Calibration of an Artificial Neural Network for Predicting Development in Montgomery County, Virginia: 1992-2001

Travis F. Thekkudan

Thesis submitted to the faculty of the Virginia Polytechnic Institute and State University in partial fulfillment of the requirements for the degree of

Master of Science
In
Geography

Dr. James B. Campbell, Chair
Dr. Laurence W. Carstensen Jr.
Peter Sforza
John D. Boyer

June 11, 2008
Blacksburg, Virginia

Keywords: Artificial Neural Network, Land Use, Modeling, Montgomery County, Urban

Calibration of an Artificial Neural Network for Predicting Development in Montgomery County, Virginia: 1992-2001

Travis F. Thekkudan

James B. Campbell, Chair
Department of Geography

ABSTRACT

This study evaluates the effectiveness of an artificial neural network (ANN) to predict locations of urban change at a countywide level by testing various calibrations of the Land Transformation Model (LTM). It utilizes the Stuttgart Neural Network Simulator (SNNS), a common medium through which ANNs run a back-propagation algorithm, to execute neural net training.

This research explores the dynamics of socioeconomic and biophysical variables (derived from the 1990 Comprehensive Plan) and how they affect model calibration for Montgomery County, Virginia. Using NLCD Retrofit Land Use data for 1992 and 2001 as base layers for urban change, we assess the sensitivity of the model with policy-influenced variables from data layers representing road accessibility, proximity to urban lands, distance from urban expansion areas, slopes, and soils. Aerial imagery from 1991 and 2002 was used to visually assess changes at site-specific locations.

Results show a percent correct metric (PCM) of 32.843% and a Kappa value of 0.319. A relative operating characteristic (ROC) value of 0.660 showed that the model predicted locations of change better than chance (0.50). It performs consistently when compared to PCMs from a logistic regression model, 31.752%, and LTMs run in the absence of each driving variable ranging 27.971% – 33.494%. These figures are similar to results from other land use and land cover change (LUCC) studies sharing comparable landscape characteristics. Prediction maps resulting from LTM forecasts driven by the six variables tested provide a satisfactory means for forecasting change inside of dense urban areas and urban fringes for countywide urban planning.

All images not cited within this document were created by the author between January 1, 2008 and July 11, 2008.

Acknowledgements

I would like to thank all the faculty members of the geography department for helping me develop the skills necessary to carry out this research effort. I owe each committee member my gratitude for their individual contributions to my work specifically, John Boyer, for providing motivation despite my early frustrations and helping me prepare my final presentation, Peter Sforza for his willingness to invest material resources and data expertise in me, Dr. Carstensen for pushing me to improve the quality of my methods and teaching me GIS, and of course my advisor, Dr. Campbell, whose experience was most valuable in guiding my research and helping me create a product I could be proud of. I also want to thank Ed Brooks from Newman Library for putting up with my constantly-changing data requests.

Dr. Dymond and the staff of the Center for Geospatial Information Technology (CGIT) at Virginia Tech deserve special thanks for offering me valuable work experience, funding much of my graduate studies, and providing a friendly environment where I could always find assistance.

Finally, I must thank my family who instilled in me the importance of education and provided me with so many opportunities that achieving academic success was almost a certainty. Their love, along with all the support shown by my friends and fellow graduate students, has definitely been my strength throughout this challenging process and for that I am forever grateful.

Table of Contents

List of Figures

List of Tables

Chapter 1: Introduction

Introduction

Urban and suburban regions in the United States began to experience uncoordinated and scattered development, known as urban sprawl, near cities and towns after World War II. Such development patterns heavily burden local governments with high financial costs due to the lower densities at which they must provide services. Today, local governments face controversy as they attempt to deal with the consequences of dispersed infrastructures, growth, and land use changes.

To prevent undesirable urban growth patterns and the economic downfalls associated with them, in 1975 the Virginia Code required municipalities to manage growth and guide new development using comprehensive plans. These plans allow for reduced costs for infrastructure improvements and development. Comprehensive plans are not always effective, however, because planners often cannot foresee those areas or land uses which are likely to be catalysts of land use fragmentation because of the complexity of factors driving urban change. Despite the importance of "visioning" within the planning discipline, few planning departments use predictive modeling. Urban growth modeling can help policy-makers weigh the effects of different planning scenarios (McCormick et al., 2001).

Analysis of land use and land cover change (LUCC) uses various modeling techniques to represent the complex nature of land use systems to evaluate spatial change, temporal change, and the sensitivity of selected socioeconomic and biophysical drivers of change (Veldkamp & Lambin, 2001; Hu & Lo, 2007). Such analyses promote understanding of the factors driving land use change so that better predictions can be made about the ecosystems around them (Berling-Wolff & Wu, 2004). Models can also examine the effects of existing land change trends on future landscapes by assessing spatial patterns of land cover change (Brown et al., 2002).

Today, urban growth modeling research depends upon continued assessment of variations in model structure, model drivers, data quality, and resolution (Allen & Lu, 2003). One type of urban growth model, the Land Transformation Model (LTM), offers methods for assessing these needs for further analysis with a flexible parameterization process using Geographic Information Systems (GIS) technologies and machine learning, specifically artificial neural network (ANN) algorithms (Pijanowski et al., 2002a). Although several deficiencies currently exist in the validation of such models, relatively new techniques such as Relative Operating Characteristics (ROC) offer a more standardized method for comparing LUCC models' ability to predict spatial change (Pontius & Batchu, 2003; Pontius et al., 2004; Hu & Lo, 2007). By testing various predictive modeling techniques, such as ANNs, LUCC modeling can work towards building a readily agreed upon and robust model for the planning community.

Statement of Purpose

The ability of modelers to interpret spatial change and development patterns has improved over the years as spatial datasets become readily available along with improved remotely sensed data (Bian & Walsh, 2002). Planners and decision-makers have the opportunity to keep pace with advancements in ecological modeling, especially in the context of urban dynamics, by utilizing the latest technologies and newest spatial datasets (Allen & Lu, 2003; McCormick et al., 2001).

This study evaluates the effectiveness of an artificial neural network to predict the locations of urban change at a countywide level by testing various calibrations of the LTM.

Analysis provides further insight for county planners into the specific dynamics of urban change in Montgomery County, Virginia, and offers an opportunity to see the utility of urban forecasting as a tool in the comprehensive planning process by applying GIS and predictive modeling techniques in the context of urban planning. This research contributes to the growing body of literature on neural network simulation as it integrates commonly used data inputs with policy-oriented land use drivers. It offers unique perspectives on land use change through various LUCC validation techniques, including the use of ROC, rigorous visual assessment, and comparison with other popular LUCC models.

Chapter 2: Literature Review

Introduction

The practice of land use and land cover change modeling (LUCC) is rooted in understanding the interaction of socioeconomic and biophysical forces factors influencing urban dynamics and establishing their connections with technology to accurately represent landscape changes. While comprehensive planning provides the context for these factors influencing development through policy in Virginia, the challenge of fitting these policies into an appropriate LUCC model still remains. The Land Transformation Model (LTM), a type of artificial neural network (ANN) utilizing machine learning, can evaluate spatially-explicit land use changes within a policy-oriented context at the county level in Virginia.

Planning

The state of Virginia requires municipalities to manage growth and guide new development by using comprehensive plans, as outlined in section 15.2223 of the Virginia Code stating:

> *The local planning commission shall prepare and recommend a comprehensive plan for the physical development of the territory within its jurisdiction and every governing body shall adopt a comprehensive plan for the territory under its jurisdiction.*
>
> *In the preparation of a comprehensive plan, the commission shall make careful and comprehensive surveys and studies of the existing conditions and trends of growth, and of the probable future requirements of its territory and inhabitants. The comprehensive plan shall be made with the purpose of guiding and accomplishing a coordinated, adjusted and harmonious development of the territory which will, in accordance with present and probable future needs and resources, best promote the health, safety, morals, order, convenience, prosperity and general welfare of the inhabitants, including the elderly and persons with disabilities.*
>
> *The comprehensive plan shall be general in nature, in that it shall designate the general or approximate location, character, and extent of each feature, including any road*

> *improvement and any transportation improvement, shown on the plan and shall indicate where existing lands or facilities are proposed to be extended, widened, removed, relocated, vacated, narrowed, abandoned, or changed in use as the case may be.*

The role of comprehensive planning in Montgomery County, Virginia has evolved over the last 30 years, yet has retained the goal of maintaining planned and orderly development of the county. The function of the plan is to state the county's goals, guide decision-making, consider long-term consequences based on current trends, promote civic communication, education, and in some cases stand as legal documentation. These functions are specifically met by establishing goals, putting forth policies to achieve those goals, and setting strategies to meet each policy. Montgomery's 1990 Comprehensive Plan provides a long-range perspective to guide the county for 5-10 years, specifically addressing environment, water and sewer, transportation, housing, economy, and community facilities.

Comprehensive plans reduce costs to infrastructure improvements and make development more predictable because they often recommend decreases in the distances public services must be stretched and layout the optimal development patterns for the planned area. Although the practice of comprehensive planning varies from community to community, all localities begin the planning process with a research phase in which data is gathered and future trends are forecast (Levy, 2006).

Planning actions form one of the main influences upon land use change through zoning, conservation easements, and in Montgomery County's case, Agricultural and Forestal Districts (AFDs) (Theobald & Hobbs, 1998). Such decisions influence development patterns by steering construction to new locations, guiding land use density, and altering development patterns (Theobald & Hobbs, 1998).

LUCC modeling research has the potential to aid planners and policymakers as a tool in forecasting future land use changes and managing urban growth. Although there is growing literature addressing the field of growth modeling, there is little published research devoted to the application of predictive modeling in policy development. In their 2003 study, Allen and Lu found that most innovative modeling techniques are confined to academic research and that predictive modeling is only just beginning in the applied urban studies field. They call for further analysis of interactions of planning and the latest predictive modeling techniques in practical applications.

Modeling Techniques

Policy-oriented Models

Lowry (1965) distinguished between descriptive, predictive, and planning models. Descriptive models attempt to replicate relevant features from an existing land use change for the purpose of formulating or justifying a theory. Predictive models forecast the consequences of implementing a theory. Planning models use predictive techniques and evaluate the results in the context of satisfying planned goals.

Good policy models must possess clarity, robustness, realistic data needs, appropriate spatio-temporal detail, and sufficient influential policy variables to evaluate policy scenarios with significant confidence (Lee, 1973). While deterministic models are commonly meaningful and operational, they often generalize reality because they fail to use a sufficient number of components and variables (Sui, 1997). Allen and Lu (2003) recognize the complexity of modeling urban land use systems and concede that it is impossible to encompass all of the elements of both natural and human elements into a single model. They call for methods that can incorporate the knowledge of planners and developers with tangible variables to create more realistic predictions using scenario-based simulations.

