

I am submitting herewith a thesis written by Steve Raymond Ahrens entitled “Land Use-Transportation Interaction: Lessons Learned from an Experimental Model using Cellular Automata and Artificial Neural Networks.” I have examined the final electronic copy of this thesis for form and content and recommend that it be accepted in partial fulfillment of the requirements for the degree of Master of Science, with a major in Geography.

Dr. Shih-Lung Shaw, Major Professor

We have read this thesis
and recommend its acceptance:

Dr. Margaret M. Gripshover

Dr. Bruce A. Ralston

Accepted for the Council:

Carolyn R. Hodges, Vice Provost and
Dean of the Graduate School

**Land Use-Transportation Interaction: Lessons Learned from an Experimental
Model using Cellular Automata and Artificial Neural Networks**

A Thesis
Presented for the
Master of Science Degree

The University of Tennessee, Knoxville

Steve Raymond Ahrens
May, 2008

ABSTRACT

Land use and transportation interact to produce large urban concentrations in most major cities that create tremendous sprawl, noise, congestion, and environmental concerns. The desire to better understand this relationship has led to the development of land use–transport (LUT) models as an extension of more general urban models. The difficulties encountered in developing such models are many as local actions sum to form global patterns of land use change, producing complex interrelationships. Cellular automata (CA) simplify LUT model structure, promise resolution improvement, and effectively handle the dynamics of emergent growth. Artificial Neural Networks (ANN) can be used to quantify the complex relationships present in historical land use data as a means of calibrating a CA-LUT model. This study uses an ANN, slope, historical land use, and road data to calibrate a CA-LUT model for the I-140 corridor of Knoxville, TN. The resulting model was found to require a complex ANN, produce realistic emergent growth patterns, and shows promising simulation performance in several significant land classes such as single-family residential. Problems were encountered as the model was iterated due to the lack of a mechanism to extend the road network. The presence of local roads in the model’s configuration strengthened ability of the model to simulate historical development patterns. Shortcomings in certain aspects of the simulation performance point to the need for the addition of a socio-economic sub-model to assess demand for urban area and/or an equilibrium mechanism to arbitrate the supply of developable land. The model constructed in this study was found to hold considerable potential for local-scale simulation and scenario testing given suitable modification to its structure and input parameters.

TABLE OF CONTENTS

I. INTRODUCTION	1
II. LITERATURE REVIEW	4
1. Land Use—Transportation Modeling	4
2. Microsimulation and CA—Beyond the Aggregate	6
3. Artificial Neural Networks and CA	9
4. Recent Land Use—Transportation CA Models	10
III. RESEARCH PLAN	12
1. Study Area	12
2. Project Data	14
3. Project Objectives	15
IV. MODEL DESIGN AND CONTRUCTION	16
1. CA Transition Formulation	16
2. Artificial Neural Networks	16
3. Model Design	19
4. Model Construction	20
5. Training Performance and Network Complexity	21
6. Computability Concerns	22
V. RESULTS	23
1. Historical Land Use Change	23
2. Test Configurations	24
3. Road Corridor	24
4. Iteration	26
5. ANN-Prediction Accuracy	30
6. Single-Family Residential	31
7. Multi-Family Residential	34
8. Agriculture/Forested/Vacant Land	35
9. Commercial	36
10. Office	38
11. Parks/Public Land/Recreation	39
VI. CONCLUSIONS AND FURTHER RESEARCH	40

BIBLIOGRAPHY.....	43
APPENDIX.....	48
VITA.....	59

FIGURES

Fig. 1 – One iteration of a cellular automata (CA)	3
Fig. 2 – The general structure of a land use—transportation (LUT) model	5
Fig. 3 – Study area	13
Fig. 4 – The perceptron or ‘processing element’	17
Fig. 5 – Multi-layer perceptron (MLP).....	18
Fig. 6 – Back-propagation (BP) training	19
Fig. 7 – ANN training results.....	21
Fig. 8 – Road corridor conversion	25
Fig. 9 – Iteration problem	27
Fig. 10 – Horizontal output glitches	28
Fig. 11 – Iterative growth	29
Fig. 12 – SFR prediction with 1993 local roads	32
Fig. 13 – SFR prediction with 2005 local roads	32
Fig. 14 – SFR prediction with no local roads	33
Fig. 15 – Railroad effect	33
Fig. 16 – Multi-family residential prediction with 1993 local roads	34
Fig. 17 – Ag/Forested/Vacant prediction with no local roads	35
Fig. 18 – Commercial prediction with 1993 local roads.....	36
Fig. 19 – Commercial prediction with 2005 local roads.....	37
Fig. 20 – Commercial prediction with no local roads.....	37
Fig. 21 – Office prediction with 1993 local roads	38
Fig. 22 – Parks/Public Land/Recreation prediction with 1993 local roads	39

CHAPTER I

INTRODUCTION

Land use and transportation infrastructure interact in a reciprocal fashion. If the land use present in a particular area demands it, a road to service that area will likely be built. The new route improves the accessibility of the area, increasing its attractiveness to potential users for a given land use. This may in turn strengthen the impetus to create additional road capacity to better serve the area's need. The cycle can also be initiated by changes in land use. This interdependence makes folly attempts to study one factor in isolation from the other.

The notion of sustainable transportation, that being the meeting of today's transportation needs without compromising the ability of future generations to meet theirs, is receiving increasing attention as the byproducts of our current transportation systems and the fossil fuels that power them mount. It can be argued that factors such as the limited nature of petroleum reserves, the negative impacts of petroleum-based emissions on air-quality, traffic-related injuries and deaths, congestion and urban sprawl make our current transportation system unsustainable (Black, 1997).

Additional pressure is exerted by policies implemented by governments in most industrialized nations. The United States' Intermodal Surface Transportation Efficiency Act (ISTEA) of 1991 specifies that "transportation plans must take into account the likely effect of transportation policy decisions on land use and development and the consistency of the transportation plans and programs with land use and development plans" (quoted in Miller, et al., 1999, p.3). ISTEA was superseded in 1998 by The Transportation Equity Act for the 21st Century (TEA-21), which also recognizes the land use and transportation relationship within a broader context of economic development and environmental issues.

Both measures require that transportation plans satisfy the requirements of the Clean Air Act Amendments (CAAA) passed in 1990. The inability of existing models to meet these legislative imperatives led the US Department of Transportation and the US Environmental Protection Agency to formulate the Travel Model Improvement Program (TMIP) of 1993 (Shaw and Xin, 2003). In Europe, the TRANSLAND research project conducted as part of the European Union's Fourth Programme Framework examines innovative transportation policies and defines future research needs.

Pressure to make astute urban policy decisions comes not only from governing bodies. Of the United States' 281,421,906 citizens, 222,360,539 reside in urban areas (SF1, 2000 US Census). Congestion is a pressing issue in most major American urban areas. While the average citizen may not be aware of the full range of causal factors for it, they can well appreciate the impact congestion has on their daily routine. And the congestion outlook is bleak, in that it will likely worsen over the foreseeable future due to continued suburbanization and increasing use of the automobile.

These desires to understand the interactive relationship between transportation and land use, progress towards sustainability, obey governmental edict and harness the political will surrounding congestion and environmental issues has led to the development of land use—transportation (LUT) models as an extension of more general urban models. The challenges presented such models are significant. The relationships

involved in land use change are complex—variable through differences in site and situation, operating on differing temporal and spatial scales, and are the composite of individual actions typically established through personal decisions considered at a large scale. Also inherent in the notion of complexity is the phenomenon of emergence, to the extent that the study of complexity has been termed the science of emergence (Krugman, 1996).

In emergent phenomena a number of rules operating at the local level can generate complex global patterns. This is the case where land use is involved, as individual decisions considered at the local level sum to the overall pattern of land use noted. In similar fashion, individual travel needs and demands conspire to shape overall patterns of travel and the networks that facilitate them. Conventional LUT models struggle to depict the dynamics of land-use change, especially emergence. Temporal aspects enter the model only through cross-sectional data, or ‘snapshots’, which bear the full task of representing temporal change. Patterns cannot appear as evolutionary arrangements of more basic components, but rather appear as interpolations of the available snapshots of the existent input data.

Microsimulation models are computer models that operate at the individual level of entities such as persons, families, or in the case of land use, parcels. Such models simulate large representative populations of these low-level entities in order to draw conclusions that apply to higher levels of aggregation such as an entire country. Cellular Automata (CA) belong to the general family microsimulation models and are gaining favor in urban modeling. In the classic CA the study area is divided into an orthogonal grid of cells. The next state of each cell is determined by simple rules repetitively applied (transition rules), typically reliant only on the states of adjacent cells, termed the neighborhood.

Figure 1 illustrates a single iteration of a simple CA. For the sake of example we will define neighborhoods as being composed of the eight surrounding cells of any given cell. The 3x4 cell block contains two neighborhoods, one with central cell ‘A’ and one with central cell ‘B’. If we define the transition rule to be ‘the central cell changes to black if it has at least four black neighbors’, cell A retains its yellow color as a result of having fewer than four black neighbors. Conversely, cell B changes to black as its black neighboring cells do number four. Re-assembling the grid gives us the 3x4 output grid shown to the right, completing one iteration of our sample CA. Obviously, cell A transitions to black on the second iteration as black cells number four in both neighborhoods after the first iteration is processed, satisfying the transition rule’s condition.

CA’s neighborhood-centric approach poses significant promise for LUT modeling. It has been recently demonstrated through statistical analysis that the land uses of neighboring land cells do influence those of central cells (Arai and Akiyama, 2004). This confirms that the basic operating principles of CA may be able to effectively model the interaction between land use and transportation at the micro level, albeit with modification to the classic CA configuration illustrated in figure 1.

Since a local rule-based calculation lies at the heart of a CA it has exceptional ability to show emergence. And while it is true that more spatial parameters than the neighborhood statistics alone need be considered, GIS provides a ready means to

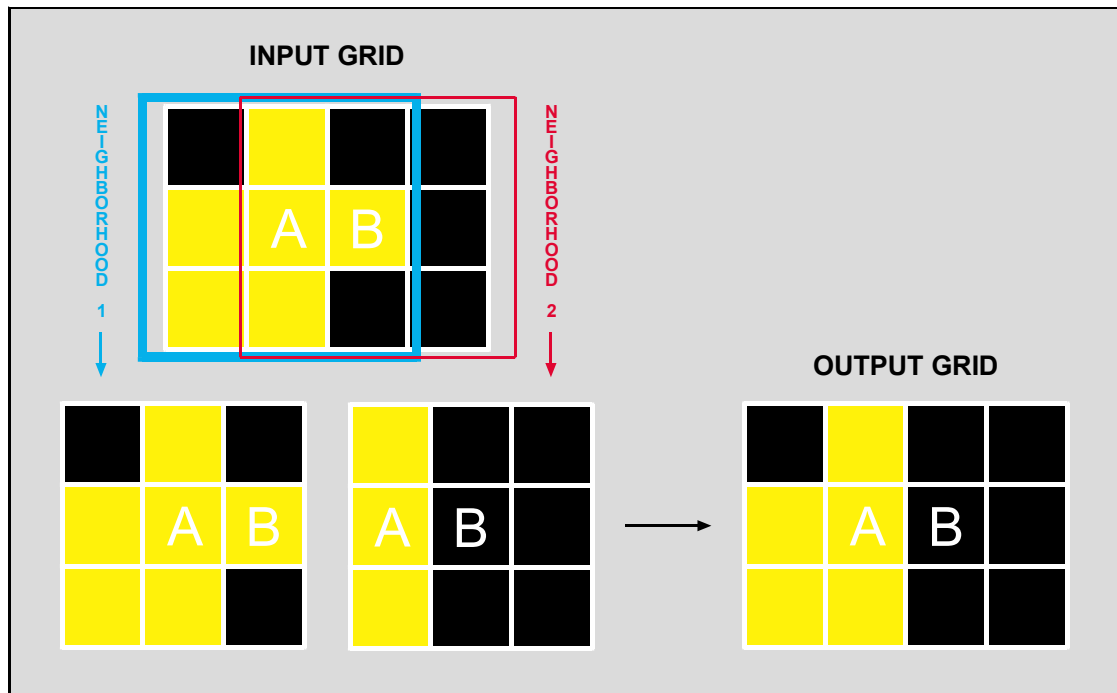


Figure 1 – One iteration of a cellular automata (CA)

generate these from existing raster data, vector data converted to raster form, or from distance calculations to vector features. CA models that consider input parameters beyond neighborhood statistics are referred to as *constrained CA* (B. Straatman et al., 2004).

The formulation and calibration of a CA-based LUT model also offers appeal as a form of exploratory data analysis (EDA). Calibration is dependant on historical data, and the formulation of a system that adequately replicates historical data emphasizes the patterns and trends present in that data. Additionally, manipulation of the historical data and/or model structure can test various assumptions made concerning the nature of the land-use change present and the responsible driving factors.

This paper details an experimental CA-based LUT model built to analyze historical land use and transportation structure at the local level. Historical land use data and road networks from a developing road corridor in southern Knoxville, TN were used to calibrate the model and assess its performance. The impact of various model inputs was explored, and implications regarding LUT model structure and historical LUT change in the study area noted. The general modeling limitations inherent in a simple model architecture with relatively few input parameters were also considered.

CHAPTER 2

LITERATURE REVIEW

1. Land Use—Transportation Modeling

Batty (1994) summarizes the origin and development of large-scale urban models by tracing the evolution of the modeling paradigm and actual resulting models from the late 1950s. He describes how computer models of land use and transportation (LUT) were first developed in the positivist age dominated by a sense that successes in science could extend to the entire realm of human experience. He follows the evolution of urban models through the decline of positivism, through the era of wholesale change that was the early 1970s, and the introduction of “normal science” to the domain. Batty goes on to describe the proliferation of comprehensive LUT models in the 1980s, the revolutionary impact of GIS in the 1990s and beyond, and the challenges inherent in implementing urban theory within a GIS. Most importantly, Batty makes the following observation, “For a hundred years or more, urban theorists have treated cities as though equilibrium were their natural condition. However, as current events increasingly demonstrate, this is less and less true.” (Batty, 1994, p. 12)

As suggested by the name, land use—transportation (LUT) models typically contain two separate systems, one considering land use and a second considering transportation. Linkages between these two systems vary, with a feedback mechanism advantageous to reflect the reciprocal nature of the land use—transportation relationship.

Figure 2 illustrates such a model, depicting an equilibrium-based land use component coupled to one containing a four-step travel demand model via feedback loops (Torrens, 2000a). These models are referred to as *composite* models, as opposed to *unified* models which have a tightly integrated structure (Shaw and Xin, 2003). The land use component of this sample LUT model employs an equilibrium device that seeks to balance the supply of land with its demand. Location and development factors are considered, but there is no guarantee that land use seeks any particular equilibrium state. Additionally, location and development factors do not influence transport directly, but only through the feedback mechanism linking the land use component to the transport one.

