics, and econometrics modeling (Chib, 1996; Lee et al., 2012). The predicted probability of the model is compared to a set of randomly simulated probability, and then the accuracy of the prediction model is calculated according to the maximum likelihood. In land use study, this method is rarely applied, since only the randomly simulated values are compared. Pontius et al. (2001) used the relative operating characteristics (ROC) as a quantitative measure to validate the model of deforestation in the Ipswich watershed, USA. The ROC curve is derived by comparing the map of real land use change to the map of model-
predicted probability. In his study, the prediction map of deforestation is calibrated with land use maps of 1971 and 1985 and validated by the maps of 1985 and 1991. The ROC for the model is calculated as 0.65.

In this paper, land use maps of 1992 and 2001 are used to generate the development model and calculate the prediction map. The land use maps of 2001 and 2010 are used to validate the model with ROC statistic. Table 1 provides the ranges of ROC in other research areas (Swets, 1988; Swets, 1986).

<table>
<thead>
<tr>
<th>Research area</th>
<th>ROC range</th>
</tr>
</thead>
<tbody>
<tr>
<td>Weather forecasting</td>
<td>0.71-0.89</td>
</tr>
<tr>
<td>Library information retrieval</td>
<td>0.75-0.97</td>
</tr>
<tr>
<td>Medical imaging diagnosis</td>
<td>0.81-0.93</td>
</tr>
<tr>
<td>Material strength testing</td>
<td>0.68-0.93</td>
</tr>
<tr>
<td>Polygraph lie detection</td>
<td>0.55-0.98</td>
</tr>
</tbody>
</table>
Chapter 3

Study Area and Data

3.1 Study Area

3.1.1 Case study 1: Hamilton County, Ohio

Hamilton County is located in the southwest corner of the state of Ohio, United States. The county has a total area of 1,068.7 km$^2$, of which 98.37% is land and 1.63% is water (Figure 1). It is a part of the Cincinnati – Middletown, OH-KY-IN Metropolitan Statistical Area. The county has 49 political jurisdictions including the city of Cincinnati and a population of 802,374 according to the 2010 Census, making it the third most populous county in Ohio. Hamilton County’s landscape consists of low, rolling hills formed by the slopes of the Ohio River Valley and tributaries. The Great Miami River, Little Miami River, and Mill Creek also contribute to the landscape. The major highways serving the county include I-71, I-74, I-75, I-471, I-275, the Norwood Lateral, and Ronald Reagan Highway. The railroad companies include CSX Transportation, Norfolk Southern, Rail America, and Amtrak.
3.1.2 Case study 2: Cincinnati metropolitan area

The second study covers seven other counties neighboring Hamilton County. The eight counties, located in the states of Ohio, Kentucky and Indiana, are collectively termed as the Cincinnati metropolitan area (Figure 2). The counties are Boone, Kenton and Campbell in Kentucky, Butler, Warren, Clermont, and Hamilton in Ohio, and Dearborn in Indiana. The Cincinnati metropolitan area has a population of 1,999,474, with the centering population of 296,943 in Cincinnati, Ohio (Census 2010). The major business districts include Hamilton and Middletown in Butler County, Lebanon in
Warren County, Cincinnati in Hamilton County, Covington in Kenton County and Newport in Campbell County.

**Figure 2**: Cincinnati metropolitan area

The purpose of investigating the Cincinnati metropolitan area is to explore the transportational factors to urban growth, which are expected to be more prominent in larger areas. More, Hamilton County alone is mostly urbanized already (with 2001 data),
and including neighboring counties with more undeveloped lands can test the model under different scenarios.

3.2 Data Management

Land use, census, road, geographic and commuting data are used in this part. Land use data for year 1992 (NLCD1992) and 2001 (NLCD2001) are downloaded from Multi-Resolution Land Characterization (MRLC) consortium, a group of federal agencies that coordinate and generate consistent and relevant land cover information at the national scale for a wide variety of environmental, land management, and modeling applications. Both datasets are at the same spatial resolution of 30 meters and based on unsupervised classifications. Note that the 1992 National Land Cover Dataset (NLCD 1992) is a 21-class land cover classification scheme and the 2001 Land Cover Dataset (NLCD 2001) is a 16-class land cover classification scheme. To keep the land use type consistent, both datasets are reclassified into four classifications, namely open water, forest, pasture and urban area. The 2010 land use map is derived from the Cincinnati Area Geographic Information System (CAGIS) vector data of impervious surface. The impervious surface is regarded as urban area, and the rest are non-urban area. The dataset is then converted to a 30m resolution raster, and reclassified into urban and non-urban land use types.

The numbers of housing units and population data are extracted from the 2000 and 2010 Censuses. The 2010 Digital Elevation Model (DEM) is used to generate
elevation and slope, with a 30m resolution consistent with the land use dataset. Commuting data for year 2000 and 2010 are downloaded from United States Census Bureau, based on the survey unit of census tract. The percentage of residents who commute out of their home to work is calculated within each municipality. Road data are downloaded from 2010 Census. Road features are then classified into three categories: highway, railroad, and local road. Downtown is also considered as an important factor that affects the land use change in this part, which is the location of Fountain Square in Cincinnati. These datasets are processed by calculating the distance between each pixel in the land cover raster dataset and the town center, the nearest highway, and the nearest railroad.