The main focus of policy-oriented models, however, is not to produce definitive results – rather, results that encourage thought and discussion among stakeholders and decision-makers (Argawal et al., 2002; Barredo et al., 2003). Scientists create these models in the hope that results will be considered during the policy process, fitting them perfectly into the research and forecasting phase of the comprehensive planning process. In his comparison of popular LUCC models, Pontius et al. (2008) noted the necessity for communicating variations in modeling techniques in ways that are scientifically detailed, yet general in nature and intellectually accessible.

Spatially Explicit Models

Spatially explicit models attempt to determine temporal relationships of a set of land classes, their spatial locations, or both temporal and spatial changes. They can be calibrated using historical land use data and/or "drivers" hypothesized from influential factors. Spatially explicit models usually produce likelihood maps that display patterns of probabilities for conversion to a certain land use class, or a map predicting land use change by location at various temporal scales.

There are two primary spatially explicit land use change models: regression and spatial-transition based (Theobald & Hobbs, 1998). Regression models attempt to determine functional relationships between variables that predict locations of spatial change using statistical regression. They assume that development patterns are a result of location factors such as distance decay. Pijanowski et al. (2002a) focused on the application of regression type models citing the ease with which they can be calculated. Landis (1995) and Allen and Lu (2003) also highlighted use of statistical methods, mainly logistic regression, to identify which variables best represent urban reality. Other LUCC modeling approaches (Hilferink & Reitveld, 1999; Pontius et al., 2001; Verburg et al., 2002; Engelen et al., 2003; McConnell et al., 2004) offer the option of using a statistical regression.

Logistic regression can represent the nonlinear nature of urban growth. Currently, IDRISI Andes 15.0 Software's LOGISTICREG module performs a binomial logistic regression, in which a binary dependent variable follows a logistic curve estimated as;

$$P(y=1|X)=\frac{\exp(\sum BX)}{1+\exp(\sum BX)} \qquad \text{[Eq. 2.1]}$$

where P represents the probability of the dependent variable being one, X stands for independent variables, and B signifies the estimated parameters.

Spatial transition-based models combine aspatial Markov-chain techniques with stochastic cellular automata (CA) approaches (Theobold & Hobbs, 1998). Markov-chain models calculate the probability of transition where land use at time $t+1$ is based exclusively on prior land use at time t (Kocabas & Dragicevic, 2006). Stochastic CA models base transition probabilities on values of neighboring cells. CA has formed the core of integrated modeling approaches, because they use data with high spatial resolution, simple rules, and dynamic behavior (Theobald & Hobbs, 1998; Barredo et al., 2003). Standard integrated CA models include Clarke et al., 1997; Silva & Clarke, 2002; and Engelen et al., 2003.

Regression and spatial-transition based models perform equally well when predicting the spatial location and proportions of discrete data (Theobald & Hobbs, 1998), although regressions often result in highly fragmented patterns, while transition models are known to produce more realistic landscape patterns (Turner, 1988).

Artificial Neural Networks

Artificial neural networks (ANNs) are software tools designed as a system of nodes and connectors which find relationships between given sets of inputs and outputs. These networks are intended to mimic the way humans solve problems through a series of repeated observations between neurons and synapses within the brain.

The first ANN was created by Rosenblatt (1958) when he introduced the single node "perceptron." This single node received outputs directly from weighted inputs and was useful when working with data operating as linear functions (Pijanowski et al., 2002a). The multi-layer perceptron (MLP), however, is the most commonly used neural net for machine learning (Skapura, 1996). The MLP is comprised of three layers: an input layer, any number of hidden layers, and an output layer (Figure 2.1). Its ability to interconnect between different nodes through the hidden layer allows it to evaluate data with non-linear functions.

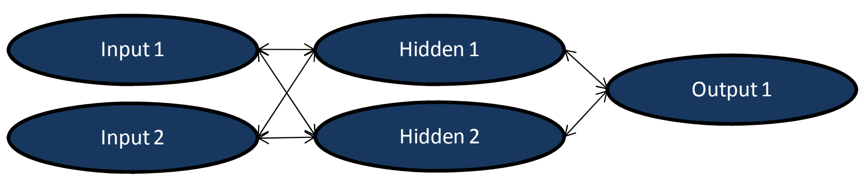


Figure 2.1 MLP showing one layer for input, hidden, and output nodes.

The MLP uses algorithms to calculate weights for input values at each node as the data is introduced to the network in a "feed forward" manner (Pijanowski et al., 2005). Output values from each node are calculated as functions of inputs, known as activation functions. These functions may have a coefficient, called a bias, used in some cases. The most common algorithm used for building a neural network is the back-propagation algorithm (Skapura, 1996).

In a network utilizing this algorithm, a Back-Propagation Network (BPN), weights are randomly assigned to the initial input connectors, propagated forward through the network, and calculated for comparison against some known output value. Differences between all the known and propagated output values for every observation are calculated as a mean squared error. Each of the values is propagated back through the network using the *generalized delta rule*, distributing the total error according to Equation 2.2.

$$\Delta w = \eta (t-u)\frac{x}{x^2} \qquad \text{[Eq. 2.2]}$$

Here network weights (Δw) are modified from this equation, where η is the learning rate, t-u is the difference between actual outputs and expected outputs, and x is the value of an input at a certain node (Pijanowski et al., 2002b).

Each time data are fed forward then back-propagated iteratively through the network (known as a cycle) errors are reduced (Pijanowski et al., 2005). This phase of presenting data in repeated cycles to the neural net is known as training. A final phase known as testing can be achieved by running the neural network using the trained network files without known output data to evaluate the effectiveness of the network's pattern development.

Although applied research on ANNs has been more prevalent in other fields, there has been a movement in the GIS and environmental research communities to study applications of machine learning for environmental purposes. In the LUCC realm, most of ANN application has been for preparation of remote sensing data. Neural networks are often used to classify land use from imagery by finding relationships between land classes and digital pixel values in an iterative process to reduce error until classification accuracy is maximized (Campbell, 2007). Of the popular LUCC models tested to date, Pijanowski's et al. Land Transformation Model (2001, 2002a, and 2005) is the only regression model that uses a neural network while other hybrid models attempt to combine ANNs with CA spatial-transition based models (Pontius et al., 2008; Yeh & Li, 2003). The SLEUTH model also uses machine learning, however, via the Monte Carlo method to carry out multiple iterations (Silva & Clarke, 2002).

Applications of neural networks offer considerable benefits to researchers. They are simple from a mathematical standpoint, because data sets are randomly weighted for input nodes, and machine learning will derive relationships between input and output layers. Neural networks do not require data sets to be perfect, because they are inherently built for error minimization. Fischer and Gopal (1994) and Yeh and Li (2003) found that ANNs perform better than conventional models because they are better suited to adjust to uncertainties in spatial data.

ANNs are best known for their ability to generalize among large sets of data. In a 2001 study, Pijanowski et al., utilized a neural network within their Land Transformation Model (LTM) to successfully transfer training patterns from one region, the Detroit Metropolitan Area (DMA), to model growth in another similar area, the Twin Cities Metropolitan Area (TCMA). Although the test did not work as well when reversed (TCMA to DMA), their model shows evidence of the neural nets' ability to generalize growth patterns for like urban areas of similar scale.

Challenges of employing neural nets include a lack of user control, lengthy processing times, and repeated trial and error. Once initial inputs are set, neural net training produces network files that are not always visible to the user until after all training is complete. Without access to these files, even a skilled user may not be able to determine the number of necessary cycles to run prior to the testing phase. Despite the speed of current processors, training of neural nets can be very time-consuming, especially when working with large networks and datasets. The practice of network simulation depends upon continued trial and error. Without familiarity with the specifically proposed application, a user may have to continually adjust the size of the network, and the number or type of inputs before the network begins to show signs of improvement.

Land Transformation Model

The Land Transformation Model forecasts changes in land use by integrating geographic information systems (GIS), artificial neural networks, geostatistics, and remote sensing technologies. The LTM can be used to study spatial interaction between various biophysical and socioeconomic drivers which have historically caused land use change. By determining the nature of the spatial interactions between these drivers through neural network technologies and machine learning, the model can simulate local land use changes.

Pijanowski et al. (2002a) used four steps to run the LTM. First, he processed and coded land uses and features into base layers through GIS to create a grid of predictor cells. Each driver for change was then quantified into a grid displaying the spatial effects of predictor cells on land use change by one of four transition rules. A "neighborhoods and densities" rule looks at effects of surrounding cells on a single cell in the middle, and their tendencies to change to other land uses. "Patch sizes" relate the values of all cells in defined zones (like parcels) to the likelihood of change. "Site specific characteristics" assign values to each cell based on its biophysical or socioeconomic characteristics. The "distance from the location of a predictor variable" rule measures Euclidean distance from the nearest predictor variable. All cells where change could not occur were included in an exclusionary layer.

In their 2002 paper, "Using neural networks and GIS to forecast land use changes: a Land Transformation Model," Pijanowski et al. used an artificial neural network to integrate predictor variables to create a map of transition probability values. The values are the combined result of executing the ANN using the back-propagation algorithm.

Pijanowski et al. (2002a) proposes a "principle index driver" or PID (Eq. 2.3) to calculate the amount of land expected to change:

$$U(t) = (\frac{dP}{dt}) * A(t) \qquad \text{[Eq. 2.3]}$$

This equation calculates the expected amount of new urban land U created, in the specified time interval t, where dP/dt is the number of new population in any area over a certain time interval and where A is the per capita requirements for urban land (Pijanowski et al., 2002a). By controlling the quantity of change, the user can set locations of spatial change as the dependent variable and forecast where land use changes will take place in the future using patterns derived from ANNs.

Chapter 3: Methods

Study Area

Montgomery County, Virginia is a 1,005 square kilometer area located in the mountains of southwestern Virginia (Figure 3.1). The county experienced urban growth in both urban and rural areas during the 1990s while attempting to follow policy-oriented development patterns laid out by its 1990 Comprehensive Plan.