The transport component illustrated is the familiar four-step travel demand model. This spatial interaction approach divides an urban area into traffic analysis zones (TAZs), effectively limiting its spatial resolution to a fairly low level. Four sub-models addressing trip generation, trip distribution, modal split, and trip assignment to the road network are positioned sequentially (Miller and Shaw, 2001). Due to the composite nature of the model, changes in transportation have a limited ability to affect individual aspects of land use, such as the demand for or supply of developable land.

Traditional LUT models also generally employ sub-models to reflect at least location and land development in addition to the equilibrium mechanism responsible for balancing the supply and demand of land. Serious concerns arise from the limitations of the linkages between component models and sub-models. The sub-models of the land use and transport component models illustrated are loosely coupled, with a unidirectional information flow. The model illustrated makes assumptions about the nature of the

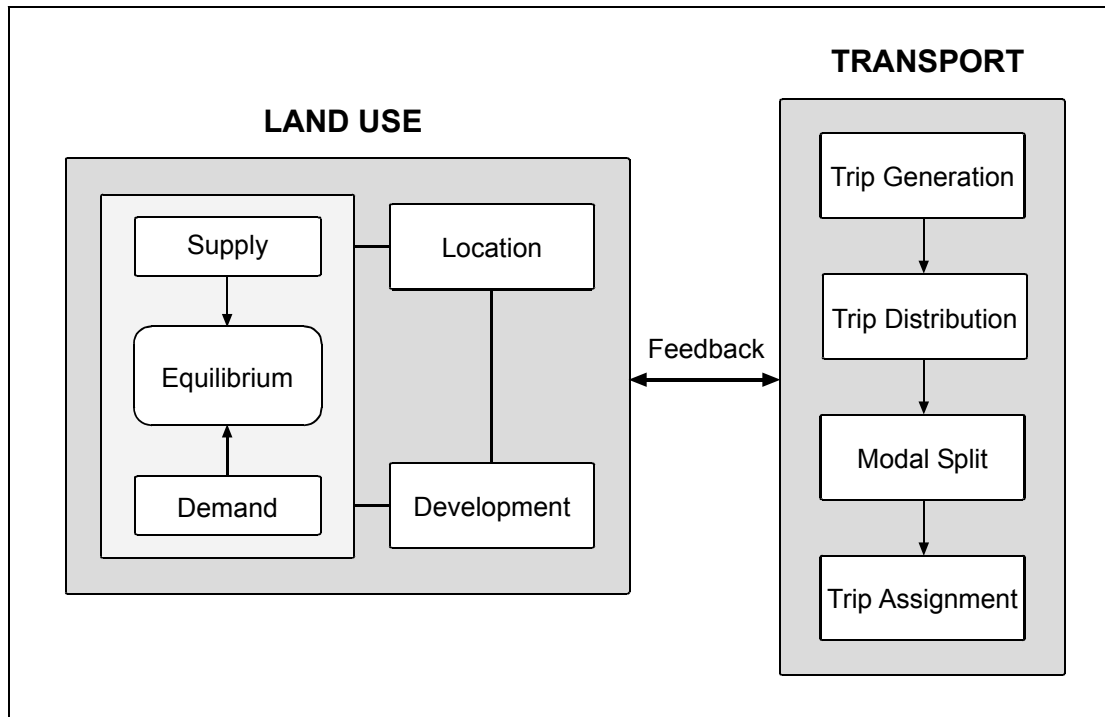


Figure 2 - The general structure of a land use—transportation (LUT) model

phenomenon being modeled, contain loosely coupled sub-models, and their generally low spatial resolution complicate the formulation of a feedback mechanism between the land use and transport components sufficient for the needs of the model (Torrens, 2000a).

Many historical examples exist of equilibrium-based LUT models (e.g. Alonso, 1964; Anas, 1982; Anas and Duann, 1986; Boyce, 1980, 1990; Hansen, 1959; Kim, 1983; Prastocos, 1986; Kim et al., 1989; and Hirschman and Henderson, 1990), and contemporary urban models often retain equilibrium concepts. Wegener (2004, pp.131-39) discusses 20 such models that have reached an operational status, having been tested, validated and used for policy analysis in actual metropolitan areas. Nearly all rely on some form of an equilibrium device, and six do not in themselves model transport but rather must be linked to existing transport models. And perhaps more importantly for the purpose of modeling interactive nature of the land use—transportation relationship, factors as large as transportation or land use may be treated as a fixed entity in the component model or models responsible for the calculation of the other.

Wegener notes a lack of spatial resolution in his survey of LUT models. Only three of the models discussed are disaggregate, the remaining aggregating urban entities into analysis zones. This approach, although necessary to make gravity and microeconomic-based models manageable, precludes exactly the type of resolution Wegener identifies as necessary to fully address the environmental concerns that in large part drive the development of such models. The typically aggregate nature Wegener reports also severely limits the ability of the model to substantially explore environmental impacts or to analyze issues of spatial equity. Most of the models seek solutions that

provide the greatest aggregate social good, which is of little use in assuring equity at the individual level and do not support the scale of analysis necessary for environmental study. However, Wegener does see the situation improving. More powerful computers allow the spatial resolution of models to be increased while preserving their computability. The advent of high-resolution spatial data promises to reduce aggregation error. GIS is identified as the mainstream data organization of future models. Also raised is the notion that aggregate probabilistic models will be replaced by disaggregate stochastic microsimulation ones, producing vastly more detailed models.

Microsimulation models operate at an individual level of entities such as persons, families, or in the case of land use, parcels. Such models simulate large representative populations of these low-level entities in order to draw conclusions that apply to higher levels of aggregation. First used in social science applications by Orcutt, et al. (1961), microsimulation techniques are gaining favor and have been adopted in three of the better-known LUT models: the US Department of Transportation's Transportation Analysis Simulation System (TRANSIMS), U.C. Santa Barbara's Slope, Land cover, Exclusion, Urbanization, Transportation, and Hillshade (SLEUTH) model, and ILUTE, the Integrated Land Use, Transportation, Environment modeling system (Salvini and Miller, 2003).

2. Microsimulation and CA—Beyond the Aggregate

Bennenson and Torrens (2004) describe spatially explicit microsimulation (or *geosimulation*) efforts in more detail. They characterize recent modeling efforts as a new wave of urban models influenced by microsimulation techniques such as cellular automata (CA) and agent-based systems. These recent models are more likely to employ individual-scale urban objects and extensively model the rules governing their behavior.

CA models were first suggested by English mathematician Alan Turing's Universal Turing Machine of the 1930s (Batty, 1997), essentially a one-dimensional CA. Two-dimensional CA were devised in the 1940s by John von Neumann, originator of game theory and pioneer in set theory, quantum mechanics, and electronic computers and Stanislaw Ulam, known for his work on Monte Carlo simulations and set theory (Torrens, 2000b). In the classic von Neuman-Ulam CA the study area is divided into an orthogonal grid of cells. The state of each cell is determined by simple rules repetitively applied (transition rules), typically reliant only on the states of adjacent cells. The selection of adjacent cells (neighborhood) can have a number of configurations, with classic CA generally formulating it as the eight surrounding cells (the Moore neighborhood), or an X-shaped configuration including the four corner surrounding cells (the von Neumann neighborhood). Because of the usual lattice-like arrangement of cell positions, CA are said to have a "natural affinity" with raster data (Couclelis, 1997). This factor produces considerable operational appeal as raster overlay computation is a core component of essentially all commercial GIS software packages.

Torrens (2000b) details the operation and suitability of cellular models for urban systems. He notes the changes in urban structure that occurred between the late 1950s era in which LUT modeling originated and the present. The essentially monocentric urban forms that most urban models presume are an extreme rarity today as cities have

spawned additional nuclei that assume CBD-like functions. This increases the importance of suburbs in the function of the city as a whole, and consequently urban modeling efforts have become outmoded in many cases. In addition to the obviously decreasing effectiveness of urban models, the early 1970s brought critical analyses of urban modeling that focused on the shortcomings of the construct of existing urban models as well as their theoretical foundation. Lee's seminal *Requiem for Large-Scale Models* (1973) received considerable notoriety. He criticized their expense, data needs, hyper-comprehensiveness, mechanical organization, resolution, transparency, dynamics, and inability to replicate their results. Some critics characterized the models of the day as error-ridden and failing to advance theory while simultaneously falling short of informing practice, and others pointed to their reliance on an assumption of predictability as a critical weakness. "CA models, while very much less than perfect, do address many of these concerns and deficiencies. In some key areas, CA models represent a significant improvement on previous generations of urban simulation models: spatiality, decentralization, affinity with new techniques for spatial analysis, attention to detail, linking function and form, dynamics, theory, simplicity, connection of micro- and macro-approaches, and visualization" (Torrens, 2000b, p. 35).

Spatial phenomena are well-represented by CA. All models are abstractions of reality, but traditional LUT models' typically aggregate approaches tend to divorce a great degree of spatial detail. CA, on the other hand, make implicit use of spatial complexity (White, Engelen, and Uljee, 1997). Beyond the resolution advantages generally enjoyed by microsimulation approaches, a well-implemented CA can better represent the actual spatial distribution of a city's elements at a given resolution level due to the absence of error inherent in aggregation and disaggregation processes.

CA promise significantly greater attention to detail. Traditional methods attempt to compute average solutions, reliant on the assumption that this average behavior is the simple sum of average individual behaviors. This is problematic to the simulation of non-linear behavior. The ability of CA to handle individual-scale dynamics while preserving computability offers the opportunity to implement more detailed models (Torrens, 2000b).

Dynamics are well represented by CA. Traditional urban models tend to treat time only through cross-sectional analyses at a limited number of discrete intervals. The jump from one interval to the next may represent years—long enough for considerable urban change to occur. This dynamic weakness is often the byproduct of calibrating the models via census demographic figures, which are collected only periodically, often every 10 years. Even if the time period between intervals is reduced, the model still is an essentially static device, as it does not proceed from one interval to the next in an iterative fashion, rather using them as a schedule for cross-sectional updates. CA models treat time much more realistically. CA models move in time intervals, but the steps are truly iterative in that the output of one model state serves as the input for the next, allowing global patterns to emerge from local ones. The interval may be decreased to as short a period of time as the dynamics of the phenomenon being modeled require. This gives CA the potential to contribute to the understanding of the evolution of urban forms and structures (Li and Yeh, 2002).

Visualization is a prime CA advantage. “CA are, by their very nature, a highly visual environment for simulation” (Torrens, 2000b, p.41). This creates user interest, as well as allowing visual interaction with the model. Identification of pattern is greatly enhanced through graphic representation, and because the model is visually dynamic the potential for clearly portraying the emergence of pattern is improved.

One potential advantage of CA models, both in comparison to other microsimulation techniques such as agent-based modeling and genetic algorithm schemes as well as equilibrium-oriented models, is simplicity. Traditional LUT models that have reached operational status tend towards extreme complexity. CA radically simplifies the model’s design. Transition rules can be derived from theoretically informed research or numerically derived from historical data and complexity is allowed to emerge from the model’s operation rather than describe the model’s structure. The simplicity of CA can be viewed as a double-edged sword as well. All models are an abstraction of reality, and as such run the risk of de-emphasizing or omitting important aspects of the complex systems they are intended to represent. “No models based on toy values and the homogeneity, uniformity, universality, etc, assumptions of classic CA can have a claim to the status of explanatory tools for real-world applications” (Couclelis, 1997, p.167).

The classic Neumann CA is too simplified and constrained to serve as an effective urban model (Torrens, 2000b). To ameliorate these concerns, Torrens details modifications to the classic Neumann CA for urban applications. No CA-based model needs pursue every possible modification (and doing so generally sacrifices the simplicity inherent in CA), but any will need some of these modifications if CA is to effectively model the specific phenomenon of urban interest. In classic CA cells adopt states from a range of like elements. It is possible to define CA such that cells can exhibit states of different forms, perhaps some binary and others integer.

A particularly compelling evolution of this idea is that of cell-state fixture (White and Engelen, 1997). Cells are regarded as ‘fixed’ or ‘functional’. Fixed cells might represent undevelopable areas, such as those covered by a body of water or other largely static element. Functional cells would form the active region of the CA, essentially deviating from the classic rectangular lattice to a spatial form that better approximates the urban area of application. This ability to be easily adapted to differing urban areas is a huge departure from the isotropic-assumptive approach of early urban models.

Neighborhoods must also be adapted. The reliance on classic CA’s entirely adjacent neighborhood definition precludes distance influences, allowing them only to propagate through the intervening cells. It is possible to introduce distance-decay effects, or apply weights to neighborhoods in the transition calculation. Neighborhoods can be extended to comprise larger spaces (e.g. White and Engelen, 1997; White, Engelen, and Uljee, 1997). And just as ‘fixed’ and ‘functional’ states can serve to remove areas from the lattice, they can also serve to remove them from neighborhoods (Torrens, 2000b).

Perhaps the biggest arena for modifications to the classic CA form is that of the transition rule. Transition rules in CA are responsible for implementing real-world behavior in the artificial world of the CA, essentially explaining how cities work (Torrens, 2000b). Classic CA use very basic deterministic transition rules, often a simple calculation based on the states of cells that constitute the neighborhood. Adapting CA to urban modeling demands serious fortification of transition rules if real-world behavior

and relationships are to be faithfully reflected. They might be reformulated as probabilistic expressions, introducing a randomness or ‘noise’. Transition rules can also be very complex mathematically or adopt methods from other simulation techniques. Genetic algorithms (see Mitchell, 1998) can be employed to give transition rules the ability to modify themselves (e.g. Colonna, et al., 1998). Economic principles such as utility maximization can also be incorporated into transition rules (e.g. Webster and Wu, 1999), as well as accessibility algorithms based on spatial interaction (e.g. White, Engelen, and Uljee, 1997). Recent CA-based urban models generally consider physical factors such as slope or soil type as well as distances calculated from geographic entities such as roads or urban centers in addition to a classic neighborhood analysis (Li and Yeh, 2001).

3. Artificial Neural Networks and CA

The responsibility given transition rules for implementing behavior creates the need for rigorous model calibration. The predicted cell states for a given CA are compared to known historical data, and the transition rules adjusted to calibrate the model. Calibration can take many forms, with brute force computation (e.g. SLUETH) and artificial neural networks (Clarke, Guan, and Wang, 2005; Li and Yeh, 2001, 2002) being workable approaches. An artificial neural network (ANN) is an artificial intelligence tool that identifies arbitrary and non-linear functions directly from experimental data (Almeida, 2002).