In the first case study of Hamilton County, Ohio, the unit of datasets is based on land use images which are 30m resolution, with the attributes of land use type in 1991, land use type in 2000, slope, elevation, distance to downtown, distance to highway, and distance to railroad. In the second case study of Cincinnati metropolitan area, the same datasets are processed for the spatial unit of census blocks, where the data of population density, number of housing units, elevation variance, local road density and percentage of workers commuting out are nested within municipality. The spatial unit and count of case items for the two case studies are shown in Table 2.
Table 2: Summary of Data Management

<table>
<thead>
<tr>
<th>Level</th>
<th>Hamilton County, Ohio</th>
<th>Cincinnati metropolitan area</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Spatial Unit</td>
<td>Count</td>
</tr>
<tr>
<td></td>
<td>Grid (30m)</td>
<td>1185442</td>
</tr>
<tr>
<td></td>
<td>Municipality</td>
<td>50</td>
</tr>
<tr>
<td></td>
<td>Census block</td>
<td>26222</td>
</tr>
<tr>
<td></td>
<td>Municipality</td>
<td>206</td>
</tr>
</tbody>
</table>


Chapter 4

Methodology

Multilevel multinomial regression is applied to model urban development from 1992 to 2001 and make predictions for 2010. There are two case studies in this research: Hamilton County, Ohio, and Cincinnati metropolitan area. The procedure of Hamilton County study is shown in Figure 3. The first step is to estimate the coefficients of the multilevel multinomial model using 1992 and 2000 land use data, 2000 census data, DEM, and road data. Secondly, intra-class correlation (ICC) is calculated to see municipal variations and justify the application of multilevel model. In the third step, the coefficients derived in first step, 2010 census data, DEM and road data are used to calibrate the prediction map of urban development in 2010. The fourth step is to validate the model result using 2010 land use map. The prediction power of the model is evaluated by Relative Operation Characteristics (ROC) statistics ranging from 0 to 1. Greater values of ROC indicate more accurate predictions.

The second case study follows the similar steps in Hamilton County case study, except for the input data shown in Figure 4. The study area is extended to the Cincinnati metropolitan area, and commuting data are added to explore the commuting effect on urbanization. The 2000 commuting data is added to the “Input Data 1”, and 2010 commuting data is added to the “Input Data 2”.
Input Data 1:
Landuse Map (1992, 2000)
DEM (2010)
Census (2000)
Road (2010)

Step 1: Estimate Multilevel Multinomial model coefficients

Step 2: Calibrate ICC

Step 3: Calibrate prediction map

Step 4: Map validation, Calculate and plot ROC curve.

Output:
Prediction Map

Input data 2:
DEM (2010)
Census (2010)
Road (2010)

Input data 3:
Landuse Map (2010)

**Figure 3**: Procedure for Hamilton County, Ohio
4.1 Multilevel multinomial logistic regression model

The format of the multilevel model is formally expressed as Eq.(1-3).

(Level-1 model) \[ Y_{ij} = \beta_{0j} + \beta_{1j}X_{ij} + \epsilon_{ij} \] (1)

(Level-2 model for intercept) \[ \beta_{0j} = g_{00} + g_{01}W_{0j} + u_{0j} \] (2)

(Level-2 model for slope) \[ \beta_{1j} = g_{10} + g_{11}W_{1j} + u_{1j} \] (3)
Eq. (1) is the level-1 model in which a single independent variable $X_{ij}$ is included to explain the dependent variable $Y_{ij}$. The symbol $i (i = 0,1,2,...)$ denotes the index of the variable at the individual level, and $j (j = 0,1,2,...)$ denotes the index of the aggregated region. $\beta_{0j}$ is the intercept of $Y_{ij}$, and $\beta_{1j}$ is the coefficient of variable $X_{ij}$. $\varepsilon_{ij}$ is the residual at the individual level. The intercept $\beta_{0j}$ and slope $\beta_{1j}$ in the level-1 model are regarded as dependent variables in the level-2 model. The level-2 model (Eq. (2-3)) considers two macro-level variables $W_{0j}$ and $W_{1j}$ in addition to micro-level variables. The $g$-coefficients are used to quantify the effect of level-2 variable. $u_{0j}$ and $u_{1j}$ represent the measurement errors. $g_{00}$ and $g_{10}$ are the intercepts of level-2 models for $W_{0j}$ and $W_{1j}$ respectively. $g_{01}$ and $g_{11}$ are the slopes of level-2 models. In this study, $\varepsilon_{ij}$, $u_{0j}$, and $u_{1j}$ are assumed to be normally distributed. A one-equation expression of the model, Eq.(4), is obtained by substituting the Eq.(2-3) into Eq.(1). The terms inside the first pair of parentheses represent the fixed effects and that inside the second one represent the random effects of the model.