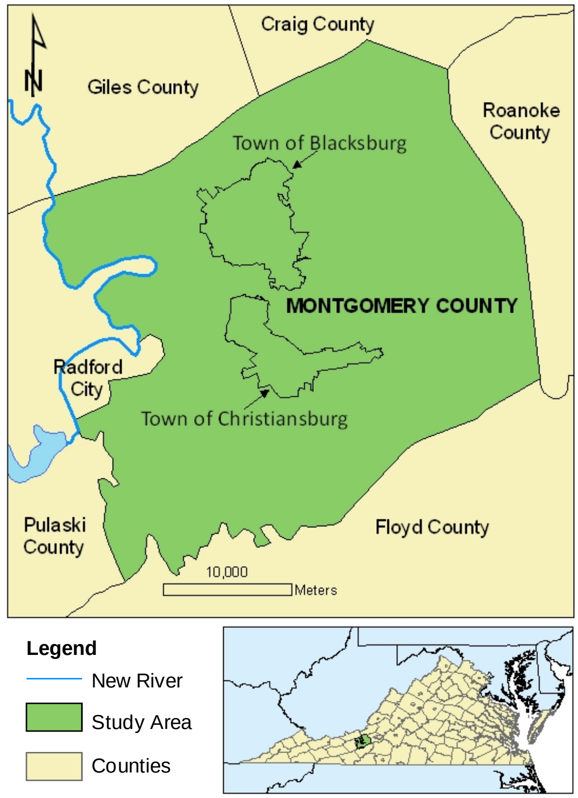


Figure 3.1 Montgomery County, Virginia and its surrounding area.

Urbanized regions are represented by, Blacksburg and Christiansburg, the two principal towns in the county. Blacksburg is home to Virginia Tech, the county's biggest employer; while Christiansburg is home to a majority of the area's commercial/retail industry. A third area, the Radford Army Ammunition Plant, has declined in significance since the mid-1980s, yet still lies as the third largest contributor of urban land area in the county.

The towns of Blacksburg and Christiansburg, combined with the City of Radford, form a socially and economically cohesive area that earned designation as a metropolitan statistical area (MSA) in the year 2003. According to U.S. Census Bureau, population estimates during the study time-period grew from 75,600 in 1992 to 84,300 in 2001.

The majority of the county's land use is devoted to forests to the north and east while agricultural lands occupy much of the county's southern and western portions. Other notable land uses include smaller urban areas “villages” located throughout the county. During the 1990’s, these satellite locations around both towns were considered urban expansion areas because of their geographical and community-defined points of centrality among the areas they surrounded. Planners designated these areas as expansion areas in the expectation that development would occur there based on their low slopes, proximity to major roads, and designation as priority areas to receive public water and sewer.

LTM

The Land Transformation Model (LTM) is a simulation program that works with distinct file formats. Each of the drivers and land use grids must be converted from their original GIS formats to ASCII files for the model to operate upon them (Pijanowski et al., 2002b). The LTM uses text files outlining network parameters and executable files allowing the network to make computations with batch processes. Currently, the publicly available download of the LTM software does not provide a graphical user interface (GUI), so users must process executable files from the Windows command prompt. The model (and video-tutorial) was obtained via download from Purdue University’s LTM website under a GNU general public license agreement.

The model can be used to train, test, and forecast land use change. It utilizes the Stuttgart Neural Network Simulator (SNNS), a common medium through which ANNs run a back-propagation algorithm, to execute neural net training (Imrie et al., 2000). The number of input, hidden, and output nodes must first be entered to build the initial neural net. Next, input parameters: the land use base, final land use layer, exclusionary layer, drivers, and any sampling boxes must be identified to produce network files in 100-cycle epochs. Testing begins once the user believes error has been reduced to sufficient levels. This phase is run by using pattern files the network created from training for each of the network files to produce residual files (transition probability maps) and prediction maps. A resulting text file also provides statistics detailing each epoch's performance in terms of a percent correct metric (PCM), Kappa values, and the numbers of cells falling into various categories of error (Table 3.1).

Table 3.1 Explanation of LTM prediction categories.

Symbol	Classification	Description
TN	True Negative	No change in both real & simulated
TP	True Positive	Change in both real & simulated
FN	False Negative	Change in real but not simulated
FP	False Positive	Change in simulated but not in real
SN	Simulated Negative	Total no change in simulated
SP	Simulated Positive	Total change in simulated
RN	Real Negative	Total no change in real
RP	Real Positive	Total change in real
GT	Grand Total	Total cells transitioned

These metrics can be used to determine the performance of the model to forecast future growth correctly. The focus of this research, however, is to explore the dynamics of each driver and how they affect model calibration.

Data

The following sub-sections provide the justification, source, and process through which each driver was created in GIS, as well as, maps displaying their distribution throughout Montgomery County. Each dataset was created using ESRI ArcGIS software to fit to Montgomery County's publicly available boundary shapefile. All of the following data were created as 1290 by 1423 (rows by columns) raster grids with a 30-meter cell size before conversion to ASCII format.

Land Use

All land classified as "built-up" or "urban" by Anderson classification level I was used to define urban areas (Anderson et al., 1976). These sites include all land covered by built structures. Urban land also includes any nonconforming use located in the vicinity of structures if the area has relatively intensive residential, commercial, and industrial uses, or provides transportation, communication, or utilities as defined by Anderson classification level II. Areas such as golf courses, parks, cemeteries, and other structure-free lands with intensive usage are also considered urban areas.

Land use for our area was derived from the National Land Cover Database (NLCD) 1992/2001 Retrofit Land Cover Change Product Multi-zone data downloaded from the Multi-Resolution Land Characteristics Consortium (MRLC) website. Although this product was designed for use with broad-scale projects (national, state, or regional) rather than county-level evaluation, it provides the most accurate publically available depiction of urban land use at Anderson level I. The limitations of this dataset are discussed further in a subsequent section.

Since Montgomery County falls on the border of the NLCD's download webpage (regions 13 & 14), a series of raster calculations were performed to merge the two files within the county boundaries. Unchanged cells where urban land uses persisted over the nine-year period were classified as urban from 1992. All cells where non-urban lands converted to urban lands were classified as 2001 urban cells. Non-urban areas were reclassified with a value of zero and urban areas a value of one resulting in the binary 1992 land use base and 2001 land use final datasets combined in Figure 3.2, Map 1.

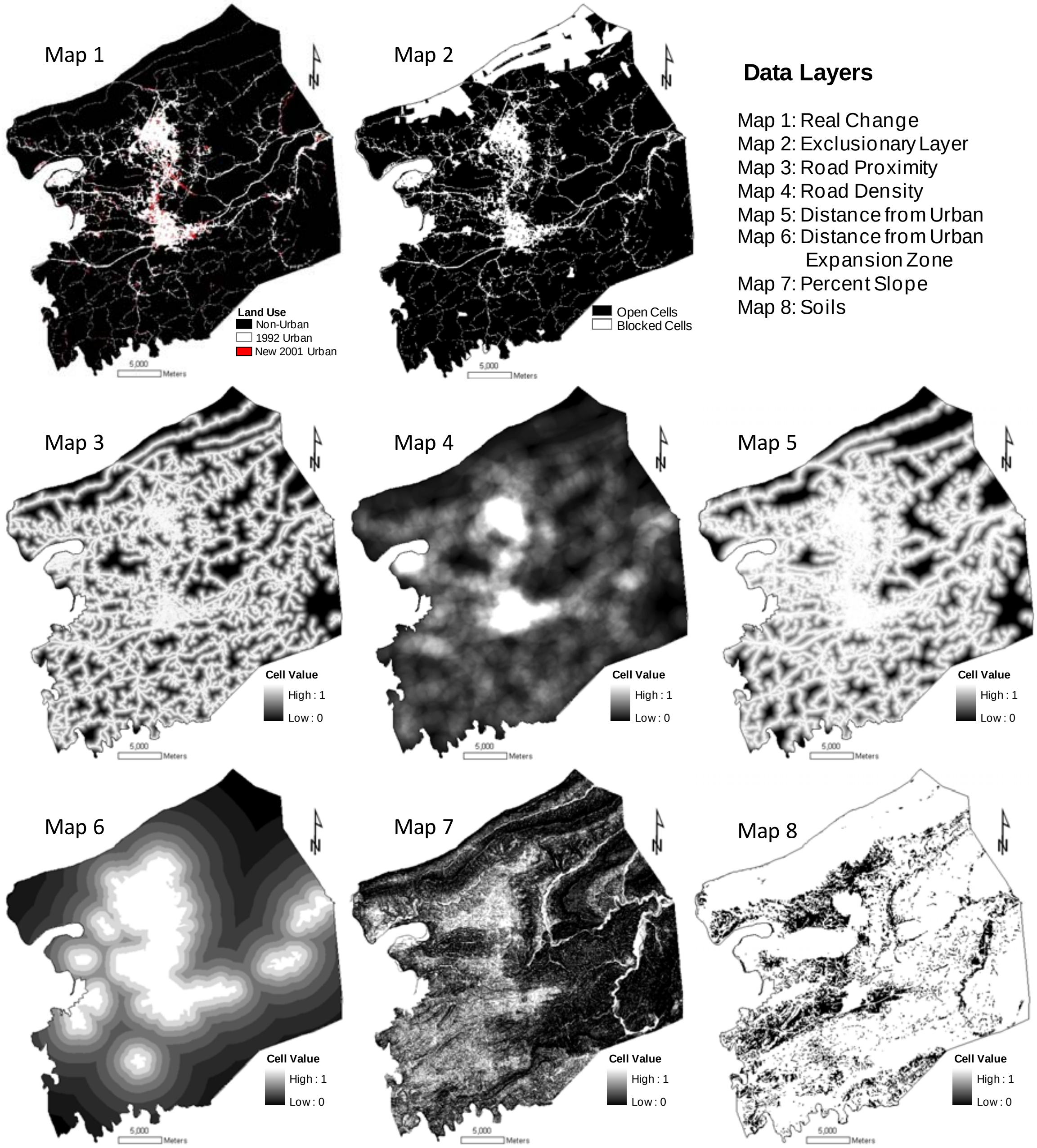


Figure 3.2 Maps representing land use change over the study period (map 1), excluded cells (map 2), and six driving variables being analyzed (maps 3-8).

Exclusionary Layer

The LTM allows the user to block out areas where development cannot occur because of either strict policy enforcement or physical constraints. Montgomery County's physical constraints include open water, existing 1992 urban (because growth cannot occur where it already exists), and cells falling within conservation easement boundaries and Jefferson National Forest to the north were also blocked. Local parks, however, were not excluded because no policy prohibiting development in these areas existed during the analysis period.

Open water and 1992 urban cells were selected out using the NLCD Retrofit dataset, while policy constraints were provided online by the county in shapefile format. A conservation easement file featured seventeen areas around the county and a zoning file providing Jefferson National Forest's boundary with its C-1 (conservation zone) classification. These sets were combined (Figure 3.2, Map 2) to form the exclusionary layer whose purpose was to block certain areas from analysis.

Accessibility

Throughout LUCC literature it is widely accepted that the building or improvement of roads will spur development because of improved accessibility. Denser urban areas which already have access to utilities like water lines, sewer lines, electricity, and cable services among others also can be expected to attract residential development because of the reduced cost of connecting to these services (Pijanowski et al., 2005, Allen & Lu, 2003). Because these areas are already filled with a high density of roads, road density can indicate presence of such urban services. Datasets related to these accessibility phenomena are standard to all LUCC models examining urban growth.