ANNs are adept at dealing with noisy and voluminous data. ANNs are formed from multiple nodes, or *neurons*, producing a network structure intended to mimic the operating principles of the human mind (Miller and Shaw, 2001). Such a network can be “trained” with historical data and “learn” the relationships present. Once trained, the network can be used to estimate conversion probabilities to the various cell states supported by the model with the maximum value determining the next cell-state rather than the output of a theory-based transition (Li and Yeh, 2001, 2002). CA models that use ANNs to process transitions are referred to as *ANN-CA*.

Brute force methods require substantial computational resources, and while neural network-based calibration offers potential relief it also entails considerable design effort and can be nearly as computationally-intensive if not implemented carefully. Although neural networks have substantial potential for automated and versatile calibration by dictating the entire transition, they in doing so turn CA into a ‘black-box’ modeling environment. The form of a theory-based transition rules can inform as to the nature of the process the CA models—if it can be calibrated. Deriving the transition rule numerically via an ANN removes this possibility, but that is not a significant drawback in all application domains. ANN-derived transitions are ideal for models built primarily to explore data, create visualizations, or perform scenario testing due to the ability to generate a new transition for various input data or model structures (Li and Yeh, 2001).

4. Recent land use—Transportation CA Models

A review of recent CA LUT implementations reveals a variety of approaches. Colonna, et al. (1998) developed a custom CA for the city of Rome, Italy that incorporates a genetic algorithm in a ‘new rules generator’ that invents possible new rules to be evaluated from the existing. Evaluation takes place in a separate sub-model called the evaluation system. This system is essentially self-calibrating, with the evaluation system serving as an internal calibration device. This custom CA model deviates from the classic CA’s orthogonal grid, allowing a CA in which cells of any shape can be assembled in any way. The classic CA’s adjacency neighborhoods are supplemented with continuous measure of distance between cells to vary their at-a-distance influence.

Candau, Clarke, and Rasmussen (2000) and Clarke and Silva (2004) used Keith C. Clarke’s Slope, Land cover, Exclusion, Urban, Transportation, and Hillshade (SLUETH) urban model (see Clarke, 1997, 1998). It was applied to the Mid-Atlantic Integrated Assessment (MAIA) study area and the metropolitan areas of Lisbon and Porto, Portugal, respectively. This popular model couples two CA and calibrates for historical time sequences using geocomputational methods (Silva and Clarke, 2002). One CA is responsible for the urban growth model and is coupled with the Deltratron Land use/Land Cover Model (DLM). Although self-modification of transition rules is permitted, the system has five transition rule parameters which must be optimized through brute force at considerable computational expense. The model does produce impressive results after calibration.

Li and Yeh (2001) present a CA model using neural networks to simulate potential or alternative urban development patterns based on different planning objectives in Dongguan, China. Transition rules are determined by an artificial neural network with state changes exported to a GIS. Li and Yeh (2002) go on to create an ANN-based CA that simulates competing multiple land uses. These models have the advantage that they calibrate themselves without human intervention. This strategy removes subjective judgment from the formulation of transition rules, but does introduce the ‘black-box’ effect previously noted. For explanatory models this is a serious drawback, but for uses such as the exploration of the patterns of land use wrought by changes to the infrastructure (or vice-versa), this is not found to be a major consideration.

Blecic, et al., (2004) create an advanced CA model for the city of Heraklion (Crete), Greece with the Cellular Automata General Environment (CAGE). CAGE provides for many advanced modifications to the classic CA formulation. These include vertical neighborhoods to allow separate model phenomenon interacting within the same physical scenario to be modeled. Transition rules can be specified at different spatial levels and can change with space and time, and can also depend on a local cell’s parameters as well as on global constraints and variables. Neighborhoods can be defined as abstract sets of cells satisfying abstract conditions and queries, and graphical representations of cells are considered attributes whose particular appearance is not subject to the constraints of spatial or temporal regularity. The CAGE development environment offers an exciting array of advanced modifications to the classic CA formulation, but calibration is not discussed in this paper.

Almeida, et al., (2005) applies the Centre for Remote Sensing of the Federal University of Minas Gerais's DINAMICA CA urban model to Bauru, Brazil. This model is based on stochastic transition algorithms and employs a different calibration technique using empirical procedures. A visual comparative analysis is employed for each type of land use change amongst general trends of preliminary simulation results and transition probability and land use transition maps created.

A sophisticated LUT model structure is presented by Guan, Clarke, and Wang (2005). This urban-growth CA of Beijing uses an artificial neural network (ANN) for internal calibration, easing the computational demands exemplified in SLUETH. The resulting ANN-Urban-CA model integrates a two-layer back-propagation (BP) neural network which estimates the probability of each cell transforming to a given land use type. This CA pairs its micro-scale geospatial model with a macro-scale socioeconomic sub-model to enhance forecast functionality.

CHAPTER III RESEARCH PLAN

1. Study Area

Land use and transportation interaction in the metro area of southwest Knox County, TN was explored. This formerly rural area south of Interstate 40 and west of US Highway 129 is now host to the southern extension of the Pellissippi Parkway, also known as US Interstate 140 (see figure 3). This major route connects Interstate 40 and the populous region of West Knoxville surrounding I-40 with US Highway 129 and the cities of Eagleton Village, Rockford, Maryville, and Alcoa that lie to the south. The corridor surrounding the new route has undergone considerable land-use change since the project's inception.

The I-140 construction project originated when officials from Blount County and the cities of Maryville and Alcoa asked the state of Tennessee to extend Pellissippi Parkway southward from its original southern terminus at I-40 in 1977. The originally proposed construction spanned 19.5 miles southward from I-40 to US 321 in eastern Blount County. By the end of 1992, the section from US 129 northward to S. Northshore Drive was complete, with the remaining distance from Northshore to I-40 finished in 1997. A small extension to the southern end was completed shortly thereafter, bringing the total length of I-140 to 11.17 miles (Center for Transportation Research, UTK 2003).

The I-140 designation stops at US 129, but an extension known as SR 162 was completed to SR 33 in 2001. Opposition from the Citizens Against the Pellissippi Parkway Extension (CAPPE) mounted a legal challenge to the completion of the project and in 2002 an injunction was granted to prohibit construction pending completion of an environmental impact study. CAPPE members cited concerns regarding sprawl, traffic, loss of farmland, and economic and environmental impacts (CAPPE, 2007). The project has been stalled since, despite the Tennessee Department of Transportation's (TDOT) September, 2004 announcement that an environmental impact statement was being prepared and that construction would continue upon its completion.

This study necessarily focuses on the northern 8.9 miles of I-140 that lie in Knox County, the segment between Interstate 40 and the Knox/Blount County line formed by the Tennessee River. This is due to the fact that parcel-level land use data is only available for Knox County. By 2002 on average nearly 39,000 vehicles traveled daily via I-140 to the Knox County border, with nearly 33,000 average daily trips between this point and the US 129 interchange (TDOT, 2002). A three-mile buffer around the 8.9 mile-long northern segment of I-140 defines the eastern and western bounds, with the northern bound established one mile north of and parallel to I-40.

This section of land offers an excellent opportunity to study land use and transportation interaction. This corridor was dominated by agricultural or undeveloped land usage until the imposition of one of the largest transportation infrastructure extensions seen in eastern Tennessee in the last three decades. The continued expansion of Knoxville to the north and Alcoa and Maryville to the south assures heavy utilization of the route. Land use conversion along its path is likely as the surrounding lands now

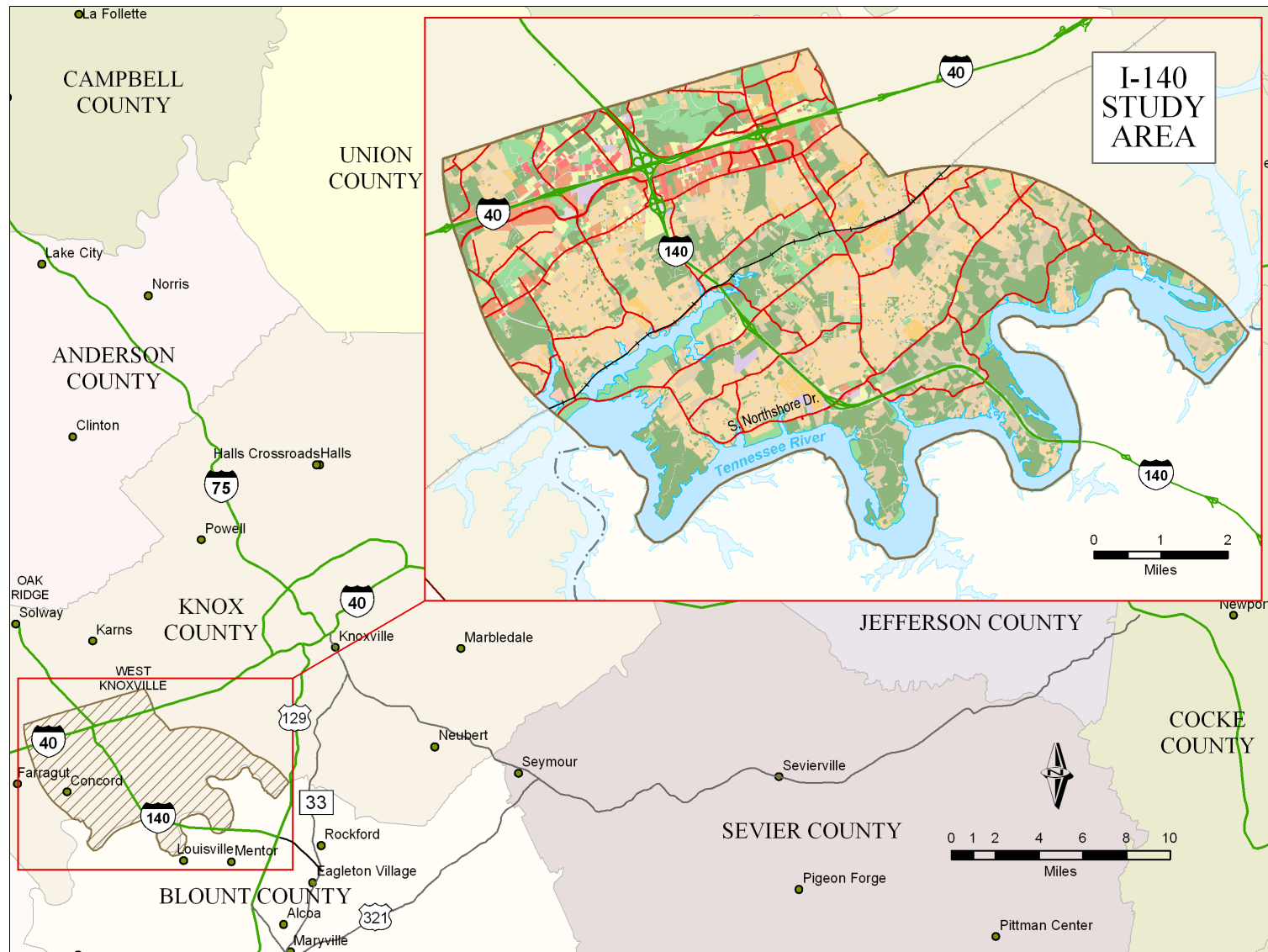


Figure 3 – Study area

enjoy excellent accessibility, with I-140 linking them efficiently to the opportunities found in the communities lying at its ends.

2. Project Data

Three primary datasets were used in this project: land use, road networks, and slope. The land use and road data were provided by the Knox County Metropolitan Planning Commission (MPC). This dataset was developed by the MPC in conjunction with the Knox County GIS (KGIS) using a best-source approach. Records from various public and private sources were used to define parcel-level polygons covering Knox County and assign a land use code to each.

Nearly complete data was available for the nominal years of 1993, 1999, and 2005. The supplied ESRI personal geodatabase feature classes were examined and corrected to make them more comparable. Slightly different classification schemes were used in the original data, but insight provided by the MPC was used to re-classify all three years to the scheme shown in Table 1. Rural residential parcels were defined in the original classification scheme as those covering an area of at least two acres, and this distinction was maintained. Class 5 includes transportation structures such as bus terminals or airports, but not the right-of-way surrounding roadways.

One anomaly affecting all three datasets was corrected, that being the classification of certain roadways. Road corridors residing on privately-owned land were not coded as such in the data, rather as the land use class of the surrounding parcels. The affected areas were compared to aerial photography and corrected. The resulting layers were converted to raster datasets with a pixel size of 50 feet, yielding overall dimensions of 1276 x 807 pixels.

The MPC also supplied road data for 1999 and 2006, with none being available for 1993. The roads were attributed with road name and class, with distinctions made for interstates, major, and local roads. A corresponding road layer for 1993 was constructed

Table 1 – Land Use Classes

LAND USE	DESCRIPTION
1	Agriculture/Forested/Vacant Land
2	Commercial
3	Industrial/Manufacturing
4	Multi-Family Residential
5	Transportation/Community/Utility
6	Single-Family Residential
7	Rural Residential
8	Road Corridor
9	Public Land/Parks/Private Recreation
10	Office
11	Under Construction
12	Water

from the 1999 one, with the appropriate segments removed via comparison with aerial photography and period road maps. The road class designation from the 1999 dataset was retained. Distances to the nearest road feature of each class were calculated for each cell and used as an additional input parameter for the CA.

As noted in the development of the SLEUTH model, slope is a key indicator for a parcel's suitability for development—to the degree that Candau, Rasmussen, and Clark (2000) state that "Slope above 21% cannot be urbanized." The USGS 10m resolution digital elevation models (DEMs) were obtained, mosaiced into a single layer, and output as an ESRI grid. From this layer a slope grid was produced retaining the 10m spatial resolution to be used as a CA input parameter.

3. Project Objectives

Several urban CA models using ANNs to process transitions have appeared (Li and Yeh 2001, 2002; Guan, Wang, and Clarke 2004, de Almeida and Gleriani, 2005). Although these models tread important ground in the design and implementation of constrained ANN-CA, they generally fail to deliver on CA's promise of adequate resolution for modeling of local interaction. These models use pixel sizes ranging from 50 to several hundred meters, a level of resolution inadequate to discern features such as a local road. These models also differ in approach, in that some include a sub-model. The 2004 Guan, Wang, and Clarke model utilizes a Tientenburg socio-economic model loosely coupled to an ANN-CA. The Li and Yeh and de Almeida and Gleriani models are heavily constrained, in that the CA used consider multiple distance and geographic factors in addition to a neighborhood analysis, but do not include sub-models to augment the CA. Like CA in general, variation amongst ANN-CA models is the rule. Little discussion exists in the literature as to the general limits of a LUT model using CA alone.