$$Y_{ij} = (g_{00} + g_{01}W_{0j} + g_{10}X_{ij} + g_{11}X_{ij}W_{1j}) + (u_{0j} + u_{1j}X_{ij} + \varepsilon_{ij})$$  \hspace{1cm} (4)$$

Binary logistic is useful in situations where the dependent variables are dichotomous. The dependent variable of the model has binary values of 1 and 0. In urban development examples, 1 means the corresponding pixel has 100% probability of urban development, and 0 means urban development will not happen. The binary logistic regression model is expressed in Eq. (5).
\[
P(Y = 1 \mid X_1, X_2, \ldots, X_n) = \frac{1}{1 + e^{\sum_{i}^{n} \beta_i X_i}}
\]  

\(P(Y = 1 \mid X_1, X_2, \ldots, X_n)\) is the probability of \(Y = 1\) given independent variable \(X_i (i = 1, 2, \ldots, n)\). The odds of \(Y = 1\) are defined as Eq. (6).

\[
\ln\left(\frac{P}{1-P}\right) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_n X_n
\]  

Binary logistic regression can be extended to multinomial logistic regression when the dependent variable has more than two possible values. If there are \(K\) candidate values for the dependent variable, then we can obtain the expression as Eq. (7).

\[
\sum_{i=1}^{K} P(Y = i \mid X_1, X_2, \ldots, X_n) = 1
\]  

Eq. (8) is obtained by substituting Eq. (5-6) into Eq. (7).

\[
P(Y = K \mid X_1, X_2, \ldots, X_n)[1 + \sum_{j=1}^{K-1} e^{\beta_j + \sum_{k=j}^{K} \beta_k X_k}] = 1
\]  

The multilevel multinomial logistic regression model can be written as Eq. (9-11).

\[
\ln(\mu_{ij}) = \beta_{0j} + \sum_{m=1}^{N} \beta_{mj} X_{mij}
\]  

\[
\beta_{mj} = g_{m0} + g_{m1} W_{mj} + u_{mj}, \quad m = 0, 1, 2, \ldots, N
\]  

\[
\mu_{ij} = \frac{\exp(\beta_{0j} + \sum_{m=1}^{N} \beta_{mj} X_{mij})}{1 + \sum_{j=2}^{K} \exp(\beta_{0j} + \sum_{m=1}^{N} \beta_{mj} X_{mij})}
\]
In Eq. (9) and (11), \( \mu_j \) is the probability of the \( i \)th observation taking the value of \( j \) where \( j = 1, 2, \ldots, K \). \( X_{mij} \) is the \( m \)th independent variable at level-1. \( \beta_{0j} \) and \( \beta_{mj} \) are the intercepts and coefficients at level-1. Eq. (10) shows the expression of the level-2 model, where \( \beta_{mj} \) is the dependent variable, \( u_{mj} \) is the level-2 residual, and \( W_{mj} \) is the independent variable affiliated to the aggregated region \( j \) with the intercept \((g_{m0})\) and the coefficient \((g_{m1})\).

### 4.2 Intra-class correlation (ICC)

Intra-class correlation (ICC) is calculated to validate the use of multiple levels for a hierarchical dataset, which is a measurement of the degree of correlation among observations within a cluster. ICC explains how much of the variance in the dependent variable in the multilevel model stems from between-municipality differences. A higher ICC implies that most of the differences across the dependent variable are stemming from municipality difference, which means the “between- municipality variance dominates the within- municipality variance”. Conversely, a small ICC indicates that most of the differences stem from individual differences within municipality.

ICC can be determined in a random-intercept model, which means the \( u \) - variables listed in Eq. (2-4) and Eq. (10) are null except for \( u_{0j} \). The variance of level-1 residual \( \varepsilon_{ij} \) is set to \( \sigma_\varepsilon^2 \), and the variance of level-2 residual \( u_{mj} \) is denoted by \( \sigma_{u0}^2 \). ICC in two-level random-intercept model is the proportion of total residual variance which is
attributable to level 2, i.e., $\sigma_{u0}^2/(\sigma_{u0}^2 + \sigma_\varepsilon^2)$. In binary response of dependent variables, the level-1 residual follows a logistic distribution with variance $\pi^2/3 \approx 3.29$ (If a probit link is used, then $\varepsilon_{ij}$ follows a normal distribution with variance 1). So the ICC can be computed as $\sigma_{u0}^2/(\sigma_{u0}^2 + 3.29)$.

Two-level multinomial model is treated as three-level binary model. The third-level is the response type, which indicates the category of output variable. Each category is treated as a binary model. So for the multinomial model that has $n$ categories for the outcome variable, we have $n-1$ binary models, and separate ICCs for each model. The way to calculate ICC in each model is the same way as in two-level binary model.