For our area, road centerline data derived from Topologically Integrated Geographic Encoding and Referencing system (TIGER) data for the year 1992 provided the base for two layers of accessibility: proximity to 1992 roads, and density of roads for 1992.

The proximity layer was created by taking the Euclidean distance from each cell to the nearest road. The grid was then divided by its maximum value and subtracted from one to show the negative relationship between road proximity and likelihood for urbanization in a 0-1 range.

Road density was found using Spatial Analyst’s simple density function with a search radius matching the grid extent (1290 cells). Dividing this grid by its maximum value provided a range of cell values from 0-1. Proximity to roads and road density maps are displayed in Figure 3.2, Maps 3 & 4.

Proximity to Urban Land

Areas closest to urban areas are likely locations for urbanization because they are assumed to require the least monetary cost for connecting to urban services such as water and sewer (Pijanowski et al., 2005). The 1991 Blacksburg Comprehensive Plan also calls for higher density cluster development. It specifically aims to implement policies encouraging new industrial sites to be located adjacent to existing industrial, commercial expansion along collector streets, and filling linear commercial expansion in vacant parcels near existing strip commercial lands.

Using Spatial Analyst’s straight line feature, the Euclidean distance was calculated from all 1992 urban cells for the entire study area. By dividing the furthest distance from the entire layer and then subtracting the layer from one, a negative relationship was calculated (Figure 3.2, Map 5).

Urban Expansion Areas

The Comprehensive Plan classified seven urban expansion areas throughout the county which were encouraged to promote medium density residential growth, community-oriented businesses, and small industries. Likewise, two rural expansion areas were also designated with similar development goals of lower intensity development and greater agricultural focus. These nine areas provide linkages between all county urban areas because of their location along major roads and are likely sites for new water and sewer utilities over the next 5-10 years.

The 1990 Montgomery County Comprehensive Plan Map provided locations of these expansion areas. The paper map was georeferenced from a scanned TIFF image using road centerlines (provided by the county) as the reference for control points. By using road intersections to create 20 control points, the map was georeferenced with a total RMS (root-mean square) positional error of 35.94 (meters). It was then rectified to an Imagine file and the seven urban expansion areas, along with the two rural expansion areas were digitized to a new polygon file. The towns of Blacksburg and Christiansburg were also pulled from the county provided boundary file and included in this polygon feature because a majority of higher intensity development (residential, commercial, and industrial) was expected within their boundaries.

Straight line distances for each cell were calculated from each town or expansion area. They were then broken down into ten classes: nine using the geometric interval method, and a tenth for all areas falling directly within towns or expansion areas. Bordering grids cells resulting from vector lines were also reclassified to a value of ten, while those furthest were reclassified as one. Multiplying the grid by 0.1 produced ten classes in a 0-1 range (Figure 3.2, Map 6).

Slope

According to the 1990 Comprehensive Plan, areas with "a natural slope less than 25%" are designated for urban expansion areas and rural expansion areas. Slopes above this range are declared not suitable for development. The Route 460/114 Corridor Development Plan and Standards (1988) broke topography down further into four classes of slope percentage: level land (0-5%), rolling land (6-10%), hilly land (11-15%), and steep land greater than 15%. In keeping with the comprehensive plan's, the 15-25% slopes were included with the lowest suitability scores other than zero.

To include the impact of elevation and slope in the analysis, a digital elevation model (DEM) from the United States Geological Survey's (USGS) National Elevation Dataset provided a grid of 1 arc-second (30-meter resolution) of elevation data. Using Spatial Analyst's slope tool, slope percentages were calculated and reclassified into their five respective slope classes listed above [4, 3, 2, 1, and 0] and multiplied by a value of .25. This process resulted in a value range from 0 – 1, level land having a value of one and unsuitable lands a value of zero (Figure 3.2, Map 7).

Soils

Montgomery County's 1983 Comprehensive Plan aspired to preserve the county's farmlands by diverting development towards other land uses, mainly forest, which does not contribute as much to the economy as do agricultural lands. Although earlier efforts were

Table 3.2
Soil potential index for 17 agricultural soils.

SOIL NAME	SOIL POTENTIAL INDEX
16B Grosclose Poplimento	100%
28 Ross	96%
15B Glenelg	92%
20B Hayter	89%
30B Unison Braddock	89%
12B Frederick Vertress silt	89%
11B Duffield-Ernest complex	87%
13 B Frederick Vertrees cherty	85%
19B Gurnsey	83%
16C Groseclose Poplimento	82%
12C Frederick Vertrees silt	81%
30C Unison Braddock	80%
11C Duffield-Ernest complex	78%
21C Hayter	77%
15C Glenelg	76%
17C Groseclose Poplimento cherty	76%
13C Frederick Vertrees cherty	75%

unsuccessful in limiting development on farming lands, the 1990 plan set forth a policy (EN 3.3.0) and strategy (EN 3.3.1) to protect prime agricultural lands in active use. Planners used the Land Evaluation and Site Assessment (LESA) system as a tool in the decision-making process concerning farmlands with the goal (EN 3.0.0) of preventing conversion to other land uses (1984). The county designated 17 soils for their physical/chemical properties, rating their potential suitability of producing an indicator crop (corn) based on LESA findings (Table 3.2).

Spatial and tabular SSURGO 2.2 files downloaded through the United States Department of Agriculture's (USDA) Natural Resources Conservation Service's (NRCS) Geospatial Data

Gateway provided information concerning the soil substratum. The 17 soils for this region were classified as inverted percentages of their respective soil potential index scores. Frederick Vertrees cherty soil (13C), for example, was assigned .25 while Grosclose Poplimento soil (16B) was assigned zero. All areas outside of the 17 classified soils were assigned values of one because policy did not affect their potential for development (Figure 3.2, Map 8).

Neural Network Parameterization

This neural network was built to examine the accuracy of the LTM's ability to locate cells of change, and to gain a better understanding of how each of the six drivers mentioned above contribute to the overall performance of the model. This model followed the LTM's recommended standard structure with one input layer, one hidden layer, and one output layer. Six nodes (one representing each land change driver) were made for both the input layer and hidden layers, while one node ("new urban") was assigned to the output layer (Figure 3.3). Each of the six drivers were randomly assigned to each input node (by the SNNS), while a "real change" binary layer was created for the output node from the LTM's subtraction of 1992 urban from 2001 urban.

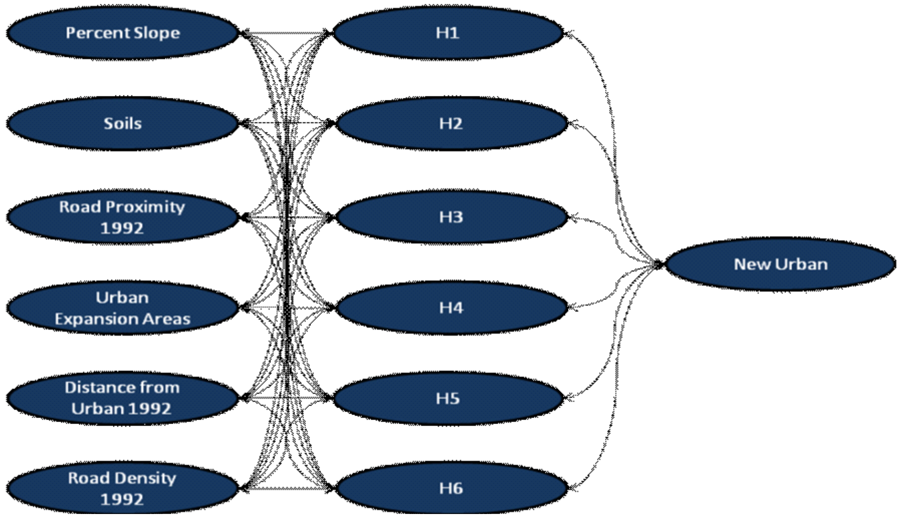


Figure 3.3 Neural network structure.

Sample areas were chosen to train the model to mimic the regional approach most LUCC models take and to assure the model was generalizing patterns. Selecting a training region from within the county was necessary to assure the model could predict urban growth patterns exclusively to Montgomery County, VA. The county was broken into four quadrants; the neural net was executed on each quadrant, stopping at the 10,000th cycle. The quadrant analysis allows us to represent nearly 20% to 30% of urban change cells in the county while offering variations in land uses matching the rest of the county's makeup. Similar methods were used by Pijanowski et al. (2005) when training the Twin Cities Metropolitan Area on 1% of the study area containing 50% of urban transition cells. The accuracy of spatial prediction produced from training each quadrant was compared and the best performing model was chosen for further analysis.

Validation Techniques

PCM's and Kappa statistics are the two automated results the LTM produces for each simulation epoch. Percent correct metrics are calculated by dividing the number of true positive cells (TP) by the total number of cells changing over the analysis period (RP). Kappa values are calculated using the values from the contingency table (Table 3.3) in the following equation:

$$K = \frac{P(A) - P(E)}{1 - P(E)} = \frac{\sum_{i=1}^{c} p_{ii} - \sum_{i=1}^{c} p_{iT} \cdot p_{Ti}}{1 - \sum_{i=1}^{c} p_{iT} \cdot p_{Ti}} \quad \text{[Eq. 3.1]}$$

where *P(A)* is the fraction of agreement (sensitivity coefficient) and *P(E)* is the expected fraction of agreement subject to the observed distribution.

Relative Operating Characteristics (ROC) is one method Pontius & Batchu (2003) offer to assess the spatial accuracy of LUCC models. The ROC module, found in IDRISI Andes, was used to predict the location of new urban cells by comparing an input suitability image of change representing the probability of urban growth to occur from the LTM being tested with a reference image showing where the real change occurred in binary format. Pontius et al. (2001) notes the utility of this statistical analysis because it allows the researcher to assess a set of maps in terms of cell location without seeking agreement between maps in the context of cell quantity.

		Real Change Categories		**Total**
LTM Change Categories		**0**	**1**	
	0	TN	FN	SN
	1	FP	TP	SP
	Total	RN	RP	GT

Table 3.3 Contingency Table comparing LTM with real change using prediction categories from Table 3.1.