This project seeks to develop an ANN-CA LUT model for the study area described at the 50ft resolution of the produced land use grids, attempting to extend the advantages of ANN-CA models to a true local level of analysis. Road distance, present land use, land use in neighboring cells, and slope will be considered. The model will be calibrated and tested with historical land use data. The importance and effect of various model design features and model parameters will be investigated. The ability of the model to simulate historical land use change will be assessed, noting the implications regarding LUT model structure and historical LUT change in the study area. Through the adoption of a minimalist ANN-CA configuration the general limits of a simple constrained CA will be explored, clarifying the concerns that motivate more complex ANN-CA designs, such as those that use linked sub-models to augment their CA.

CHAPTER IV MODEL DESIGN AND CONTRUCTION

1. CA Transition Formulation

The relationships present in LUT models are complex by their very nature. Local and global considerations combine to shape the actions of individuals, and these individual actions sum to form the overall patterns of land and transportation use. LUT models that discern many multiple land use classes, as opposed to binary urban development models, increase complexity. By definition a binary model supports but two classes, greatly simplifying the necessitated model structure via the reduction of solution space. The 12 land use classes used in this study produce 144 possible land use conversions, increasing the complexity of the problem and requiring vastly greater computational ability from the model.

The simple CA illustrated in the introduction of this paper (see figure 1) has a very simple transition rule—that being that the central cell turns to black when the neighborhood count of black cells reaches four. While this transition rule was arbitrarily chosen, its selection is somewhat illustrative in that this is the primary means by which the transition rule for any theory-based CA is chosen. It is very difficult to know in advance exactly what a transition rule’s formulation should be, or even the variables it needs consider. While theory may inform the choice, more often considerable trial and error will be required to identify a workable transition formulation. It is necessary to calibrate the model with historical data and then verify its performance on unseen data. Success in verification confirms the viability of the transition rule used and input parameters considered.

The important implication of a theory-based rule is that its form can inform as to the process driving the patterns seen in the historical data. But there is no guarantee we can construct such a rule for a given phenomenon, and no example of an entirely theory-based CA urban model discerning multiple land use types was found in a survey of the literature. Due to the large solution space and highly non-linear associations present, the formulation of a CA’s transition is an excellent application for an artificial neural network (ANN). A universal definition of an ANN does not exist, with most attempts referring to physical structure or function of the network as much as defining the principle of operation. Perhaps the most relevant functional description of an ANN is that they are “associators”, in that they can associate a given input pattern with a given output (Abdi, 1994). The model built in this study attempts to make associations between the input parameters with the various land use conversions present in the historical data, complex though they may be.

2. Artificial Neural Networks

Considerably more agreement exists on the structure of neural networks than does on their definition. The basic building block is the *perceptron*, also known as a *processing element* (PE) or artificial neuron (see figure 4). The perceptron accepts multiple inputs, with a corresponding weight w adjusting the signal flow to the input

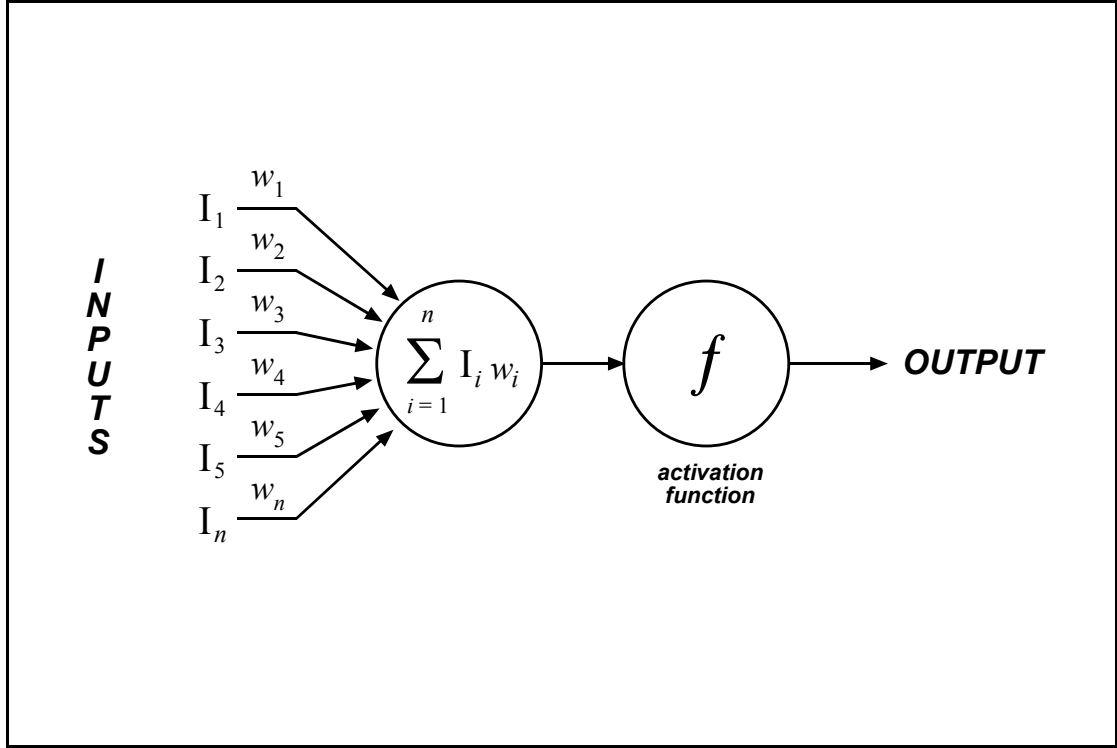


Figure 4 – The perceptron or ‘processing element’

node, where the weighted input signals are summed and used as the input for an activation function (Abdi, 1994). The resultant output is useful for solving simple problems, but is more often used as the input activation, or signal, for another perceptron in a network, creating a *feed-forward* ANN. A variety of functions may be used, with the sigmoid or logistic (Equation 1) and tanh (Equation 2) being the most common.

$$y = \frac{1}{1 + e^{-x}} \quad (1)$$

$$y = \tanh(x) \quad (2)$$

The sigmoid function spans a range of [0, 1], and the inputs are scaled to that range for use with it. Similarly, the inputs for a network using the tanh activation function will be scaled to a range of [-1, 1] to correspond with the range of the tanh function.

Figure 5 shows a number of PEs arranged in a multilayer network structure composed of three layers: input, hidden, and output. This general class of networks is known as a multi-layer perceptron (MLP). Why the increase in complexity? A perceptron can only discriminate linearly-separable functions, and it is difficult to calibrate its weights. The addition of a hidden layer or layers and an increase in the number of processing elements solves both problems (Abdi, 1994). With proper network

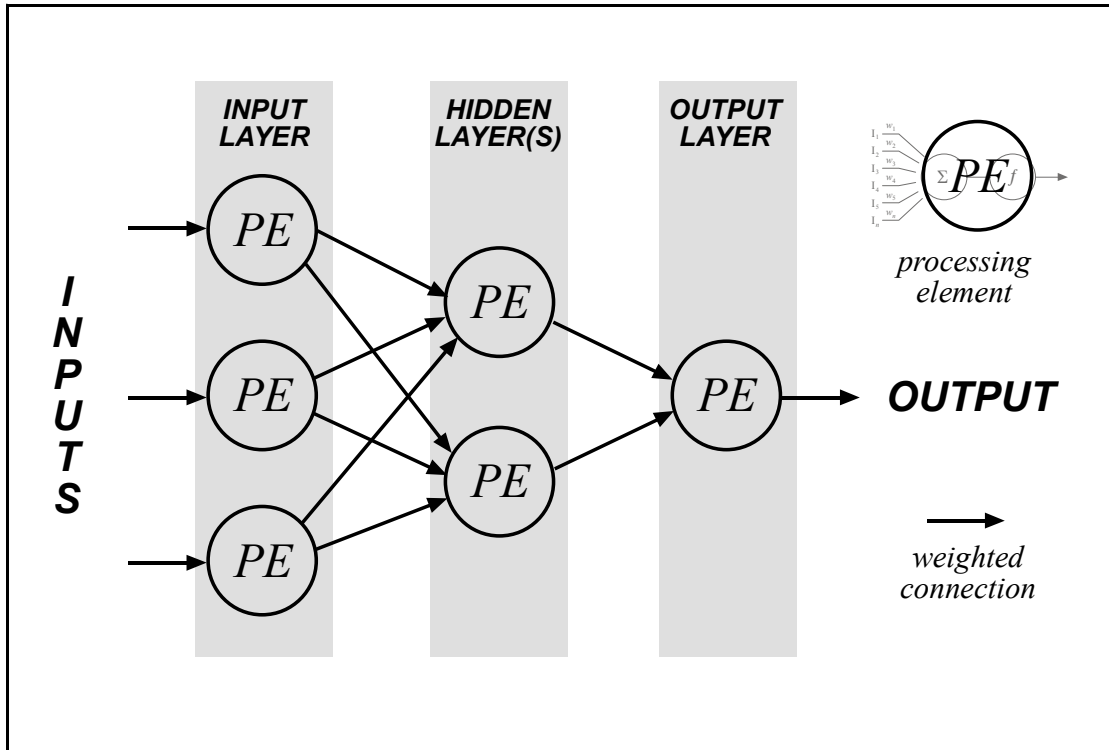


Figure 5 – Multi-Layer Perceptron (MLP)

configuration it is possible to represent any continuous function provided a sufficient number of inputs is provided (Li and Yeh, 2001). The additional weighted connections between layers also enable the implementation of the popular *back-propagation* (BP) training approach illustrated in figure 6.

BP training compares the actual output of the network to the desired output defined by historical data, and then adjusts the network weights to minimize this differential by propagating the error measurement back across the network. By processing many records from the historical data the network can “learn” the relationships present by adjusting its weights from an arbitrary initial selection to those that best allow replication of historical data. The relationships present in the training data are “learned” by the selection of weights that minimize error during training. The trained ANN is now fed data that it has not previously seen, with the output being the network’s estimation of the consequences of the initial conditions defined by that input data acted on by the relationships learned from the training data.

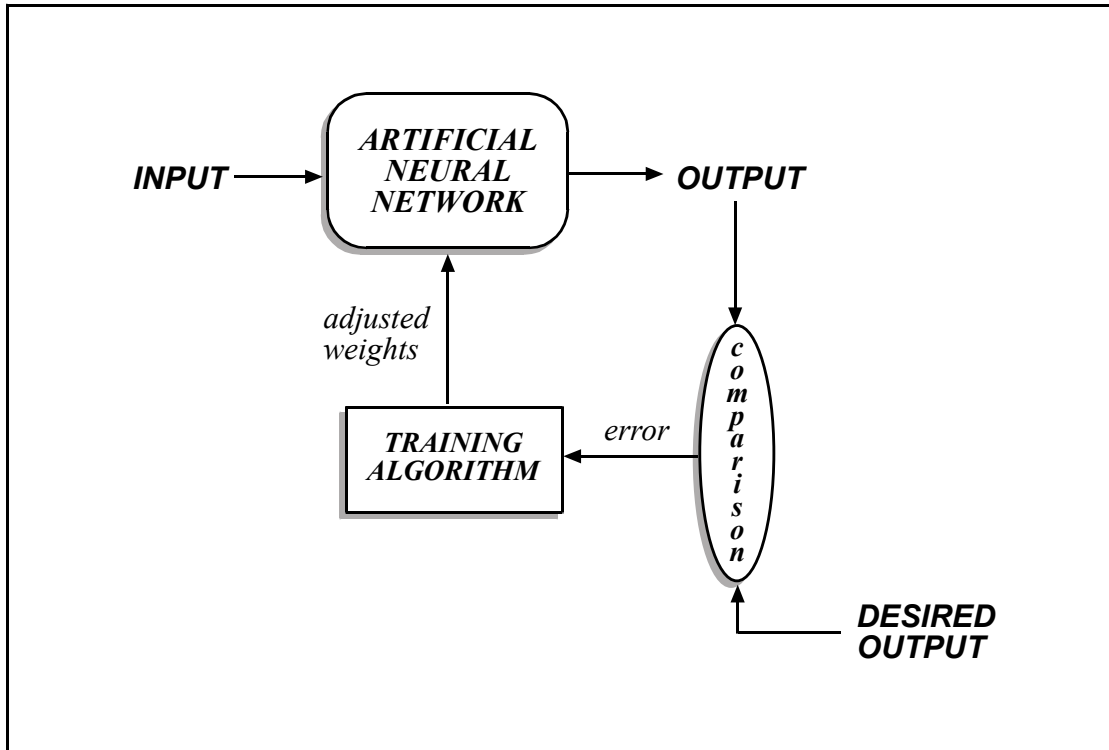


Figure 6 – Back-propagation (BP) training

3. Model Design

Table 2 shows the input parameters considered by the constructed ANN-CA model. Since the current land use of a given cell is one of 12 discrete values, this parameter is expanded to 12 binary indicators corresponding to the 12 possible land use classes. The binary indicator corresponding to a given class will contain a one, the rest zeros. This presents the ANN with 28 actual inputs—12 for the categorical land use value of a given cell, 12 for the cell counts of each of the 12 classes neighboring that cell,

Table 2 – Model input parameters

PARAMETER	DESCRIPTION
slope	slope derived from USGS 10m DEMs
NEAR_I	distance to nearest interstate
NEAR_MR	distance to nearest major road
NEAR_LR	distance to nearest local road
LU_VAL	current land use of cell
N1-N12 (12 total)	cell count of each land use class in a ¼ mile radius

and four for the slope and road-distance parameters. Different ANN configurations were tested through variation in the number of PEs and hidden layers and the choice of activation function used, with two hidden layers and the tanh activation function producing the best simulation performance.

Upon completion of training, the network is presented with the initial condition of the input parameters of the historical data, cell by cell. These are generated from the 1993 land use grid and road layers in this case. The ANN returns a corresponding list of conversion probabilities, 12 for each cell. These probabilities indicate the network's estimation of the likelihood of land use conversion to each of the 12 classes. The greatest of these values is the network's prediction of the successive land use for a given cell. These maximum probabilities are compared against a threshold value used to control the rate of conversion in the model via limiting the number of converted cells per iteration. This threshold value is experimentally chosen relative to the goals of the model, with values in the 0.7-0.8 range often useful given adequate ANN performance (Li and Yeh, 2002). Cells with maximum probabilities lower than the threshold retain their existing land use values.

A threshold of 0.7 was used in the iterative experiments in this study. Once the new cell states are determined, a new land use grid can be constructed from them, completing the first iteration of the model. The second iteration restarts the process, re-sampling the input parameters, sending them to the ANN for evaluation, resulting in another output land use grid.