### 4.3 Model validation

Pontius (2001) introduced the relative operating characteristic (ROC) as a measure to validate the land-cover change model. ROC is a statistic that derived from a series of two-by-two contingency tables corresponding to different simulated scenarios of predicted land cover change. Each contingency table is structured in Table 3 where “Reality” refers to the map of reality used for validation and “Model” refers to the calculated prediction map. The A, B, C, and D represent the number of grid cells in a map of reality versus a map of a modeled scenario.
To validate the model prediction, two basic datasets are required for ROC calculation. The first one (row) is a predicted probability map showing the likelihood of urban development. The second one (column) is the real land use change map in which each cell containing the binary value of change or non-change. To obtain the contingency tables, the probabilities at grid cells are first sorted in the descending order. Then, the probability map is reclassified to ten percentiles with the sorting. Thirdly, multiple sets of A, B, C, and D are determined by using the reclassified probability map and the real land use change map. In this step, eleven contingency tables are created by setting ten thresholds from 0% to 100%. The probability thresholds are set to be 0%, 10%, …, and 100%, with which reclassified grid cells can be categorized as “change predicted by the model” or “no change predicted by the model”. In each run, the binary information is compared against the reality to obtain a different set of A, B, C, and D constituting a contingency table. Two other important measures can be calculated from the contingency tables. True-positive (TP) ratio is the proportion of cells marked as “change” in both real and predicted land use maps. False-positive (FP) ratio is the proportion of cells marked as “change” in prediction maps but not changed in reality. Because the numbers of true-positive and false-positive cells correspond to A and B in the table, the true positive ratio
equals to $A / (A + C)$, and the false positive ratio equals to $B / (B + D)$. An example of ROC curve is drawn by plotting the eleven sets of TP and FP ratios (Figure 5). The area under the curve is defined as the ROC statistics. The formal expression of ROC statistics is shown in Eq. (12).

$$ROC = \sum_{i=1}^{n} [x_{i+1} - x_i][y_{i} + y_{i+1} - y_i / 2]$$  \hspace{1cm} (12)$$

In Eq. (12), $x_i$ is the false-positive ratio for scenario $i$, and $y_i$ is the true-false positive ratio for scenario $i$. ROC statistics ranges from 0 to 1, with greater values representing most accurate predictions. A plain random land use change model has the expected ROC statistic of 0.5, shown in Figure 5 as the dashed line.

![Figure 5: ROC Curve example](image_url)
Chapter 5

Results and Analysis

5.1 Case Study 1: Hamilton County, Ohio

5.1.1 Modeling result

A multilevel multinomial regression model is developed in this study to describe the urban development in periods of 1992-2000 and 2000-2010 in Hamilton County, Ohio. To utilize the available 2000 census data, DEM data, and 1992 and 2000 land use data, the model is constructed as Eq. (13-21). The level-1 model is expressed as Eq. (13-15).

\[
\ln[P_1/P_4] = \beta_{01} + \beta_{11}S_{ij} + \beta_{21}E_{ij} + \beta_{31}D_{ij} + \beta_{41}H_{ij} + \beta_{51}R_{ij} \\
\ln[P_2/P_4] = \beta_{02} + \beta_{12}S_{ij} + \beta_{22}E_{ij} + \beta_{32}D_{ij} + \beta_{42}H_{ij} + \beta_{52}R_{ij} \\
\ln[P_3/P_4] = \beta_{03} + \beta_{13}S_{ij} + \beta_{23}E_{ij} + \beta_{33}D_{ij} + \beta_{43}H_{ij} + \beta_{53}R_{ij}
\]

In Eq. (13-15), \(P_1, P_2, P_3,\) and \(P_4\) are the probabilities of four types of land use changes, namely open water to urban, forest to urban, pasture to urban, and no change. The forth type is used as the reference. \(S_{ij}, E_{ij}, D_{ij}, H_{ij},\) and \(R_{ij}\) represent the variables of slope \(\left(S_{ij}\right),\) elevation \(\left(E_{ij}\right),\) distance to downtown \(\left(D_{ij}\right),\) distance to highway \(\left(H_{ij}\right),\) and distance to railroad \(\left(R_{ij}\right)\) respectively. The subscript \(i\) denotes the index of the
variable at the grid level, and \( j \) denotes the index of the municipality region. The \( \beta \)-variables are coefficients of the level-1 model. Each of them is treated as dependent variables in the level-2 model.

Eq. (16) shows the level-2 model for variable \( \beta_{01} \sim \beta_{51} \), which is a random-intercept multilevel model. The Level-2 models for other \( \beta \)-variables can be established likewise. \( T_j \), \( V_j \), \( U_j \), and \( L_j \) represent the municipality-level variables road density \( (T_j) \), elevation variance \( (V_j) \), population density \( (U_j) \) and number of housing units \( (L_j) \) respectively. The \( g \)-variables are the coefficients of level-2 model. \( u_{001} \) is the residuals at municipality level.

\[
\beta_{01} = g_{001} + g_{011}T_j + g_{021}V_j + g_{031}U_j + g_{041}L_j + u_{001}
\]  

(16)

Table 4 shows the fixed effects of the model for three categories of urban development generated by the MLwiN statistic software. Coefficients and p-values for the level-1 and level-2 variables are included. It can be noticed that all level-2 variables show significant effects at the significance level of 0.05.
Table 4: Fixed effects of Hamilton County, Ohio