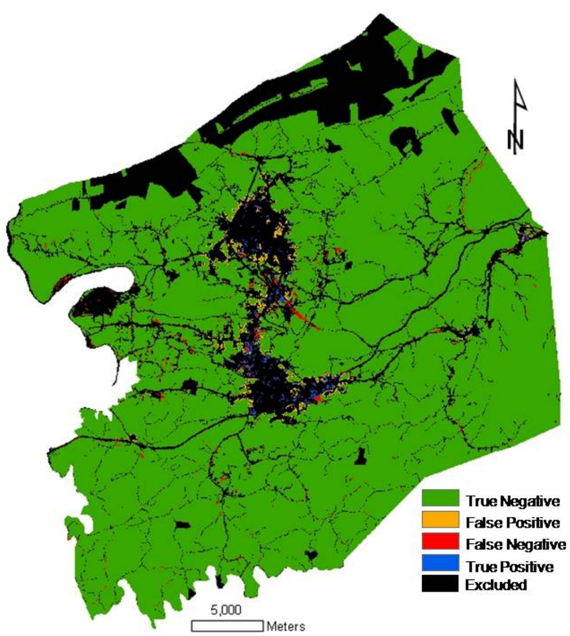
The best performing LTM model was compared to results from IDRISI's Logistic Regression (LR) model to assess how the neural net compares to a common LUCC model. The IDRISI drivers were assigned equal weights to assure the model gave each driver the same consideration as the LTM. Since IDRISI's logistic regression module did not offer a choice for defining a training region, a stratified random sample of urban and non-urban cells nearly equivalent to the LTM's training quadrant was used to train the LR model.

The LTM was also executed six separate times, each without one of the driver variables to assess the relative sensitivity of each driver's contribution to the overall performance of the model. Along with this sensitivity analysis, the resulting prediction and suitability maps from the highest PCM model were used to assess the model's performance with varied land uses (land use assessment) and investigate the causes of error for each predicting category (TN, FN, FP, TP).

Visual assessment identified patterns in resulting prediction maps and examined the real changes that occurred over the nine-year study period. Aerial images from 1991 National Aerial Photography Program (NAPP), 2002 Virginia Base Mapping Program (VBMP), and 2005 National Agriculture Imagery Program (NAIP) provided the basis for assessing land use changes at each predicted category, and aided in accuracy assessment of the original land use data. First-hand knowledge of the study area also contributed to understanding of observed changes.

Chapter 4: Results

The southwestern quadrant yielded the highest PCM and Kappa values of the four initial runs of the LTM. Unless otherwise noted, this model, trained on the southwestern portion of the county, will represent the LTM model for the remainder of this analysis. The suitability map (Figure 4.1, right) shows the rank of each cell's likelihood for converting to urban land use from a non-urban use between 1992 and 2001 according to the LTM model. After the highest ranking 11,823 (number of cells matching "real change" over nine years) were selected, a categorical error map (Figure 4.1, left) shows the accuracy of the model's prediction by representing correctly predicted cells as "true" (blue and green) and incorrectly predicted cells as "false" (orange and red).



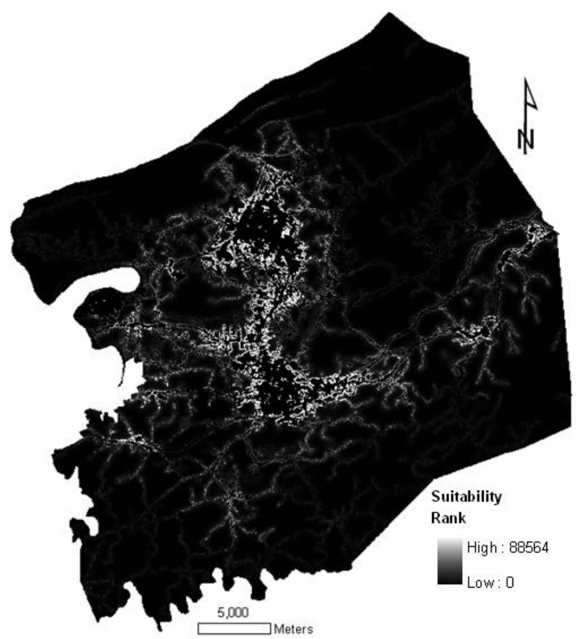


Figure 4.1 LTM predicted categories (left) and ranked cell suitability map (right).

Results show a PCM of 32.843% and a Kappa value of 0.319. A ROC value of 0.660 calculated from the highest ranking cells in the suitability map showed that the model performed better than chance (0.50). The error matrix (Table 4.1) for the categorical map from Figure 4.1 shows the totals for each predicted category used to calculate the Kappa.

		Real Change Categories		Total
LTM Change Categories		0	1	
	0	877289	7940	885229
	1	7940	3883	11823
	Total	885229	11823	897052

Table 4.1 Comparison of LTM prediction cells and real change cells used to calculate Kappa.

A logistic regression model, trained on a stratified random sample of 36.3% of both urban and non-urban cells resulted in a PCM of 31.752%, a Kappa of 0.308, and a ROC of 0.654. In 1977, Landis and Koch divided KHAT values into three ranges: >0.80 (strong agreement), 0.40-0.80 (moderate agreement), and <.40 (poor agreement). Although these Kappa values represent poor agreement between prediction and real change, the similar performance using logistic regression indicates consistency between the two model types (Landis and Koch, 1977; Congalton and Green, 1999).

The sensitivity analysis (Table 4.2) shows the resulting PCM, Kappa, and ROC values. Each row shows the performance of the full model in the absence of a designated driver and the difference from the LTM for each statistic.

Driver Removed	PCM	Kappa	ROC
LTM (Full)	*32.843*	*0.319*	*0.66*
Road Density	27.971 (-4.872)	0.270 (-0.049)	0.635 (-0.025)
Distance from Urban	28.030 (-4.813)	0.271 (-0.048)	0.635 (-0.025)
Distance from Urban Expansion Zone	29.908 (-2.935)	0.290 (-0.029)	0.645 (-0.015)
Percent Slope	31.566 (-1.277)	0.307 (-0.012)	0.653 (-0.007)
Road Proximity	32.987 (0.144)	0.321 (0.002)	0.660 (0)
Soils	33.494 (0.651)	0.326 (0.007)	0.663 (0.003)

Table 4.2 Model performance statistics in the absence of the driver listed in first column and the difference from the model run with all six drivers.

PCM and Kappa values maintain poor agreement during each test of the sensitivity analysis. ROC values ranging from 0.635-0.663 show that model performance is better than chance. Although these statistics seem to maintain consistency with full runs of the LTM and LR models, note that the model actually shows improved performance in the absence of the soils driver.

Several tests of both models using the given drivers seem to show that predictability could be maximized at approximately 33%. ROC values consistently around 0.65 also indicate that although the spatial prediction is better than chance, there is no indication of strong predicting ability. This neural network model performed similarly to LTM models done at regional scales. In their calibration of the Twin Cities and Detroit metropolitan areas, Pijanowski et al. (2005), calculated Kappas from error matrices of the LTM model resulting in ranges from 0.12-0.30 and 0.20-0.26 respectively for six simulations tested on each site.

None of the 1992 urban cells converted to other land uses, while together, barren, forest, and agricultural lands shifted 11,823 cells to urban by 2001 (Table 4.3). Agricultural land gained 10,323 cells (a 1.2% change) despite forfeiting 3,307 to urban uses. Forests (which make up over 60% of the county's land use) experienced a -2.3% change. Of the lost forested land, over 42% (8,520 cells) converted to the urban classification.

		1992 Land Use							
		Water	Urban	Barren	Forest	Grassland/Shrub	Agriculture	Wetlands	Totals
2001 Land Use	Water	3489	0	0	1	0	165	0	3655
	Urban	0	128608	10	8520	0	3307	0	140445
	Barren	0	0	123	390	0	40	0	553
	Forest	0	0	0	704718	0	3800	0	708518
	Grassland/Shrub	0	0	0	1297	1139	8	0	2444
	Agriculture	0	0	0	10323	0	251704	0	262027
	Wetlands	0	0	0	0	0	0	177	177
	Totals	3489	128608	133	725249	1139	259024	177	1117819

Table 4.3 Contingency table displaying the number of cells for each land use class.

Chapter 5: Discussion

Analysis

During the nine-year study period, our analysis of NLCD Retrofit data shows the county experienced a 1% increase in urban land use. These changes can be characterized as either of two main types: infilling and expansion (Civco et al., 2002). During the analysis period, *infilling* accounted for a majority of the growth. This type of growth, characterized by the urbanization of smaller non-urban areas surrounded by existing urban areas, primarily occurred within the county's two towns and to a lesser extent within urban expansion areas. Growth included 2,020 new urban cells in Blacksburg, 2,870 in Christiansburg, and 1,479 spread among urban and rural expansion areas. Major urban growth also included *expansion* along roads, including the Smart Road, a proposed connector to I-81 southeast of Blacksburg and a new road near the county's eastern border. These new roads contributed 421 and 376 new urban cells respectively to the real change occurring during the study period.

The LTM performed well at predicting infill development (Figure 5.1). Most false positives (orange) occurred on the fringes of towns, while false negatives (red) occurred along roads, as the model did not predict growth there. This pattern is not surprising given that Blacksburg's 1991 Comprehensive Plan called for higher density clustered development for new industrial areas to be located adjacent to existing industrial uses, commercial expansions to occur along collector streets, and filling linear commercial expansion in vacant parcels near existing strip commercial.

Relative to the amount of urban growth occurring in most LUCC studies, Montgomery County showed little urban development. Pontius et al. (2008) hypothesizes that reference maps showing larger amounts of net change in land use offer LUCC models better opportunities to detect and predict change purely due to the increased probability of finding change when there is a greater quantity of change. In his evaluation of nine models, Pontius et al. (2008) fails to mention that models tested on developing countries or developing regions seem to perform far better than those applied to areas already extensively developed. A model run on a large region, with ample land availability open to development, certainly seems to have little chance of performing at the same level as a model in a smaller area where larger proportions of land are subject to development. This quantity issue, coupled with America's sprawling patterns, often characterized by leapfrogging of development over open lots to reach land further from existing development, greatly complicates the effort to model new urban land.

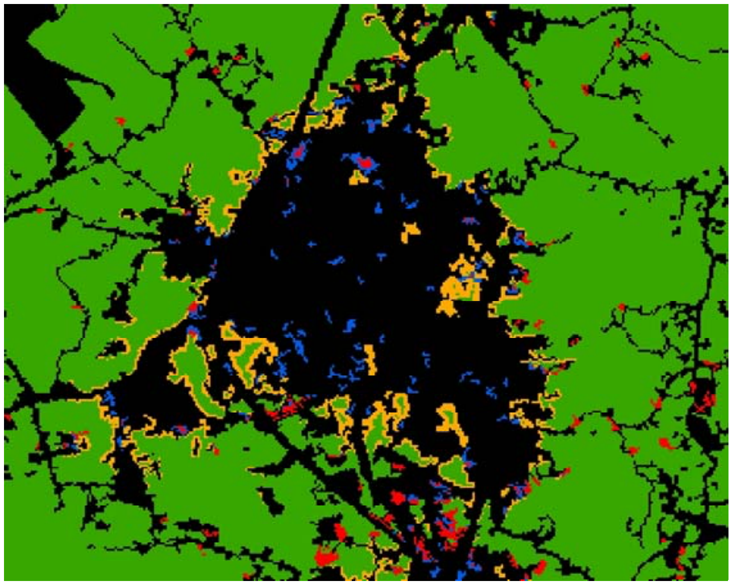


Figure 5.1 Categorical assessment of Blacksburg and surrounding area.