4. Model Construction

Training data were sampled via ESRI's ArcGIS software using a random, stratified strategy. Two thousand point samples were selected from each of the 11 land use groups. The 12th group, water, was assumed un-developable and was left out of the training data for this reason. It was, however, included in the neighborhood counts to inform the network of the proximity to water enjoyed by a given parcel. After trimming the sample points that yielded no data on one of the various layers, 19,556 samples were retained. ArcGIS's 'extract values to points' tool was used to sample the slope and land use grids. Since ArcGIS's built-in tools have no local functions that permit cell counts within a neighborhood, the "landscape characterization" function of the freeware ArcGIS 9x extension Hawth's Tools was used. This extension also contains a point-intersection tool that proved faster than ArcGIS's 'extract values to points' tool for grid sampling.

The ANN development package selected was NeuroDimensions' NeuroSolutions 5. This fully featured ANN development environment eases network design, training, and production via its advanced graphical interface and various design aids such as automated normalization of inputs, expansion of symbolic variables, and randomization of the records in the training data. It does, however, accept only text files as inputs, and the header format differs from that of the text table outputs from ArcGIS. A Microsoft Access database was used for various data processing tasks related to the data exchanges between the GIS and the ANN development environment.

After the training was completed, input for the network was created by repeating the parameter collection routines in the GIS for every non-water cell in the grid, as

opposed to only those selected for training. At the 50ft. cell-size used, this resulted in nearly 460,000 samples. Once the initial state data was generated it was processed by the ANN, and the resultant conversion probabilities evaluated in ArcGIS using the VBA pre-logic function in the field calculator. Scripts were written to select the maximum conversion probability, determine the land use class that probability corresponded to, and filter the indicated land use with respect to the threshold chosen. After the new land use for each cell was determined, ArcGIS's 'feature to raster' tool was used to quickly convert the values into an output land use grid, completing the iteration.

5. Training Performance and Network Complexity

Standard statistical measures are used to assess the performance and terminate the training of an ANN at an appropriate place. The most succinct of these measures is the correlation measure r calculated from a comparison of the network's predicted output to the desired response as defined by historical data. There is no established optimal structure for ANNs, and the precise configuration used varies greatly with application. In general, it is best to use the least complex neural network possible to improve comput-

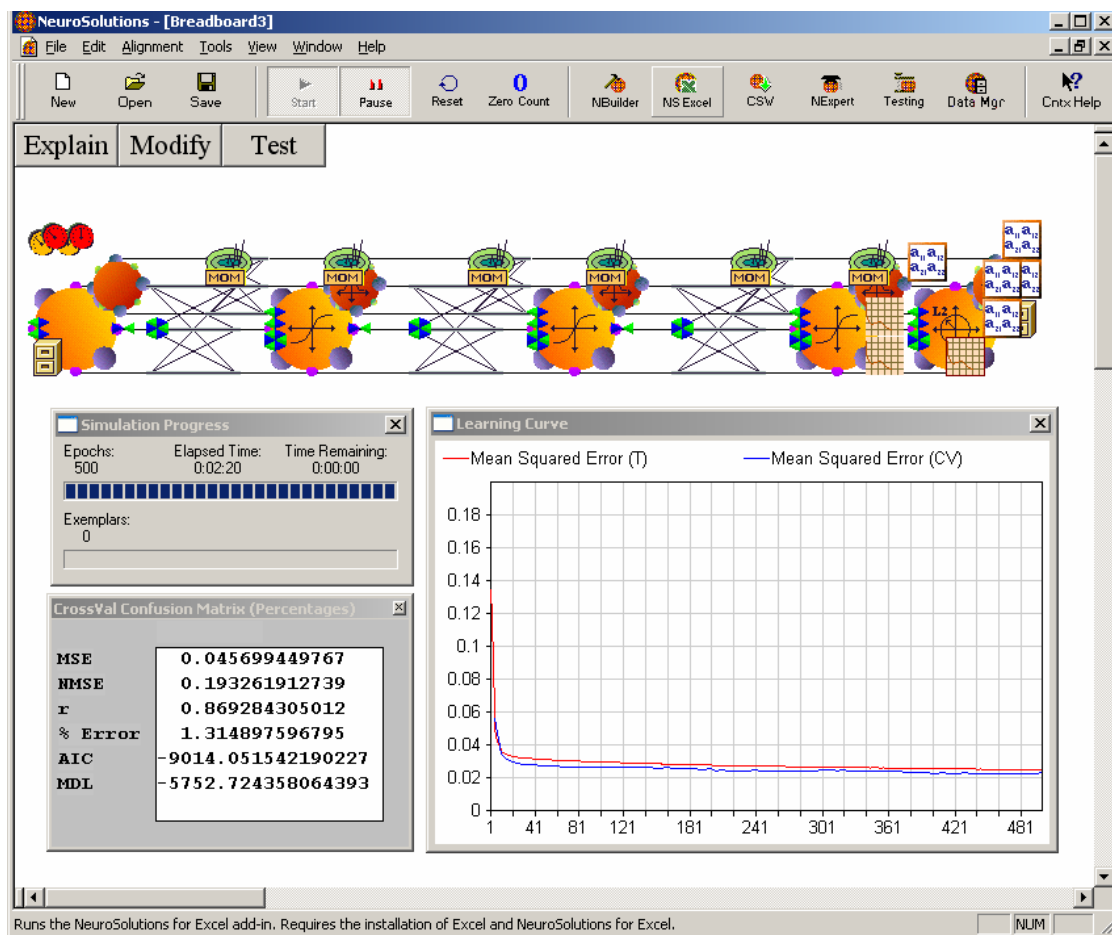


Figure 7 – ANN training results

ability and preserve the greatest degree of generalization possible (Li and Yeh, 2002). For this reason, initial training attempts were conducted with a very simple one hidden-layer design using the sigmoid activation. Training results with this configuration proved poor, struggling to reach an r value of 0.2. Analysis of the output data revealed the network was only able to assign probabilities above 0.7 for cells that did not change in the historical data, indicating it was not “learning” anything regarding the patterns of change. The relationships are too faint or complex in the historical data to be captured by the simple ANN network topology used.

The network topology was increased in complexity twice, arriving at a two hidden-layer topology with 30 PEs in the first hidden layer and 15 in the second. The tanh activation function was used. As shown in figure 7, this topology was able to capture the complex relationships present, achieving an r value of nearly 0.87. All further experiments used this network structure, or a variant thereof adjusted for experiments with fewer input parameters.

6. Computability Concerns

The 50ft. cell size used was chosen to allow the model to discern individual land parcels and streets. As noted above, this generated nearly a half-million grid points for the study area—raising computability concerns. While the computational resources needed to prepare land use grids, create the training sets, train the network, and process the production sets are minimal, the calculation of the road distances and collection of the point samples for each of a half-million points is demanding. Using a powerful custom workstation PC, collection of the road distance measures for each cell in the study area takes several hours, and the extraction of the point values associated with each iteration of the model consumes nearly an hour.

Although CA can model large-scale interaction and global patterns simultaneously, there are considerable practical performance limitations to the application of this approach—sure to be eased as technology continually provides faster computational platforms. Further discussion of measures that might be taken to ease computability concerns are found in Chapter VI of this paper.

CHAPTER V RESULTS

1. Historical Land Use Change

Although 12 land use classes were preserved in the historical data, more notable change inevitably occurs in some classes than others. Table 3 summarizes the absolute and percent change for each land use class between the 1993 and 2005 datasets. The ‘trans/communication/utilities’ class includes all transportation structures that are not roads, such as bus terminals or airports. Due to the relative lack of change and/or samples, specific performance regarding the ‘under construction’, trans/communication/utilities’, and the ‘industrial/manufacturing’ categories will not be discussed. Although the water category did vary slightly in the historical data, water will be treated as a static quantity for the purposes of the model. The sample points that have land use value of water in any input layer or layers are excluded from the simulation, implementing a “fixed” cell state modification to the CA.

An inconsistency in the classification schemes used in the historical data complicates comparison of performance regarding the rural residential class. Although the data supplier defined rural residential as residential parcels greater in area than two acres, the 1993 dataset contains many parcels smaller than two acres marked rural residential. This inconsistency was discovered but time did not permit correction of the input data for the purposes of this paper. Modeling performance for the remaining six land use classes will be detailed later in this chapter, with maps of the models’ predicted output overlying maps of the historical data for each land use class.

Table 3 – Historical Land Use Change

LAND USE	1993	2005	DIFFERENCE	% CHANGE
Single-Family Residential	95511	125414	29903	31.3%
Parks/Public/Recreation	20589	34584	13995	68.0%
Road Corridor	48534	59506	10972	22.6%
Office	6074	14147	8073	132.9%
Multi-Family Residential	8961	15925	6964	77.7%
Commercial	13252	20068	6816	51.4%
Rural Residential	42993	47265	4272	9.9%
Under Construction	2791	3386	595	21.3%
Trans/Com/Utilities	2308	2653	345	14.9%
Water	86250	85913	-337	-0.4%
Industrial/Manufacturing	3376	2984	-392	-11.6%
Ag/Forest/Vacant	215231	133828	-81403	-37.8%
TOTAL	545870	545673	-197	-0.04%

2. Test Configurations

Four model configurations were evaluated, varying in their choice of input parameters or initial land use grids. All configurations consider slope, distance to interstates, distance to major roads, and the cell counts of each land use type present in a quarter-mile circular radius from each cell's center. The first configuration considers the existent 1993 local roads, and allows the model to convert cells to the 'road corridor' land use class—the only configuration to do so. The second configuration is the same, except conversion to the road corridor class is prohibited. The third configuration does not consider local roads, and no conversion to road corridor is permitted. The final configuration imposes the full 2005 road network on the 1993 land use as the initial condition rather than using the 1993 road network as do the first three configurations. The first configuration was used only to demonstrate road corridor conversion limitations. The first-iteration performance in the other three configurations (henceforth referred to as 'with 1993 local roads', 'no local roads', and 'with 2005 local roads') will be detailed for the six selected land use classes identified in Section 1 of this chapter by comparing selected maps of the ANN predicted output for a land use class with one of the same class from the 2005 historical data, as well as characterized statistically.

3. Road Corridors

Due to the cellular organization inherent in CA and the limited input parameters considered by the model, it is not possible to effectively model the underlying network structure that translates to the road corridor cells in a given land use grid. Network representation in GIS typically is predicated on the presence of nodes and connections between those nodes, and these structures do not exist in CA. Although the output of the ANN may indicate a suitable conversion probability to the road corridor class, the cell in question will only by chance fall as an extension to the existing road network, as opposed to exhibiting an apparently random placement. Figure 8 illustrates a close-up of the I-40/I140 interchange area. Note the orange cells; they do in fact change to road corridors by 2005. The ANN also indicates the pink cells transition to road corridor class, but this is wholly incorrect. Interestingly, several cells line up to form an apparent on-ramp to I-40, but these are just placements that satisfy the parameter combinations identified in the historical data rather than an attempt to construct linear features, much less ones part of a network.

Various adjustments to the input data were tried in an attempt to correct for this factor. The first variation was to simply not allow the model to convert cells to the road corridor class. Should the ANN dictate conversion to road corridor, the model would defer to the land use class with the second-highest conversion probability. And although this proved a workable adjustment for the first iteration, continued iteration of the model proved troublesome (see Section 4). The second method tested was to eliminate the local roads in the model due to the fact that they contain the vast majority of the yearly road change in the study area for this period. This led to a loss of model accuracy for some land use classes, as well as its own iteration side-effects. The third adjustment tried was to superimpose the complete 2005 road corridor class on the 1993 land use data as the

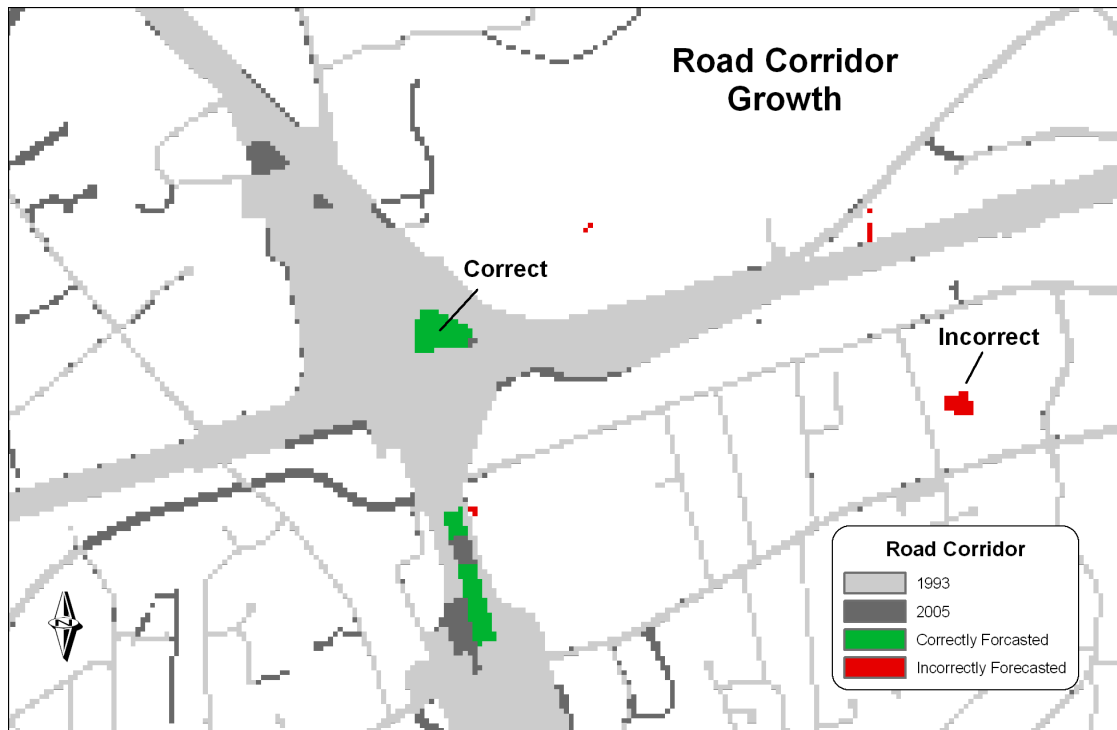


Figure 8 – Road Corridor Conversion

initial condition for the iterative process, with road distance figured from the 2005 road features. This eased, but did not eliminate, the iteration problems and introduced a skew to the modeling performance in some classes. A vector road growth sub-model could produce accurate simulation results over multiple iterations of the model, as the road network and attendant road corridors must evolve gradually and realistically in order for the model's training to be valid across multiple iterations.

A simpler scheme would be to establish the date of construction of the road features. From this data one could advance the road network to an appropriate level for each successive iteration of the model. This approach is more data intensive, but will produce more accurate simulation results through the elimination of the error inherent in a sub-model responsible for the estimation of road growth. The road data supplied did not establish a date of construction for the various road features. Period road maps and aerial photography were sought to establish the construction dates of the roads, but those found to exist for the area were too few in number to establish a date estimation for the features more finely-grained than the 6 years indicated by the nominal dates of the road datasets. This technique is not applicable to forecast usage, however, and for this reason as well as time constraints was not pursued.