<table>
<thead>
<tr>
<th>Level</th>
<th>Variables</th>
<th>Open water to urban</th>
<th>Forest to urban</th>
<th>Pasture to urban</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>coeff</td>
<td>p</td>
<td>coeff</td>
</tr>
<tr>
<td>1</td>
<td>Slope</td>
<td>0.010</td>
<td>0.000</td>
<td>0.018</td>
</tr>
<tr>
<td></td>
<td>Elevation</td>
<td>-0.008</td>
<td>0.247</td>
<td>-0.001</td>
</tr>
<tr>
<td></td>
<td>Dis2Downtown (km)</td>
<td>0.040</td>
<td>0.000</td>
<td>0.077</td>
</tr>
<tr>
<td></td>
<td>Dis2Highway (km)</td>
<td>0.035</td>
<td>0.589</td>
<td>0.016</td>
</tr>
<tr>
<td></td>
<td>Dis2Rail (km)</td>
<td>-0.074</td>
<td>0.000</td>
<td>-0.062</td>
</tr>
<tr>
<td>2</td>
<td>Pop_den</td>
<td>0.039</td>
<td>0.000</td>
<td>0.015</td>
</tr>
<tr>
<td></td>
<td>HH_units</td>
<td>-0.134</td>
<td>0.000</td>
<td>-0.098</td>
</tr>
<tr>
<td></td>
<td>Ele_var</td>
<td>0.032</td>
<td>0.000</td>
<td>0.090</td>
</tr>
<tr>
<td></td>
<td>Road_den</td>
<td>-0.010</td>
<td>0.000</td>
<td>-0.003</td>
</tr>
</tbody>
</table>

(coeff: coefficient; p: p-value)

For the geographic factors of slope (Slope), the positive effect indicates that sloping areas are more likely to have land use types changed. This effect matches the positive coefficient of elevation variance (Ele_var) in level-2 model. The negative effect of elevation (Elevation) shows that higher elevated regions are less prone to new development. Although lower elevation area has better accessibility to public transportation and infrastructure, most of the suitable areas have already been urbanized. In order to expand the urban city, the higher sloping areas are chosen.

The road network also has significant effects on urban development. Both the distance to downtown (Dis2Downtown) and the distance to highways (Dis2Highway) have positive effects on the model. The further away an area is from the downtown or highway, the more chance the area will get developed during the studied period. This is
because the downtown of Cincinnati is already developed, but other municipalities are still building new residential and commercial zones. The census data also shows that the population of Cincinnati is decreasing over the last fifty years. New cities and townships in Hamilton County developed more rapidly than the central Cincinnati area, for example, Anderson Township. The same effect of distance to highway \((\text{Dis2Highway})\) also indicates the fact that the surrounding areas of highway are already been urbanized. Highway is the main transportation and traffic method in Hamilton County and it provides great accessibility to infrastructure consistently. This results in the effect that the closer area to highway has higher priority to get urbanized. The significant but negative effect of distance to railroad \((\text{Dis2Rail})\) shows that the closer to railroad, the higher probability of urbanization. Although the facility of railroad is often neglected, it is playing an increasingly important role in urban development.

Demographic factors including the population density \((\text{Pop\_den})\) and the number of housing units \((\text{HH\_units})\) show significant impact on urban development. The model result shows that urban development is more likely to happen in areas with larger population density, but with smaller number of housing units. Areas with greater population density have more demand of resources, which drives the urban development. However, in the areas with larger number of housing units, the development is mature and limited.
5.1.2 ICC

Intra-class correlation (ICC) is calculated for each category (Table 5). “Null model” is the empty model without any variable, “Level-1 model” is the model with only level-1 variables, and the “Full model” is the model shown in Eq. (13-16). In the category of open water to urban, 57% of the total variance is explained by the difference between municipalities before adding variables. In other word, 57% of total variability in urban development can be attributed to the municipality and 43% is within municipalities. After adding level-1 variables into the model, the percentage increased to 61.1%, showing that the 4.1% (61.1% - 57% = 4.1%) of the total variability is explained by the added level-1 variables. There are still considerable differences between municipalities that can potentially be explained by level-2 variables. This assumption is validated by the decrease of ICC from “Level-1 model” to “Full model” in which level-2 variables are added.

For the categories of forest to urban and pasture to urban, the ICCs have similar change after adding level-1 variables and level-2 variables. The ICCs increased after adding level-1 variables means part of the variability in urban development is explained. The high values of ICC suggest the application of multilevel approach to model urban development. The ICC decreased in the “Full model” validates the use of multilevel approach, showing that the added level-2 variables help explain the difference between municipalities in urban development.
Table 5: ICCs for Hamilton County, Ohio

<table>
<thead>
<tr>
<th></th>
<th>Open water to urban</th>
<th>Forest to Urban</th>
<th>Pasture to Urban</th>
</tr>
</thead>
<tbody>
<tr>
<td>Null model</td>
<td>0.570</td>
<td>0.232</td>
<td>0.755</td>
</tr>
<tr>
<td>Level-1 Model</td>
<td>0.611</td>
<td>0.296</td>
<td>0.807</td>
</tr>
<tr>
<td>Full model</td>
<td>0.484</td>
<td>0.216</td>
<td>0.412</td>
</tr>
</tbody>
</table>