Our model may have predicted patterns correctly based on the policy-oriented rules set for the region, while developers may have behaved irrationally in decision-making for land development or (more accurately) the model may not have implemented the same logic employed by developers. It is possible the LTM acted as a suitability model identifying optimal development sites while developers applied a different mix of criteria selecting land for development. Much of this disparity may be due to the absence of economic drivers for the network (Irwin & Geoghegan, 2001). Ideally, this research would also include a normalized dataset for land value. Another useful driver would likely be a constraint that allows only parcel size units to change rather than individual cells. These drivers, in conjunction with a broader time-span, could allow the model to identify situations in which "leapfrog" development might occur, due to variations of land value and lot size. Areas such as the new residential subdivisions might have been identified with additional data; however, neither of these sets was available for the time of analysis.

Sensitivity Analysis

Six suitability maps showing the 11,823 cells most likely converting to urban land use from the LTM were produced to compare visually the prediction patterns of models executed missing one driver with the patterns produced by a model using the full complement of drivers and a map of the actual change that occurred. Each of these maps (Figure 5.2) will be referred to by their respective map numbers. Comparison of these maps (5-10) also offers a means for assessing of the relative sensitivity of each driver.

Figure 5.2 Map of "real" new urban cells (Map 3) and predicted urban change maps (Maps 4-10.)

Different types of growth patterns can be recognized in visual assessment of maps. The real change map (Map 3) displays dense growth with several clusters visible in Christiansburg and some in Blacksburg, while more scattered growth is evident throughout the rest of the county.

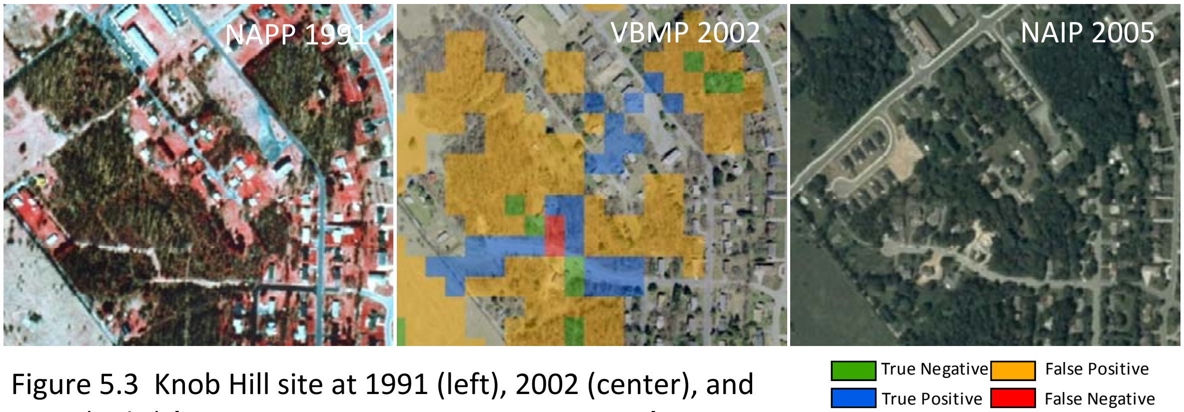


Figure 5.3 Knob Hill site at 1991 (left), 2002 (center), and 2005 (right) **(NAPP 1991, VBMP 2002, NAIP 2005)**.

The absence of the road density driver (Map 5) creates the most dispersed pattern of growth and shares the poorest spatial predictability scores with the model leaving out "distance from urban" (Map 6) according to resulting ROC values (Table 4.2). Conversely, "distance from urban" sensitivity shows the most cohesive grouping of prediction cells with tightly clustered growth evident from the dense white patches in and around the towns and village fringes. Map 5 also incorrectly predicts development in a rural expansion zone south of Christiansburg -- while no other model displays a cluster of prediction cells there. Distance from urban expansion areas may have constrained growth to towns and villages, because in its absence real change cells missed by all other models, predicted the growth of new urban cells for a residential area just east of Blacksburg. Surprisingly, lack of a distance from urban expansion zone driver still shows growth in villages. The distance from urban driver probably makes up for the expansion zone absence, because it recognizes the urban cells in these smaller satellite communities.

Large clusters of prediction cells are positioned around the middle of Blacksburg just inside the eastern border, known today as the Carriage Court and Knob Hill subdivisions (Maps 4 and 6-10). This area included consistently high values for each road density pixel, causing the area to be nearly non-existent in Map 5. It is evident from 2002 aerial imagery that growth did not occur here, despite strong suitability values among all drivers. However, 2005 NAIP imagery reveals that these areas were later developed into small cookie-cutter subdivisions characteristic of modern residential development (Figure 5.3). This phenomenon indicates the model has the ability to forecast future development sites based on given suitability inputs, but in this instance, development did not occur within the nine-year study period.

Similar sets of prediction patterns are found in and around towns and villages (Maps 7-10), indicating that interaction of the distance from urban and road density drivers balance the two types of patterns the model displays: (1) clustered-dense growth from the road density driver and (2) dispersed-spotty growth as the model attempts to place new growth near prior urban cells.

Land Use Assessment

During the study period, 10 barren cells, 8,506 forest cells, and 3,307 agriculture cells converted to urban land use. The LTM predicted 30% of barren cells and 25% of forest cells correctly, but over 50% of agriculture cells correctly (Table 5.1). This large disparity in predictive ability could be attributed to the larger percentage of agricultural cells available for training in the southwest corner. According to Nefeslioglu and Gokceoglu's 2008 review of neural network literature, approximately 80% of all data is usually needed to train networks before testing final models. The LTM tutorial, on the other hand, recommends using sample boxes covering about 10% of data. Nevertheless, training the network on the entire county yielded testing results of approximately 33% PCM. It should be noted that of the four quadrants tested as training areas, the chosen southwest quadrant contained the most "new urban" cells, indicating that running the LTM with higher quantities of "known change" cells may provide greater predicting power.

	False Negative	True Positive
Barren	7	3
Forest	6284	2222
Agriculture	1649	1658

Table 5.1 Numbers of cells in each land use predicted correct and incorrectly.

Agriculture and forest contributed the most land to new urban growth over the nine-year span of this study. After the county's second comprehensive plan in 1983, planners found that, while agricultural lands were more economically profitable, they were being converted to urban areas, despite the availability of less expensive of forested land. In response to this finding, the 1990 comprehensive plan called for increased protection of agricultural soils, specifically citing the 17 soils found in the 1984 Land Evaluation Study and Assessment (LESA) to be the most agriculturally suitable. Therefore, it is not surprising that 72% of land use converted to urban came from forested lands. This figure signifies that the comprehensive plan may have been effective in preserving some agricultural lands -- a significant finding.

Categorical Error Assessment

Comparisons of locations and quantities of cells falling into each category of prediction with their representations on aerial imagery offered site-specific ground conditions which may have affected model behavior. Generally, true positives fall within dense urban areas, false positives fall at the fringes of towns, and false negatives are dispersed throughout the county along urban corridors.

True Positives

Most true positive (TP) cells (where the model predicted change correctly) occur deep within towns and villages. Of the 3,883 TP cells the LTM produced, 2,132 fell in Christiansburg, 1,205 in Blacksburg, and 498 within urban and rural expansion areas, leaving only 1.2% (48) of true positive cells outside planned urban areas. These figures demonstrate that the model can be expected to perform best when predicting open cells in previously urbanized areas.

Figure 5.4 Pheasant Run housing complex in Blacksburg 1991 (left), and 2002 (right) **(NAPP 1991, VBMP 2002)**.

New residential sites were well-predicted. Housing developments infilling prior developments comprised a majority of TPs in Blacksburg as well as Christiansburg which also correctly predicted some commercial and industrial developments upon visual inspection. Figure 5.4 shows one site (Pheasant Run) in Blacksburg where townhomes were developed in close proximity to previous residential communities.

False Positives

False positives (FPs), cells in which the model predicted change when it did not actually occur, were mostly located at the fringes of towns or along the interior rims of undeveloped patches (Figure 5.5). A majority of the 7,940 FP cells fell within Blacksburg (3,320) and Christiansburg (2,781), while 1,517 fell within urban and rural expansion areas. Therefore, only 4% (322) of missed predictions fell outside previously or expected areas of development.

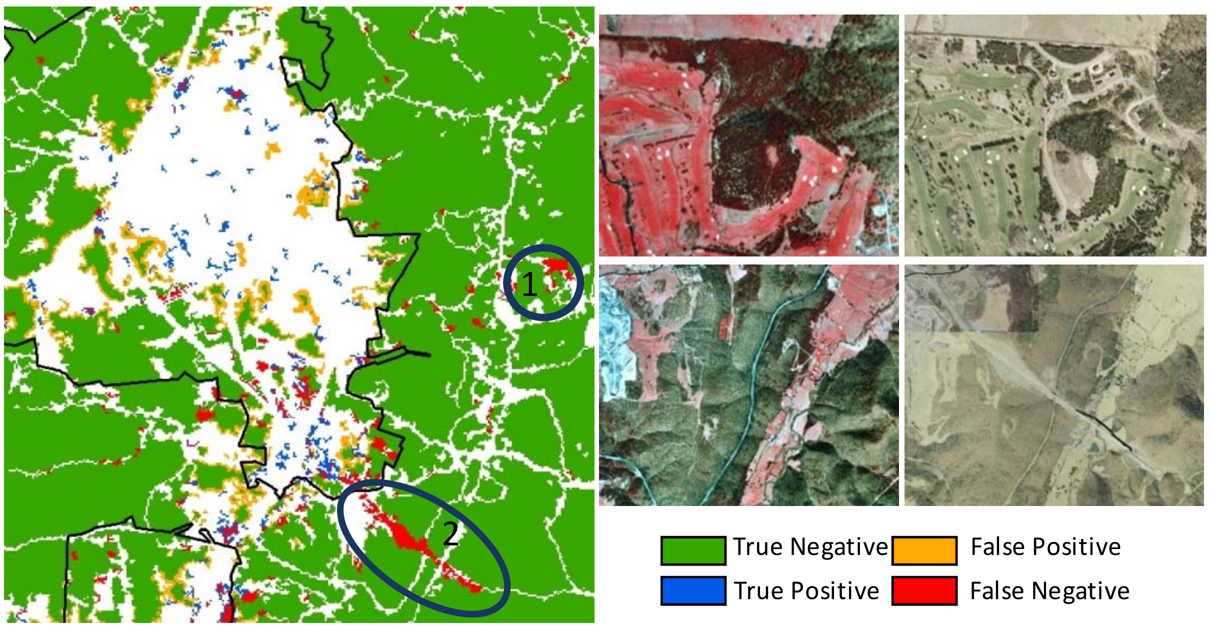


Figure 5.5 Context of FNs for Blacksburg County Club Estates (1, top) and the Smart Road Corridor (2, bottom); aerial images of 1991 (left) and 2002 (right) **(NAPP 1991, VBMP 2002)**.