4. Iteration

Two difficulties were discovered when multiple iterations were produced. First, as noted previously, there is no way to correctly transition road corridor cells and advance the road features. A skew in the neighborhood counts for the road corridor class is introduced and the road-distance parameters are left static in succeeding iterations. This has the effect of limiting conversion overall, as nearness to road features is indicated in the historical data for cells that transition to many land use classes. Most of the cells that lie appropriately close to roads convert on the first iteration, leaving fewer that meet those same conditions on the succeeding iterations. Consequently, the conversion that does happen in later iterations has an unnatural focus on areas whose road network is already saturated. Figure 9 shows close-ups of the same sections of the 1993 and 1999 historical data. Note the development in the areas marked A and B. In the first iteration from 1993, the balance of development that exists between them resembles that seen in the historical data. By the second iteration, however, development has stalled in area B but accelerated in a circular cluster in area A. These crops are taken from the configuration using the 2005 roads imposed as the initial condition, nevertheless, development is skewed towards area A. This is because area A better satisfies the conditions for development noted in the historical data—nearness to roads and high neighborhood counts of developed land types. Although the complete 2005 road network is given as the initial condition, the low neighborhood counts of developed land use types in area B leave it less highly rated for development by the model, a disadvantage that is maintained as iteration continues. This same effect was noted to a greater degree in the ‘with 1993 local roads’ configuration—where the initial local roads are 1993’s. Nearly all the development present in the second iteration was located in area A.

The second difficulty noted was the presence of glitches in the output of the model. One might refer to them as graphical glitches, but in this case, of course, they reveal errors in the output of the model. No cure was found, but since the difficulties were proved sensitive to the exact stopping point of the ANN training it would appear that they are a by-product of the complexity of the ANN used. Li and Yeh (2001, 2002) state that simpler ANNs are suitable for iterative models, but do not specify the exact root of this recommendation. Simpler ANNs do have a greater ability to generalize (Abdi, 1994) though, and this factor would reduce the sensitivity to training variation. Figure 10 shows a horizontal glitch visible as a line running horizontally across the land mass. No glitches were encountered on the first iteration of any network, but some were visible in the second iteration’s output of all but one configuration. No explanation was found for the exact location of the affected areas. This is additional impetus for the adoption of additional input parameters and/or sub-models to augment the CA, as the relationships present in the slope, road distance, and neighborhood factors used here are faint and complex enough as to mandate a very complex ANN to discern them.

In spite of the difficulties encountered, ability of the model to generate emergent growth patterns is promising. Figure 11 shows a section of the first three iterations of a simulation. Note the development of the single-family residential mass (shown as light

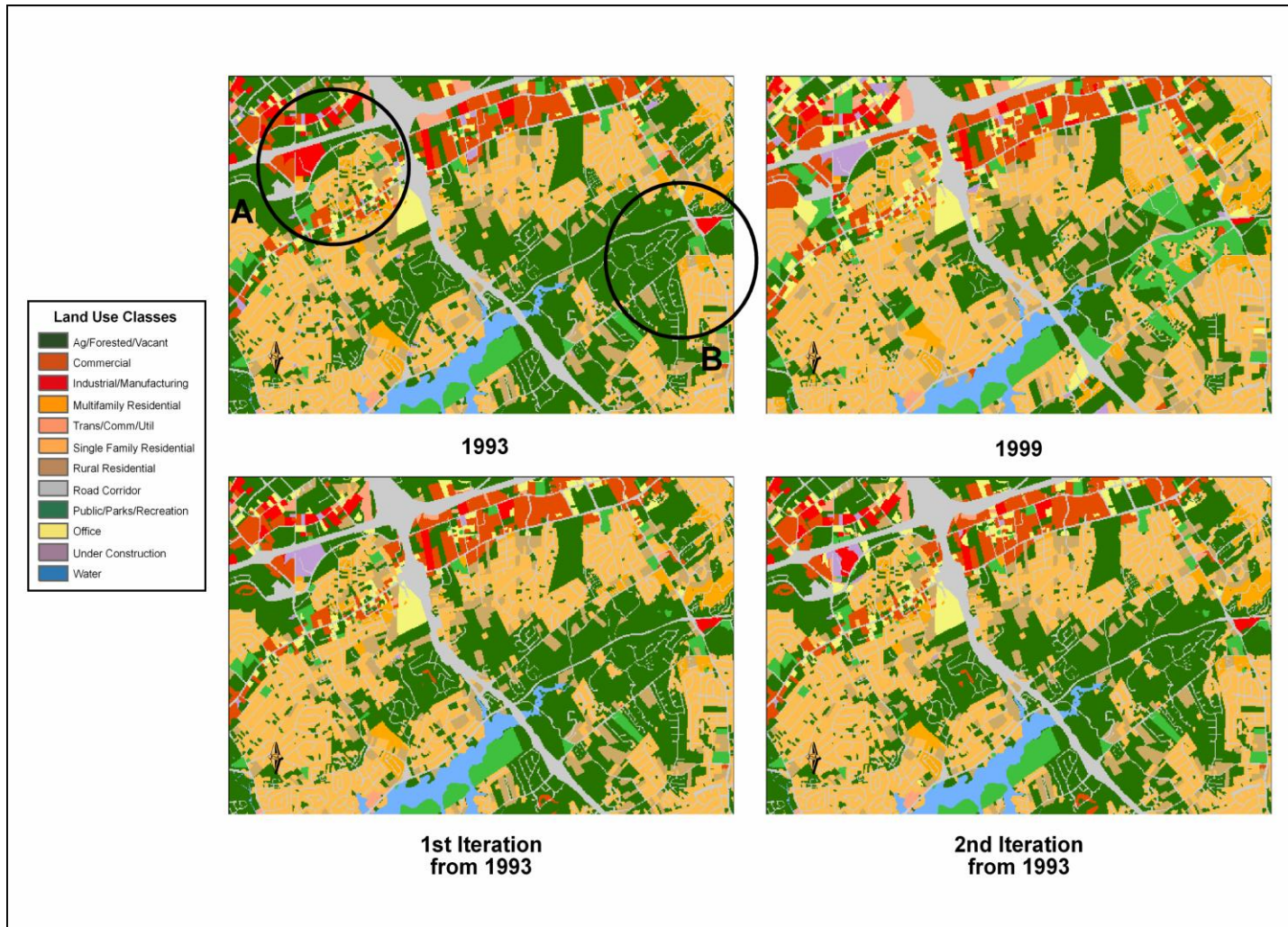


Figure 9 – Iteration Problem

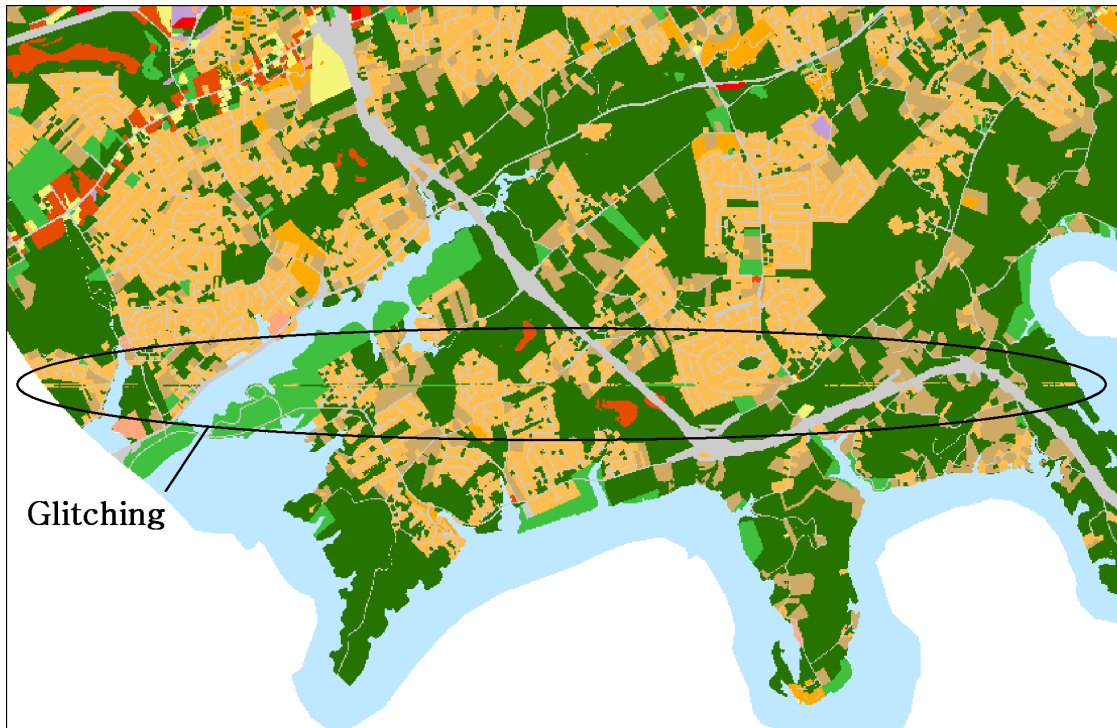
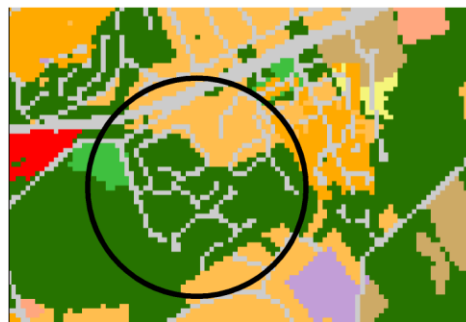
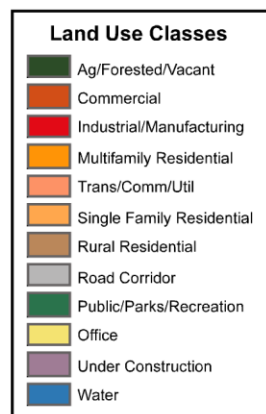


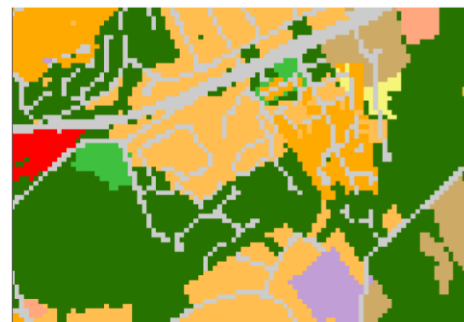
Figure 10 – Horizontal output glitches

orange cells) circled in the first iteration. As the model iterates, the group of cells spreads to surround the road corridors, joining the bright green (parks/public lands/recreation) cells to the west by the third iteration. The fourth frame shows the first iteration with no threshold applied, which shows an extension of the same cluster of single-family residential land use. Note that this form extends farther southward and does not reach as far west as the park. This is a demonstration of CA's ability to provide emergent growth patterns rather than being limited to an interpolation or extrapolation of the historical data. The fourth frame depicts an essentially interpolated intermediate point between the earlier and later historical data snapshots, where the third one shows a true intermediate stage of development.

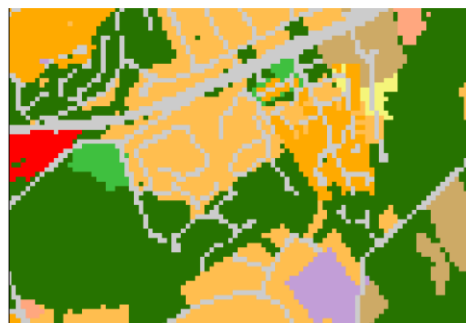
Maps of the first iteration of all three model configurations, the first three iterations of the model configuration with the 2005 roads superimposed, and the 1993, 1999, and 2005 historical data are shown in the Appendix for closer examination. As it was used as the initial condition for the three-iteration sequence shown, the map of the 1993 land use data with the 2005 roads superimposed is also included prior to that sequence.



1st Iteration



2nd Iteration



3rd Iteration



**1st Iteration,
no threshold**

Figure 11 – Iterative Growth

5. ANN-Prediction Accuracy

Due to the iterative difficulties discussed in the preceding section, the first-iteration performance with no development threshold applied will be examined to assess the models' comparative potential to simulate development of the six selected land use classes identified in Section 1 of this chapter. Table 4 shows three performance measures for each of the six classes. The first indicates the cells in the model's output that match the 2005 historical data for that land use class. The second shows the percentage of cells in the predicted output exhibiting a false positive (Type 1 error), with the third indicating false negatives (Type 2 errors).

The model configuration using the 1993 local roads as the initial condition showed the highest accuracy, with an average 51.5% of its output across the six

Table 4 – Model Accuracy Summary

LAND USE CLASS	Percentage of Model Output Matching 2005	Percentage of cells exhibiting a false positive	Percentage of cells exhibiting a false negative
With 1993 Local Roads			
Single-Family Res.	71.3%	13.5%	15.2%
Multi-Family Res.	45.1%	10.2%	44.7%
Ag/Forest/Vacant	60.2%	28.7%	11.1%
Commercial	54.7%	29.8%	15.5%
Office	42.7%	14.2%	43.1%
Public/Parks/Rec.	57.3%	14.2%	28.5%
AVERAGE	51.5%	21.5%	27.1%
With 2005 Local Roads			
Single-Family Res.	61.2%	17.3%	10.9%
Multi-Family Res.	42.2%	8.7%	45.3%
Ag/Forest/Vacant	61.1%	24.3%	14.6%
Commercial	42.7%	23.5%	23.7%
Office	35.6%	15.0%	44.0%
Public/Parks/Rec.	55.6%	6.3%	34.6%
AVERAGE	48.4%	19.7%	29.3%
No Local Roads			
Single-Family Res.	70.8%	11.7%	17.5%
Multi-Family Res.	46.0%	7.8%	46.2%
Ag/Forest/Vacant	59.1%	32.3%	8.6%
Commercial	59.7%	15.4%	25.0%
Office	41.1%	16.7%	42.2%
Public/Parks/Rec.	57.3%	13.3%	29.4%
AVERAGE	50.1%	24.5%	25.4%

selected land use classes matching the historical data. Using the 2005 road network as the initial condition lessened model accuracy some 2.1% through an increase in Type 2 error. This is not surprising, as this configuration was chosen for its possible improvement in iterative performance, rather than the expectation of short-term accuracy. In this regard this modification is a failure, in that it only barely eases the difficulties encountered during iteration while generally degrading accuracy potential. The introduction of the 2005 road network as the initial condition is not consistent with the historical data, and this discrepancy lessens the relevance of the parameter combinations the ANN has perceived in the historical data during training. The third configuration, no local roads, also proves less successful than the first configuration, this time through an increase in Type 1 error. In this configuration the model sees parameter combinations that are consistent with the historical data, but the exclusion of the local roads parameter diminishes the reliability of using them to infer land-use change. Stated another way, the inclusion of local roads improves the model's descriptive power.