5.1.3 Prediction map

The generated coefficients are used to calculate the urban change probability from 2000 to 2010 by incorporating the 2010 census data, DEM, and road network data into the model (Eq. 13–16). Each grid cell in the land use map has four possible changing patterns with probabilities of $P_1, P_2, P_3$ and $P_4$. The corresponding probability for specific pixel is selected based on the land use type of that pixel in 2000. For example, if the land use type for a pixel in 2000 is open water, the urban development probability is $P_1$. A prediction map showing the probabilities of the non-urban to urban use is shown in Figure 6 where darker colors indicate higher probability of urban development, while lighter colors mean lower chance of change. The cells that indicates the urban area is set as the lightest color.
To investigate the relationship between urbanization and road transportation, the counts and percentages of mutable cells within different buffer areas of downtown, railroad and highway are summarized in Table 6 and Table 7. A mutable cell is defined as one that is predicted be urbanized with the probability of $P_1$, $P_2$, or $P_3$ greater than 0.5, as well as the urban cells in 2001.
Table 6: Mutable cells in buffer area of downtown

<table>
<thead>
<tr>
<th>Distance (mile)</th>
<th>0.5</th>
<th>0.5-1</th>
<th>1-1.5</th>
<th>1.5-2</th>
<th>2-2.5</th>
<th>2.5-3</th>
<th>3-3.5</th>
<th>3.5-4</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Count</td>
<td>1189</td>
<td>1671</td>
<td>1814</td>
<td>2715</td>
<td>2795</td>
<td>2108</td>
<td>1885</td>
<td>1848</td>
<td>16025</td>
</tr>
<tr>
<td>Percentage (%)</td>
<td>0.94</td>
<td>1.33</td>
<td>1.45</td>
<td>2.17</td>
<td>2.23</td>
<td>1.68</td>
<td>1.50</td>
<td>1.47</td>
<td>12.77</td>
</tr>
</tbody>
</table>

Table 6 shows the count and percentage of mutable cells within different buffer distances from downtown. Only 12.77% of the mutable grids are located within 4 miles of the center of downtown Cincinnati. More urbanization activities are predicted to happen in areas further away from downtown. The number of mutable cells increases until the distance away from downtown reaches 2.5 miles. The phenomenon of distance decay shows up after that when the count of mutable cells decreased. This observation corresponds to the prediction map of Figure 6, in which only a small proportion of darker points are within a small range in downtown but largely distributed away from downtown.

Table 7 is the summary of the count and percentage of mutable cells within certain distance of highway and railroad. It can be seen that the proportion of the mutable grids located within the 3-mile radius of highways is very large. This indicates a positive correlation between the presence of a highway and urban development. 85.6% of mutable cells are located within 3 miles of railroad. Most mutable grids are observed in the 1-mile proximity of railroads. This result is consistent with multilevel model, in which the variable of distance to railroad shows significant negative correlation with urban development.
Table 7: Mutable cells in buffer area of highway and railroad

<table>
<thead>
<tr>
<th>Distance (mile)</th>
<th>Highway</th>
<th></th>
<th>Railroad</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Count</td>
<td>Percentage</td>
<td>Count</td>
<td>Percentage</td>
</tr>
<tr>
<td>0-1</td>
<td>79144</td>
<td>63.1%</td>
<td>78160</td>
<td>62.4%</td>
</tr>
<tr>
<td>1-2</td>
<td>23059</td>
<td>18.4%</td>
<td>17022</td>
<td>13.5%</td>
</tr>
<tr>
<td>2-3</td>
<td>14106</td>
<td>11.3%</td>
<td>12096</td>
<td>9.7%</td>
</tr>
<tr>
<td>Total</td>
<td>116309</td>
<td>92.8%</td>
<td>107278</td>
<td>85.6%</td>
</tr>
</tbody>
</table>

5.1.4 ROC statistic

To validate the prediction map, ROC statistic is computed using the method described in the previous sections. Figure 7 is the ROC curve for the prediction map of land use change in Hamilton County, Ohio during 2000-2010. Each point on the curve corresponds to a pair of TP and FP ratios calculated using a probability threshold. The ROC statistic is calculated as 0.816, showing good prediction power of the method. It suggests that the prediction result of urban development in Hamilton County is comparable with other fields (see Table 1 for reference).

In order to see the detail prediction for the three categories of urban development from open water, forest and pasture to urban, ROC statistic is also applied. Table 8 shows the percentage of cells and ROC result for each category of urban development. It can be noticed that the values are higher in the category of forest to urban and pasture to urban, which means the model predicts these two types of change better than the category that
changes from open water to urban. There is a need to improve the modeling of the urban development from open water. Since the percentage of the cells that changes from open water to urban is relative low, the less satisfied prediction result does not affect much of the whole prediction for Hamilton County, Ohio. However, there is still a need to improve the model to predict the urbanization of open water.