Since false positive cells seem to fall into thin segments along the borders of already urbanized areas, a change in unit of analysis may improve performance in these areas. If units for study were parcels, for example, entire parcels would be predicted to convert land use rather than small sections of parcels**.**

False Negatives

False negatives (FNs) are areas in which the model failed to predict cells that experienced urban growth. FNs seemed to mainly fall close to roads, usually near intersections. Of the 7,940 cells falling into this category, 5.3% (421) fall into the Smart Road Corridor (Figure 5.5), and 4.7% (376) come from a new road near the eastern border of the county. The model failed to predict a single cell of change within these corridors falling outside town limits. In both cases, FNs are clearly cells representing paved roads. Visual inspection on these sites, however, shows many cases where roads show no signs of improvement or widening (represented by small red speckles seen throughout Figures 4.1 and 5.5), a finding that illustrates the limitations of the land use data.

The model missed some residential developments dispersed throughout the county, and one new school. Blacksburg Country Club Estates (Figure 5.5) represents one of these larger developments. Although this site seems to have moderately high suitability scores for four drivers, a relatively low score for distance from urban expansion areas and its proximity to agricultural soils may have prevented this area from being predicted correctly. These characteristics are similar to the other large FN patches found throughout the county.

Figure 5.6 Collegiate Suites apartment complex in Blacksburg 1991 (left), and 2002 (right) **(NAPP 1991, VBMP 2002)**.

For our area, the model does not predict the transfer of large contiguous areas to urban land use. The LTM never produced one instance in which true positive cells fell more than three cells deep into a patch of new development because these cells are not in close proximity to already urbanized lands. An example of this phenomenon is shown at the Collegiate Suites apartment complex, built in 1998, showing that true positive cells form a ring at the edge of prior urban cells, but false negatives fill the inside of the ring (Figure 5.6). In this example, the model clearly recognizes that new development will occur because drivers display high suitability there, with the exception of soils which can unrealistically weight the suitability values of neighboring cells.

This phenomenon is repeated in the middle of Christiansburg where two true positive rings (blue) surround large patches of false negative cells at the Vistavia and Oak Tree

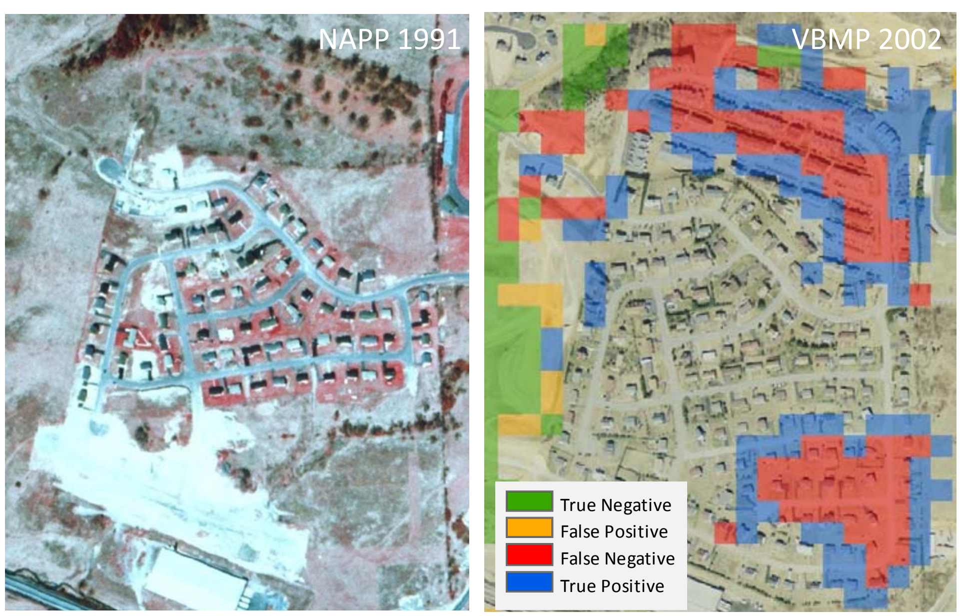


Figure 5.7 Vistavia and Oak Tree Townhomes subdivisions in Christiansburg 1991 (left), and 2002 (right) **(NAPP 1991, VBMP 2002)**.

Townhomes subdivisions (Figure 5.7). Although the model shows the ability to recognize that change is likely to occur in areas surrounded by true positive rings, the model will never perfectly predict the location of a large development using this study's unit of analysis.

Running ANNs may be helpful for identifying areas a GIS model fails to recognize. For example, if errors were consistently detected at the locations of schools for resulting ANN maps, a distance from schools driver could be added to the GIS to determine if it could improve that model. It seems new roads, road extensions, or road improvements were not predicted by the model, as is evident from the number of false negative cells located near these features. The inability of the model to predict change in these areas suggests to a planner/modeler that a new driver may be needed to address this failure. Although the 1990 comprehensive plan only addresses road improvements in the document's text, new comprehensive plans could improve modeling by including maps displaying locations of future road improvements. A new driver could be created from this future road map in the same fashion as the "distance from urban expansion areas" grid to provide one method for improving model calibration for a neural network in Montgomery County, Virginia.

Limitations

Data and software limitations created some inconsistencies across the county. The metadata indicates retrofit sets were derived from Landsat ETM+ (TM) scenes from 1991, 1992, and 2000. Most of the land use classification for the early period was created from a late September image (09/30/1991) during the leaf-off season, while a smaller portion came from the leaf-on season of spring (05/11/1992). Land use classifications for the end of the study period were derived from imagery taken during leaf-on, June 10, 2000. Spectral differences resulting from varying seasons could have contributed to irregularities in land use classification.

Each of the pixels used in retrofit creation were produced first using Anderson level I classifications and filtered in cells falling within a 70% confidence level of both classification years (1992 and 2001). Areas not meeting this confidence level were refined using an 8-neighbor clump and 5-pixel sieve process in which only contiguous areas of 5 pixels or more were retained. Since these refined cells were evaluated using areas approximately equivalent to one acre, they did conform to our 30-meter unit of analysis during preparation. Therefore, it is possible that areas where a relatively small amount of urban change at a 30-meter resolution may not have been detected by the original coarser resolution retrofit data. Visual inspection of retrofit data sets, often showed new urban cells falling along roads where not previously represented as continuous segments in 1992 land use data. Several false negative cells were likely attributable to these areas.

Various methods of supervised and unsupervised land use classifications were used during our initial data compilation, however, these techniques often resulted in unrealistic classifications displaying either speckled urban patches throughout the county, often in deep forest areas or over simplification where several urban cells were removed from growing urban areas. While the MRLC warns against the use of retrofit data at local levels, this set provided easily accessible data with the greatest relative accuracy for this study. Although full accuracy assessment for this region is not currently available, initial assessments of the decision-tree classification method used to produce the retrofit found a 94.2% agreement against change/no-change pixels identified using traditional methods (Homer et al., 2007).

Chapter 6: Conclusions

Further Research

The calibration of the LTM in this analysis presents a useful set of driving factors to forecast future development for Montgomery County, Virginia. An urban growth forecast based upon census bureau population projection data, could be made using Pijanowski's principle index driver (2002a) (*Eq. 2.3*). Further research projecting development patterns beyond the testing period, however, should be done cautiously. Although the model performed well at predicting infilling and fringe development at previously urbanized sites, it performed poorly in tests at forests, which make up a majority of the county.

Although this study reasonably generalized urban growth patterns from the southwest quadrant of the county for the entire area, this training region should be applied to other similar landscapes to further assess the LTM's generalization ability. Different grid resolutions and numbers of neural network simulation cycles also need to be tested, as well as, additional drivers and variations in the types of drivers as new datasets become available.

The current structure of the LTM neural network only allows us to evaluate land use change for one time interval rather than evaluating change in a time-step manner such as is permitted by Markov-chain techniques. As current LUCC modeling research moves toward the use of hybrid models, researchers should investigate integration of the use of neural networks with cellular-automata models to account for land use changes at various temporal scales.

Applications of ANNs for Planners

To some degree, the ambiguity of neural networks may be its greatest asset. The inter-workings of the model are largely unknown to the analyst as data are provided for a so-called "black box;" however, the LTM's ability to randomize variable weights without user-defined weighting offers a relatively hands-free user experience. A model that works independently of user bias may currently be the best modeling option for the field of urban planning, characterized by technological capabilities still in a state of infancy.

Development of LUCC models is not a competition to determine which model performs best, rather to gain a better understanding of the factors at play guiding urban dynamics (Pontius et al., 2008). Although the model presented here shows no significant advantage over a logistic regression model, the LTM does show some ability to predict infilling and fringe development at the county-level. Perhaps such models would be better suited for a more densely packed urban area or large city with traditional growth patterns. Additional research is needed to assess the model's ability to predict change at various scales and with the use of with different drivers.

Summary

The LTM (utilizing an ANN) and a logistic regression appear to perform almost exactly the same when compared to each other after training on nearly a third of the study area's urban land. Results of testing one full run of the LTM using six variables and six variations of the model with five variables show that the model performs consistently, albeit poorly, by Kappa agreement standards. When comparing the results of this countywide study with other variations of popular regional models, the LTM shows accuracy consistent with those models where similar economic and biophysical landscapes were most similar (Pontius et al., 2008). The use of ANNs provides a satisfactory technique for assessing and understanding the spatial patterns and dynamics of urban systems in Montgomery County, Virginia by current LUCC standards.

References

Agarwal, C., Green, G. M., Grove, J. M., Evans, T. P., & Schweik, C. M. (2002). *A review and assessment of land-use change models: dynamics of space, time, and human choice.* Retrieved. from http://www.treesearch.fs.fed.us/pubs/5027.

Alexandridis, K. (2006). Calculation of the Kappa Coefficient for Binary Land Use Change in GIS Maps.