The relative performance of the three model configurations are now discussed for the six land use classes selected in Section 1, with comparison maps of various notable model outputs presented. The 2005 historical patterns are superimposed as white outlines on top of colored pixels that comprise the model's output for the land use class discussed.

6. Single-Family Residential

First iteration performance on the single-family residential class was uniformly the best of the six classes examined, with three tested configurations averaging 67.8% predicted cells matching the 2005 historical pattern. Variation exists between the tested configurations, however, with the configuration using the 1993 local roads producing the best statistical fit and apparent pattern (see figure 12). It produces a slight increase in statistical accuracy as compared to the configuration not considering local roads. Inspection of the available historical aerial photography makes clear that residential construction often starts before the road network is extended to that area, and this is borne out by the competitiveness of the 'no local roads' configuration's performance. It would appear that the assumption made is that the attendant roads will be constructed later, and if the attendant roads are not constructed the development may not continue. The accuracy problems encountered iterating the models without road growth tends to support this assertion, in that it was characterized by a lack of development in the later iterations. Superimposing the 2005 roads as the initial condition produces an approximate 10% drop in accuracy over the other two configurations, with the output marked by false positives (see figure 13).

Figure 15 shows a cropped section of figure 13, the single-family residential predicted output using the 2005 local roads. This area lies just north of the center of the study area, and shows a railroad corridor. The historical data reveals lessened single-family residential development in this area, whereas the model's output favors it for development. This is a shortcoming in the model to be addressed, as a railroad distance parameter and railroad land use class could be added and would likely improve the

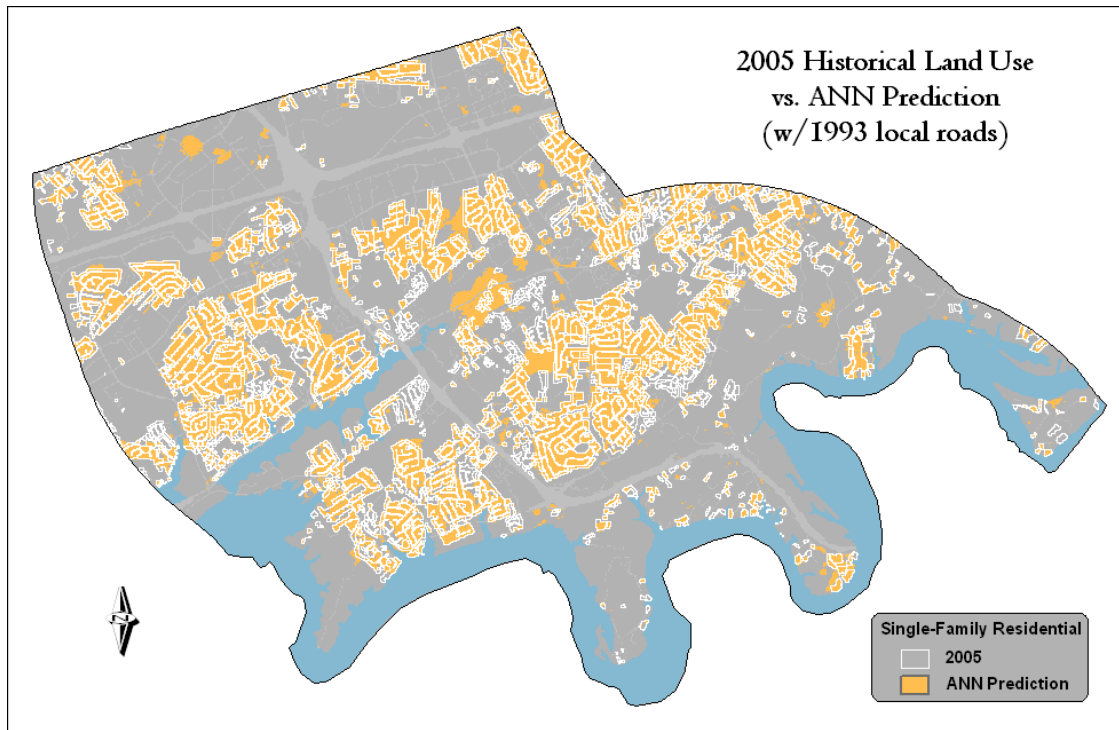


Figure 12 – SFR prediction with 1993 local roads

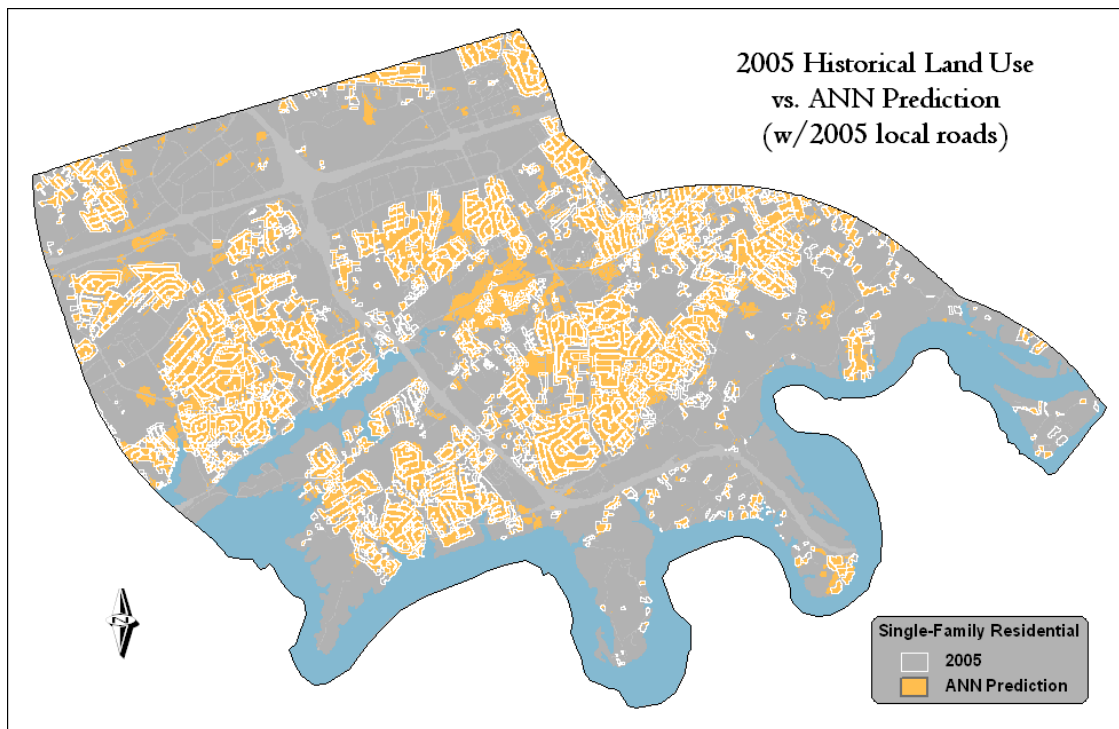


Figure 13 – SFR prediction with 2005 local roads

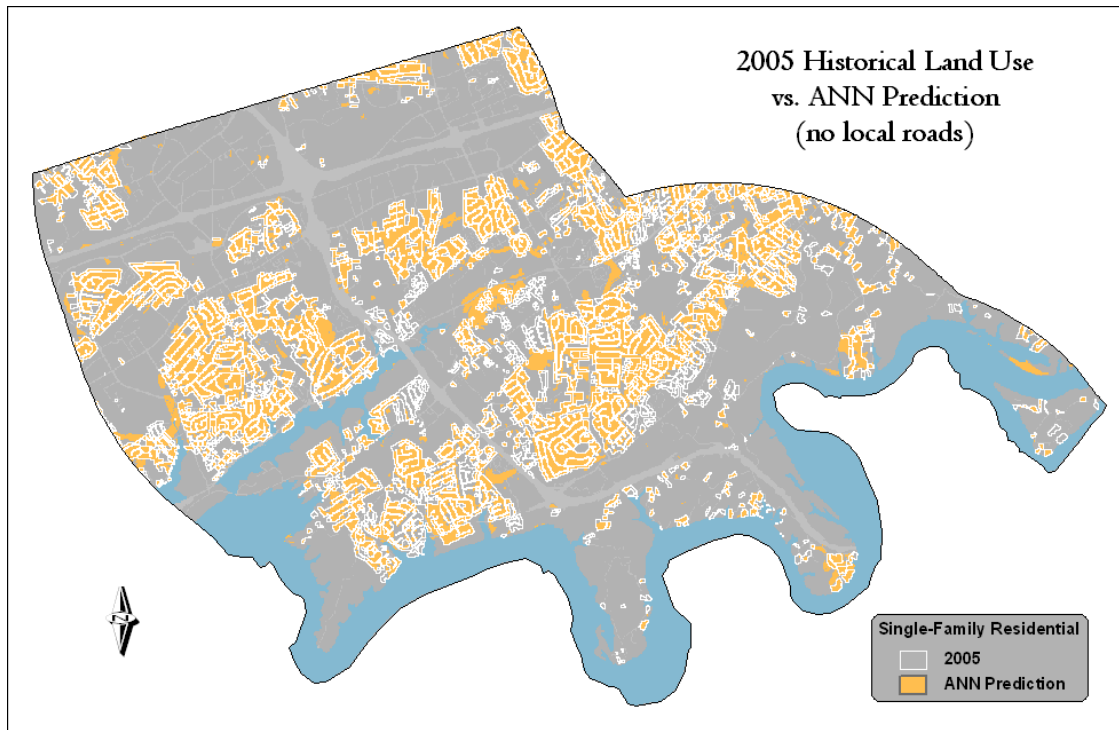


Figure 14 – SFR prediction with no local roads

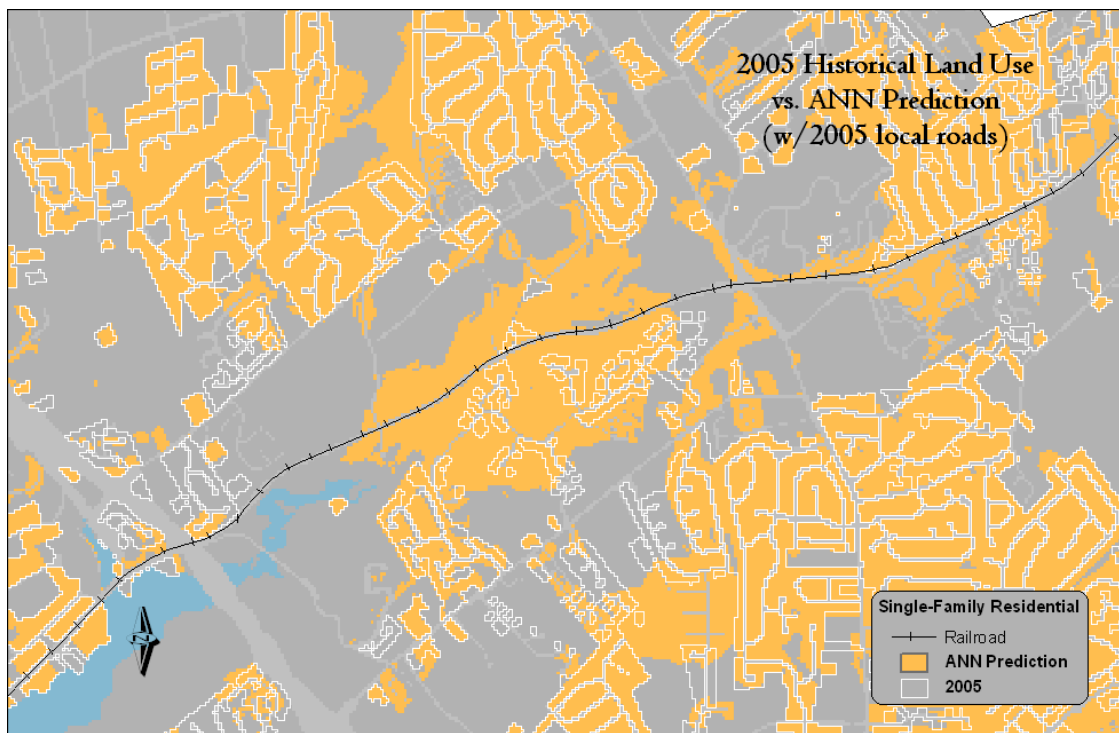


Figure 15 – Railroad Effect

model's performance near railroad features. This limitation is common to all three model configurations with the effect most visible in the output pattern shown.

7. Multi-Family Residential

Multi-family residential (MFR) performance is poor in all three configurations, with an average matching percentage of 44.5%. The bulk of the new MFR parcels are not reflected in the ANN predicted output, with Type 2 error rates similar to the matching rate. This suggests that MFR as a land use class is dominated by external factors not addressed by this model's design. The pattern is not spread-centric enough nor related uniquely enough to its road distance or neighborhood parameters for this model to effectively model this class. A sub-model may be required to adequately address this concern, as significant inability to predict the typically disconnected new MFR sites is prevalent in the output of all three configurations. Guan, Wang, and Clarke (2005) augment the CA of their "ANN-Urban-CA Model" with a Tietenberg macro scale socio-economic model to quantify the demand for urban space. This resource economics model can solve the problem of sustainable resource consumption, in this case land (Tietenberg, 1992). This model treats land a finite resource consumed by urbanization. Inclusion of this information could inform the model of the severe underproduction in this land use class relative to the indicated demand for urban space. Figure 16 shows the output of the first model configuration (using 1993 roads) and is representative of the output of the other two model configurations.