![ROC curve of prediction map of Hamilton County, Ohio](image)

**Figure 7:** ROC curve of prediction map of Hamilton County, Ohio

<table>
<thead>
<tr>
<th>Category</th>
<th>Percentage (%)</th>
<th>ROC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Open water to Urban</td>
<td>(1.2%)</td>
<td>0.687</td>
</tr>
<tr>
<td>Forest to Urban</td>
<td>(32.78%)</td>
<td>0.828</td>
</tr>
<tr>
<td>Pasture to Urban</td>
<td>(10.62%)</td>
<td>0.818</td>
</tr>
</tbody>
</table>

(All values are rounded to the nearest thousandth)

**Table 8:** ROC statistic for three categories of urban development in Hamilton County, Ohio

(%=percentage of cells for each category in prediction map)
5.2 Case Study 2: Cincinnati metropolitan area

5.2.1 Modeling result

Due to the large datasets and computing power of statistical software, the level-1 unit is set to census block of 2000. The level-2 unit stays same as municipality. The level-1 variables in this study are the same as Case Study 1, which is shown in Eq. (13-15). The level-2 variables are modified by adding a commuting factor, \( O_j \), which represents the municipality-level variable percentage of residents who commute out of their home to work. The equation is shown in Eq. (17). Other parameters as the same they are in Eq. (16).

\[
\beta_{01} = g_{001} + g_{011}T_j + g_{021}V_j + g_{031}U_j + g_{041}L_j + g_{051}O_j + u_{01}
\]  

(17)

Table 9 shows the fixed effects of the model for three categories of urban development from MLwiN statistic software. Coefficients and p-values are included for the level-1 and level-2 variables.
According to Table 9, the effects of the geographic, demographic and socio-economic variables in this case study of Cincinnati metropolitan area are consistent with those in the model of Hamilton County case study, except for the Dis2Downtown variable in the Category 1 (open water to urban). The negative effect of Dis2Downtown variable in Category 1 (open water to urban) indicates that the closer to downtown, the larger probability for the urbanization of open water. The dark grids in Figure 8 show the locations of where this type of urban development happens from Year 1992 to Year 2010. It can be noticed that a proportion of grids are located close to the Cincinnati downtown, closing to the Licking River in southern Cincinnati. The reason behind the fact is that there are more intensive human activities that motivate urban development. Similarly, the
spots close to William H Harsha Lake and Caesar Creek Lakes are changed into development due to human activities. Since the urbanization of water is restricted to specific environment, like creek or river, it is not sufficient to adjust the explanation of urbanized open water locations to their geographical position towards downtown. It is well-explained by adding more demographic variables into the model.

Figure 8: Change map of Open water to Urban
From the corresponding rows for geographical factors of slope *(Slope)*, elevation *(Elevation)* and elevation variance *(Ele_var)* in Table 9, urban development has higher probability to happen in sloping but lower elevated area. This effect is similar to that in the Hamilton County case study. Parameter estimates of demographical variables of population density *(Pop_den)* and *(HH_Units)* show that area with larger population density and less housing units are more likely to be urbanized. The road network has significant effects. The variable of distance to downtown *(Dis2Downtown)* in the categories of “Forest to urban” and “Pasture to urban” shows that these two kind of urban development is less likely to happen in locations closer to downtown. The variables of distance to highway *(Dis2highway)* and distance to railroad *(Dis2Rail)* also represent significant effects on Cincinnati metropolitan area’s urban development.

The newly added commuting factor *(Commute_out)* shows significant negative effect in the models. Specifically, lower ratio of workers commuting out of residence means more working opportunities and better job accessibility in the area, which in turn drives the urban development.

### 5.2.2 ICC

Intra-class correlations (ICC) for the three categories of urban development are also shown in Table 10. It compares the “Null model” that without any variable, the “Level-1 model” with only level-1 variables and “Full model” with all associated variables. In the category of open water to urban, 70.8% of the total variance is explained
by the difference between municipalities before adding variables. In other word, 70.8% of total variability in urban development can be attributed to the municipality and 29.1% is within municipalities. After adding level-1 variables into the model, the percentage increased to 71.9%, showing that the 1.1% (71.9%−70.8%=1.1%) of the total variability is explained by the added level-1 variables. The value of ICC decreased from “Level-1 model” to “Full model” with the level-2 variables added showing that the level-2 variables help explain the differences between municipalities in urban development.

For the categories of forest to urban and pasture to urban, the ICCs also increased after adding level-1 variables means part of the variability in urban development is explained. The high values of ICC suggest the application of multilevel approach to model urban development. The decrease of ICC in the “Full model” validates the use of multilevel approach, showing that the added level-2 variables help explain the difference between municipalities in urban development.