Allen, J., & Lu, K. (2003). Modeling and Prediction of Future Urban Growth in the Charleston Region of South Carolina: a GIS-based Integrated Approach. *Conservation Ecology, 8*(2), 2.

Anderson, J. R., Hardy, E. E., Roach, J. T., & Witmer, R. E. (1976). *A Land Use And Land Cover Classification System For Use With Remote Sensor Data*. Retrieved. from http://landcover.usgs.gov/pdf/anderson.pdf.

Barredo, J. I., Lavalle, C., Demicheli, L., Kasanko, M., & McCormick, N. (2003). *Sustainable urban and regional planning: The MOLAND activities on urban scenario modelling and forecast*: Institute for Environment and Sustainability.

Berling-Wolff, S., & Wu, J. (2004). Modeling urban landscape dynamics: A review. *Ecological Research, 19*(1), 119-129.

Bian, L., & Walsh, S. J. (2002). Characterizing and Modeling Landscape Dynamics: An Introduction. *Photogrammetric Engineering & Remote Sensing, 68*(10), 999-1000.

Brown, D. G., Goovaerts, P., Bumlckl, A., & U, M. (2002). Stochastic Simulation of Land-Cover Change Using Geostatistics and Generalized Additive Models. *Photogrammetric Engineering & Remote Sensing, 68*(10), 1051-1061.

Campbell, J. B. (2007). *Introduction to Remote Sensing* (Third Edition ed.). New York: The Guliford Press.

Clarke, K. C., Hoppen, S., & Gaydos, L. (1997). A self-modifying cellular automaton model of historical urbanization in the San Francisco Bay area. *Environment and Planning B: Planning and Design, 24*(2), 247-261.

Congalton, R. G., & Green, K. (1999). *Assessing the Accuracy of Remotely Sensed Data: Principles and Practices*. New York: Lewis Publishers.

Engelen, G., White, R., & Nijs, T. (2003). Environment Explorer: Spatial Support System for the Integrated Assessment of Socio-Economic and Environmental Policies in the Netherlands. *Integrated Assessment, 4*(2), 97 - 105.

Fischer, M. M., & Gopal, S. (1994). ARTIFICIAL NEURAL NETWORKS: A NEW APPROACH TO MODELING INTERREGIONAL TELECOMMUNICATION FLOWS*. *Journal of Regional Science, 34*(4), 503-527.

Hilferink, M., & Rietveld, P. (1999). LAND USE SCANNER: An integrated GIS based model for long term projections of land use in urban and rural areas. *Journal of Geographical Systems, 1*(2), 155-177.

Homer, C., Dewitz, J., Fry, J., Coan, M., Hossain, N., Larson, C., et al. (2007). Completion of the 2001 National Land Cover Database for the Conterminous United States. *Photogrammetric Engineering & Remote Sensing, 73*(4), 337-341.

Hu, Z., & Lo, C. P. (2007). Modeling urban growth in Atlanta using logistic regression. *Computers, Environment and Urban Systems, 31*(6), 667-688.

Imrie, C. E., Durucan, S., & Korre, A. (2000). River flow prediction using artificial neural networks: generalisation beyond the calibration range. *Journal of Hydrology, 233*(1-4), 138-153.

Irwin, E. G., & Geoghegan, J. (2001). Theory, data, methods: developing spatially explicit economic models of land use change. *Agriculture, Ecosystems & Environment, 85*(1-3), 7-24.

Kocabas, V., & Dragicevic, S. (2006). Coupling Bayesian Networks with GIS-Based Cellular Automata for Modeling Land Use Change. In *Geographic, Information Science* (pp. 217-233).

Land Evaluation and Site Assessment (LESA) (1984).

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

Landis, J. D. (1995). Imagining land use futures: applying the California urban futures model. *Journal of the American Planning Association, 61*, 438-457.

Lee, D. B. (1973). Requiem for large-scale models. *AIP Journal*, 163-177.

Levy, J. M. (2006). *Contemporary Urban Planning*. Upper Saddle River, NJ: Pearson/Prentice Hall.

Lowry, I. S. (1965). A SHORT COURSE IN MODEL DESIGN. *Journal of the American Planning Association, 31*(2), 158 - 166.

McConnell, W. J., Sweeney, S. P., & Mulley, B. (2004). Physical and social access to land: spatio-temporal patterns of agricultural expansion in Madagascar. *Agriculture, Ecosystems & Environment, 101*(2-3), 171-184.

McCormick, N., Lavalle, C., Kasanko, M., Demicheli, L., & Turchini, M. (2001). Modelling the impact of land use planning and management practices on the fragmentation of urban landscapes. In *IEEE/ISPRS Joint Workshop on Remote Sensing and Data Fusion over Urban Areas* (pp. 315-319). Rome.

Montgomery County Planning Commission (1983). *Montgomery County Comprehensive Plan 1983*.

Montgomery County Planning Commission (1990). *Montgomery County 1990 Comprehensive Plan*.

Nefeslioglu, H. A., Gokceoglu, C., & Sonmez, H. (2008). An assessment on the use of logistic regression and artificial neural networks with different sampling strategies for the preparation of landslide susceptibility maps. *Engineering Geology, 97*(3-4), 171-191.

Pijanowski, B., Shellito, B. A., Bauer, M. E., & Sawaya, K. E. (2001). *USING GIS, ARTIFICIAL NEURAL NETWORKS AND REMOTE SENSING TO MODEL URBAN CHANGE IN THE MINNEAPOLIS-ST.PAUL AND DETROIT METROPOLITAN AREAS.* Paper presented at the ASPRS Proceedings 2001.

Pijanowski, B. C., Brown, D. G., Shellito, B. A., & Manik, G. A. (2002a). Using neural networks and GIS to forecast land use changes: a Land Transformation Model. *Computers, Environment and Urban Systems, 26*(6), 553-575.

Pijanowski, B. C., Shellito, B., Pithadia, S., & Alexandridis, K. (2002b). Forecasting and assessing the impact of urban sprawl in coastal watersheds along eastern Lake Michigan. *Lakes & Reservoirs: Research & Management, 7*(3), 271-285.

Pijanowski, B. C., Pithadia, S., Shellito, B. A., & Alexandridis, K. (2005). Calibrating a neural network-based urban change model for two metropolitan areas of the Upper Midwest of the United States. *International Journal of Geographical Information Science, 19*(2), 197-215.

Pontius, R. G., Cornell, J. D., & Hall, C. A. S. (2001). Modeling the spatial pattern of land-use change with GEOMOD2: application and validation for Costa Rica. *Agriculture, Ecosystems & Environment, 85*(1-3), 191-203.

Pontius, R. G., & Batchu, K. (2003). Using the relative operating characteristic to quantify certainty in prediction of location of land cover change in India. *Transactions in GIS, 7*(4), 467-484.

Pontius, R. G., Huffaker, D., & Denman, K. (2004). Useful techniques of validation for spatially explicit land-change models. *Ecological Modelling, 179*(4), 445-461.

Pontius, R., Boersma, W., Castella, J.-C., Clarke, K., de Nijs, T., Dietzel, C., et al. (2008). Comparing the input, output, and validation maps for several models of land change. *The Annals of Regional Science, 42*(1), 11-37.

Rosenblatt, F. (1958). The perceptron: a probabilistic model for information storage and organization in the brain. *Psychological Review, 65*, 386-408.

Route 460/114 Corridor Advisory Planning Council (1988). Route 460/114 corridor development plan and standards.

Silva, E. A., & Clarke, K. C. (2002). Calibration of the SLEUTH urban growth model for Lisbon and Porto, Portugal. *Computers, Environment and Urban Systems, 26*(6), 525-552.

Skapura, D. M. (1996). *Building Neural Networks*. New York: ACM Press.

Sui, D. Z. (1997, June 4-7). *The syntax and semantics of urban modeling: Versions vs. visions.* Paper presented at the Proceedings of USGS-NCGIA Workshop in Landuse Modeling, Sioux Falls, SD.

Theobald, D. M., & Hobbs, N. T. (1998). Forecasting rural land-use change: a comparison of regression- and spatial transition-based models. *Geographical & Environmental Modelling, 2*(1), 65-82.

Town of Blacksburg (1991). *1991 Blacksburg Comprehensive Plan*.

Turner, M. G. (1988). A spatial simulation model of land use changes in a piedmont county in Georgia. *Applied Mathematics and Computation, 27*, 39-51.

Veldkamp, A., & Lambin, E. F. (2001). Predicting land-use change. *Agriculture, Ecosystems & Environment, 85*(1-3), 1-6.

Verburg, P. H., Soepboer, W., Veldkamp, A., Limpiada, R., Espaldon, V., & Mastura, S. S. A. (2002). Modeling the Spatial Dynamics of Regional Land Use: The CLUE-S Model. *Environmental Management, 30*(3), 391-405.

Virginia Code (2007). Comprehensive plan to be prepared and adopted; scope and purpose. 15.2-2223.

Yeh, A. G.-O., & Li, X. (2003). Simulation of Development Alternatives Using Neural Networks, Cellular Automata, and GIS for Urban Planning. *Photogrammetric Engineering & Remote Sensing, 69*(9), 1043-1052.

Electronic Sources

Land Transformation Model. (2008). *Purdue University* Retrieved May, 2007, from http://ltm.agriculture.purdue.edu/default_ltm.htm

Montgomery County GIS Layers. (2008). *Newman Library* Retrieved January, 2008, from http://www.lib.vt.edu/

NLCD 1992/2001 Retrofit Land Cover Change Product. (2008). *Multi-Resolution Land Characteristics Consortium* Retrieved January, 2008, from http://www.mrlc.gov/

The National Map Seamless Server. (2008). *United States Geological Survey* Retrieved December, 2007, from http://seamless.usgs.gov/

Road Centerline Files. (1992). Topologically Integrated Geographic Encoding and Referencing System. Retrieved December, 2008, from CD-ROM.

VBMP Disclaimer

ALL DATA IS PROVIDED "AS IS." THE COMMONWEALTH OF VIRGINIA, THE VIRGINIA GEOGRAPHIC INFORMATION NETWORK AND THEIR RESPECTIVE, OFFICERS, AGENTS AND EMPLOYEES JOINTLY AND SEVERALLY DISCLAIM ANY AND ALL REPRESENTATIONS OR WARRANTIES, EXPRESS OR IMPLIED, WRITTEN OR ORAL, IN FACT OR ARISING BY OPERATION OF LAW, INCLUDING ANY WARRANTIES OF MECHANT ABILITY, FITNESS FOR A PARTICULAR PURPOSE, ACCURACY, CURRENCY, COMMERCIAL VALUE, OR FREEDOM OF DATA FROM INFRINGEMENT OF ANY THIRD PARTY INTELLECTUAL PROPERTY RIGHTS.