**Table 10: ICCs for Cincinnati metropolitan area**

<table>
<thead>
<tr>
<th></th>
<th>Open water to Urban</th>
<th>Forest to Urban</th>
<th>Pasture to Urban</th>
</tr>
</thead>
<tbody>
<tr>
<td>Null model</td>
<td>0.708</td>
<td>0.188</td>
<td>0.277</td>
</tr>
<tr>
<td>Level-1 model</td>
<td>0.719</td>
<td>0.194</td>
<td>0.443</td>
</tr>
<tr>
<td>Full model</td>
<td>0.357</td>
<td>0.089</td>
<td>0.261</td>
</tr>
</tbody>
</table>
5.2.3 Prediction map

Figure 9 is the prediction map of urban development from Year 2001 to Year 2010 in Cincinnati metropolitan area. Since this research is based on the assumption that developed area would not transfer back to non-urban area, the grids that represent the developed area are set as null. The grids showing darker color indicate higher probability of urban development and vice versa. The pattern of urban development shows an obvious trend of keeping away from downtown area.

Figure 9: Prediction map of Cincinnati metropolitan area
5.3.4 ROC statistic

The ROC statistic is computed using the same method as in the case study of Hamilton County, Ohio. Figure 10 is the new ROC curve for the prediction map of land use change in Cincinnati metropolitan area during 2000-2010. The ROC statistic is calculated as 0.859, suggesting that the prediction map is valid.

In order to see the detail prediction for the three categories of urban development from open water, forest and pasture to urban, ROC statistic is also applied. Table 11 shows the percentage of cells and ROC result for each category of urban development. It can be noticed that the values are also higher in the category of forest to urban and pasture to urban, which indicate better prediction than the change from open water to urban. The prediction for Cincinnati metropolitan area would not be much affected by it since the percentage of the cells that changes from open water to urban is relative low. However, there is still a need to improve the model to predict the urbanization of open water.
Figure 10: ROC curve of prediction map of Cincinnati metropolitan area

Table 11: ROC statistic for three categories of urban development in Cincinnati metropolitan area

<table>
<thead>
<tr>
<th>Category</th>
<th>ROC</th>
<th>ROC</th>
<th>ROC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Open water to urban</td>
<td>0.584</td>
<td>0.827</td>
<td>0.854</td>
</tr>
<tr>
<td>(1.14%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Forest to urban</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(39.17%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pasture to urban</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(35.40%)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Chapter 6

Discussion and Conclusion

Multilevel multinomial regression was applied to identify and improve our understanding of the urban development in Hamilton County, Ohio and Cincinnati metropolitan area driven by geographical, demographic, transportation and commuting factors. The models were also used to make predictions of future urbanization. The ROC statistics suggest that the models predict the urban development reasonably well.

The municipality-level variables included are road density, elevation variance, population density, number of housing units, and the percentage of residents who commute out of their home to work. Parameter estimation shows that elevation variance and population density at the municipality level are positively correlated with all three categories of urban development. Number of housing units and road density are negatively related with urban development, indicating that the urbanization is less prone to happen in areas with larger number of housing units and greater road density. The commuting variable added to the second case study, the percentage of residents who commute out of their home to work, shows that lower ratio of workers commuting out of residence would drive urban development.

The level-1 variables include slope, elevation, distance to downtown, distance to highway, and distance to railroad. From the parameter estimation, slope, distance to downtown and distance to highway have positive effect on urban development, while the elevation and distance to railroad have negative effect on that. The only exception is the
negative effect of distance to downtown in the case study of Cincinnati metropolitan area. The urbanization category of open water to urban is modeled to happen closer to downtown. The further analysis of the prediction map suggests that most of the urbanization area is far away from downtown, but close to highways and railroads. The result that more than 85% of potential urbanization areas are within the 3-mile buffer areas of highways and railroads shows road networks are very important for urbanization.

This study has also shown that much of the heterogeneity can be properly accounted for with a hierarchical structure. This paper has demonstrated that contextual variability in cell-based (or block-based) change can be quantified and statistically tested in a multi-level multinomial logit model. The multilevel structure is particularly desirable in studying the contextual effect of the built environment on urban development. It is found that cell-based (or block-based) land use change (e.g. urban development) varies considerably across different municipalities. However, significant variability among municipalities is still present, suggesting other variables in municipal level urban development.

The topic of multilevel context is rarely examined from the perspective of land use change and urban development. This omission might lead to critical errors in policy. Since the built environment of municipality has been shown to shape micro-level (cell or block) urban development, municipal land use/transport planning efforts must work in co-ordination with the regional planning effort to provide more coherent land use/transport planning at both local and regional levels. The impact on the urban development will come from refocusing the policy on the transportation factors that
determine land use changes, including public transit support policies, smart urban growth implementations, and local policies that encourage mixed land use. In locations where land use changes urban developments, policymakers pay careful attention to the policies that may cause significant shifts in the urban development by land use changes. The study findings have important implications for integrated land use, transport planning, and urban development. Urban and regional planners need to consider contextually when viewing land-use policy as a way to manage urban development.

It is important to point out that the study findings must be interpreted within the context of the study limitations, due to the complex relationships between the built environment and urban development. A set of available census block variables are assembled for analysis. However, it is not intended to imply that the assembled variables have fully characterized the built environment of the municipal level. There may be variability due to micro level factors that are not captured at the census block level. Due to the limitation of the currently available data, future investigation on this subject is necessary. Other measures such as, job accessibility, journey to work mode, regional transport forecasting and municipal level land use policy would be of great interest. Further research effort is needed.
Bibliography