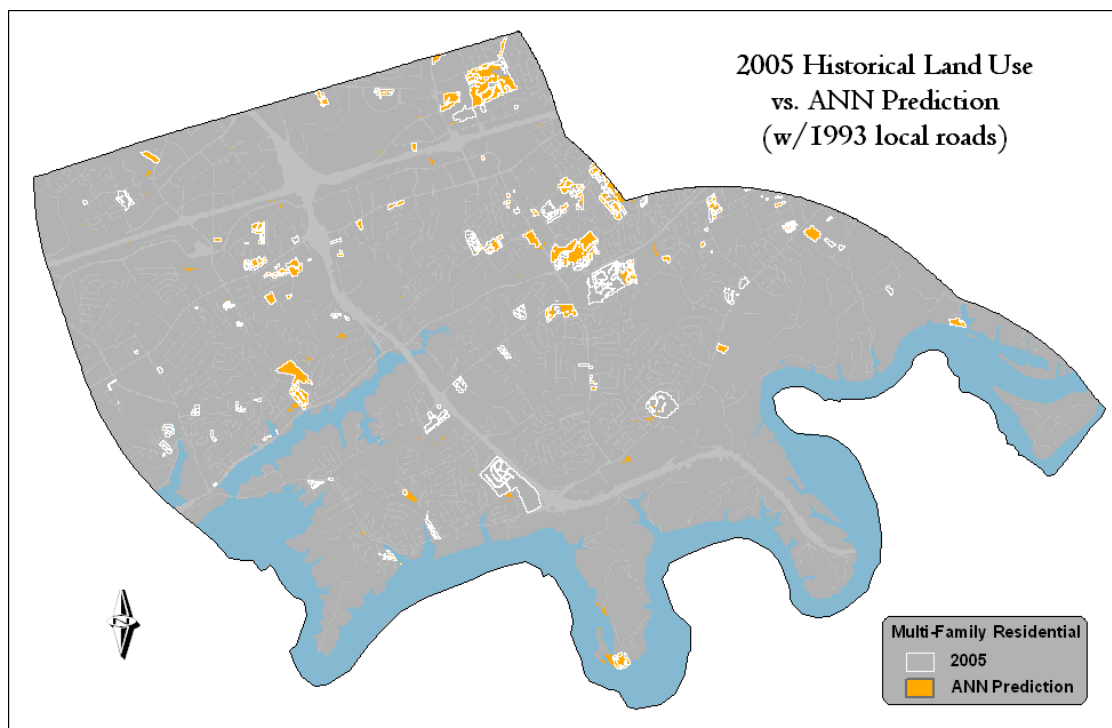


Figure 16 – Multi-family residential prediction with 1993 local roads

8. Agriculture/Forested/Vacant Land

The primary stock being consumed by land-use conversion in the study area is the agriculture/forested/vacant land class (see Table 3). The remains predicted by the model emphasize the model's ability to discern development as whole. All three model configurations hovered around a 60% match rate, with notable differences in the balance of the errors. The configuration considering no local roads produced a Type 1 error rate of 32.3%; it is more often than the other configurations predicting development in the wrong place. This reflects a decrease in the statistical significance of the parameters this configuration considers. Quite simply, distance to local roads improves the accuracy of land use-change predictions, all other factors being equal. The configuration considering the 1993 road network decreases the Type 1 error to 28.7%, with the configuration using the 2005 road network as the initial condition further reducing Type 1 error to 24.3%.

Imposing the 2005 road network as the initial condition is a modification intended to inform the model of the extent of future road development as the model tended to predict less development than noted in the historical data. It had this effect, allowing more accurate identification of locations that have poor potential for future development. This does not mean the model knows what to convert more likely developable location to, as evidenced by the poorer short term accuracy indicated by the first iteration performance. The differences are more difficult to discern visually, and thus only the least-accurate prediction is shown (see figure 17).

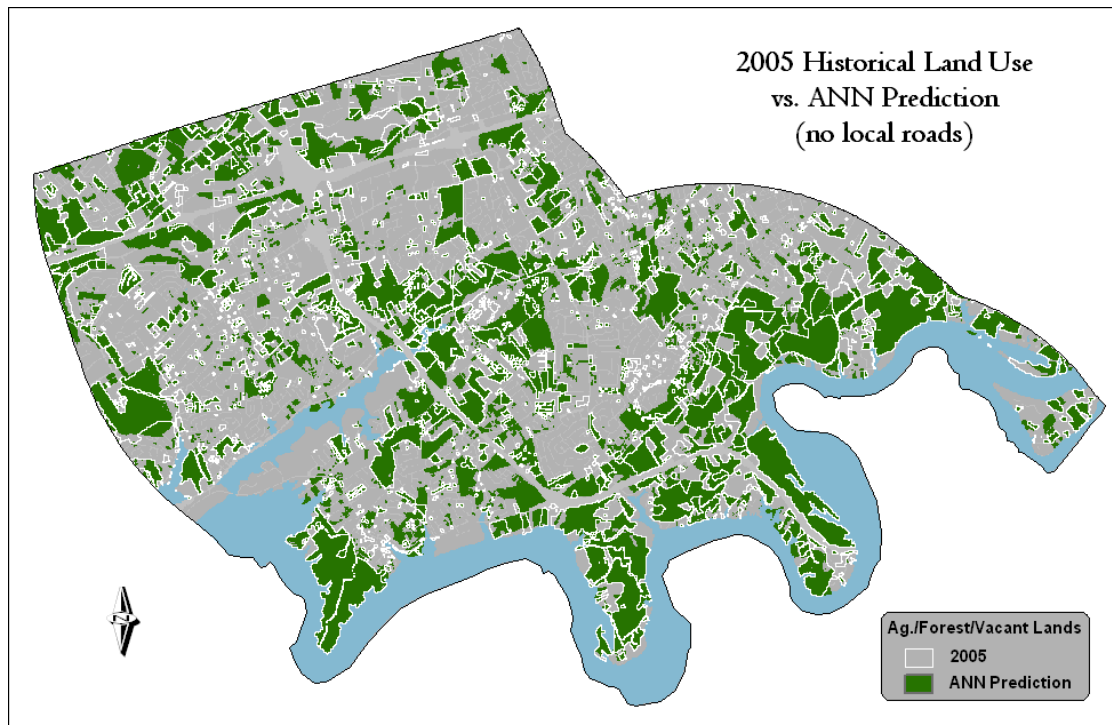


Figure 17 – Ag/Forested/Vacant prediction with no local roads

9. Commercial

Statistical performance in the commercial land class varies widely, but the patterns produced are similar (see figures 18-20). The statistical accuracy and pattern produced in the 2005 local road configuration (figure 19) was poorer, as the premature road development produced by the imposition of the 2005 road network made the sites that would eventually transition to commercial land uses prime targets for prior conversion to other classes such as single family residential. All three configurations predicted commercial development east of I-140 halfway between I-40 and the Knox County line that do not appear in the historical data, indicating that although these sites fit the historical patterns regarding road distance, neighborhood and slope factors, these factors poorly indicate likelihood for conversion to commercial use in this study area. It would appear other considerations figure as prominently in the determination of commercial land use. Incorporation of zoning data or other commercially restrictive legislative factors would allow the model to reject the false positives more effectively and thus more closely simulate the historical development pattern.

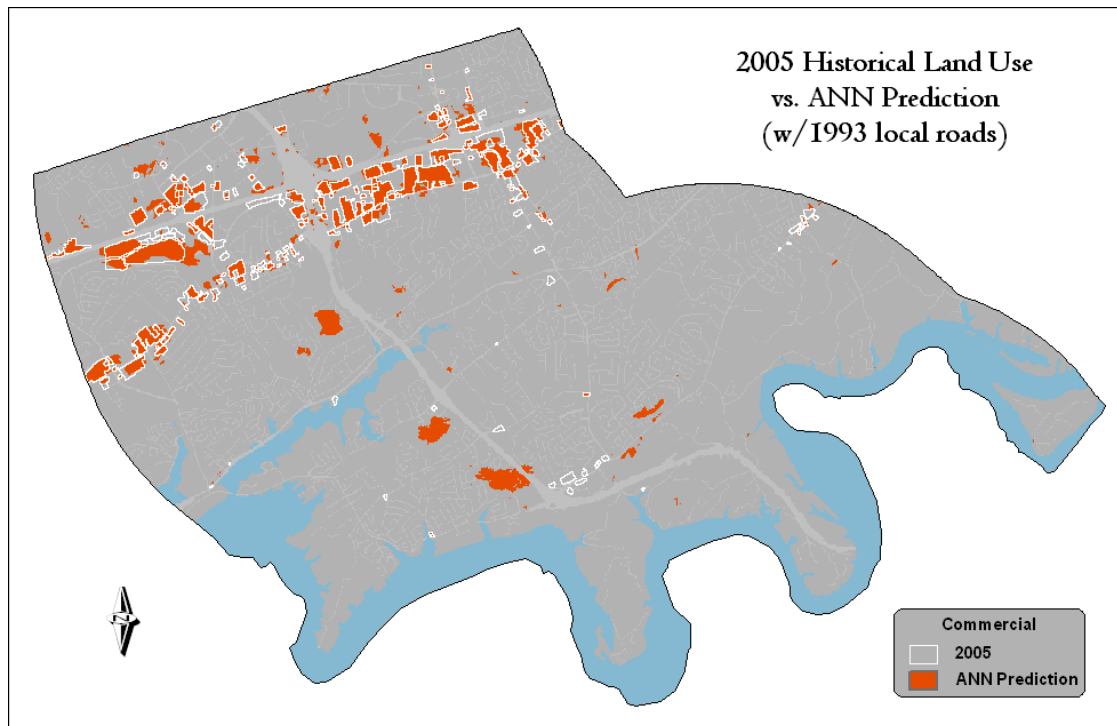


Figure 18 – Commercial prediction with 1993 local roads

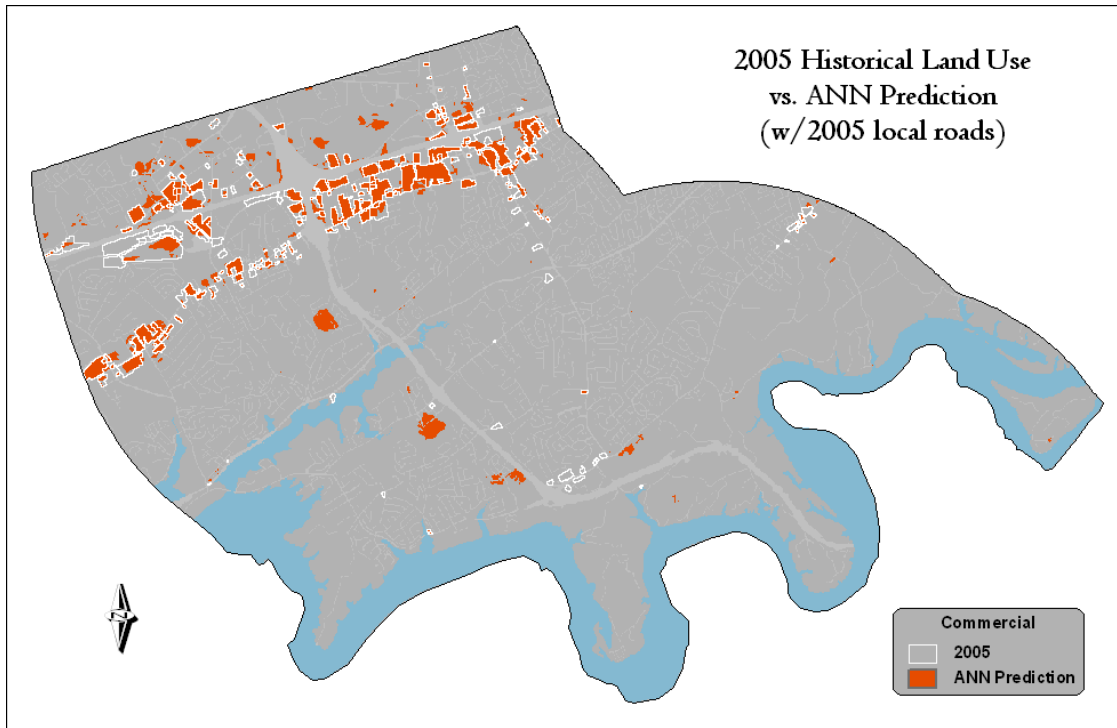


Figure 19 – Commercial prediction with 2005 local roads

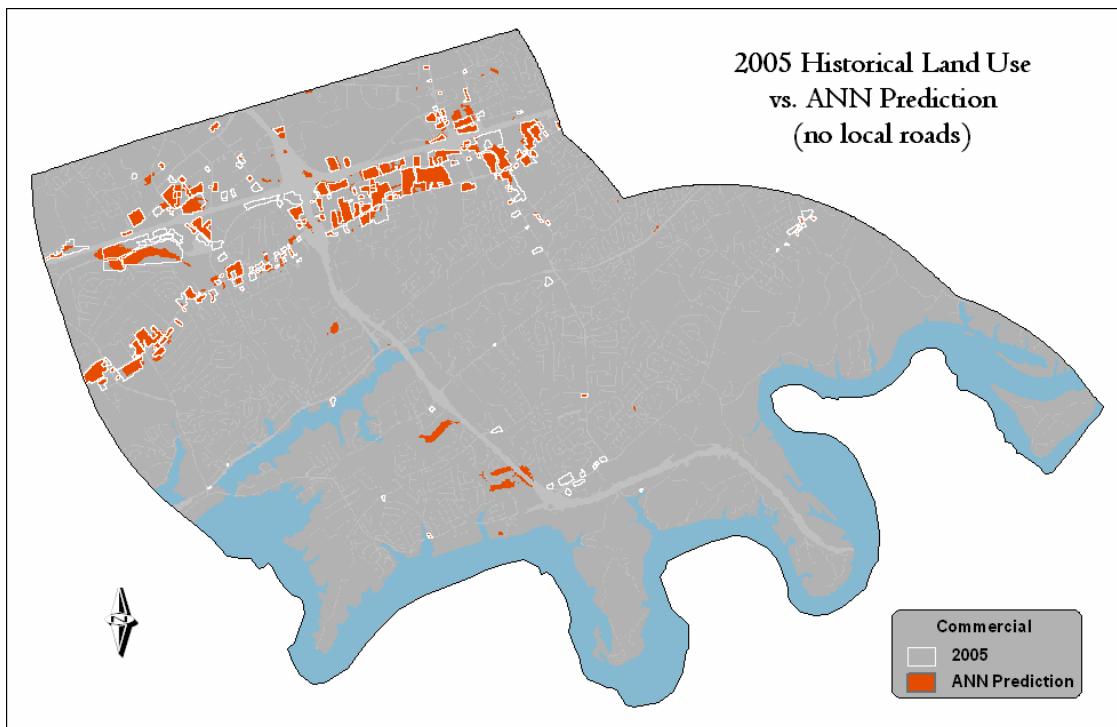


Figure 20 – Commercial prediction with no local roads

10. Office

The performance on the office land use class was the poorest of the six selected classes, with an average matching rate of 39.8% among the three model configurations. Little difference was noted between the three configurations statistically or graphically, thus only the pattern for the first configuration is shown (see figure 21). This lack of variation demonstrates there is little sensitivity to the presence or absence of local roads. The similarity of overall patterns indicate the spatial preference for office and commercial land use to be similar, with neither sensitive to local roads. Performance on either could surely be improved by considering business-centric factors such as communication or zoning concerns, but there is undoubtedly more difficulty in predicting office locations than others. Type 2 error dominates in the office class results, indicating there simply isn't much of a historical trend in the model parameters considered for the ANN to discover.

Note that the performance on office land use prediction in the northern section of the study area is not nearly as accurate as is the commercial class's, suggesting that distances to interstates or major roads don't dictate office locations as directly as they do commercial ones—a notion consistent with the goods-transport needs of most commercial endeavors.

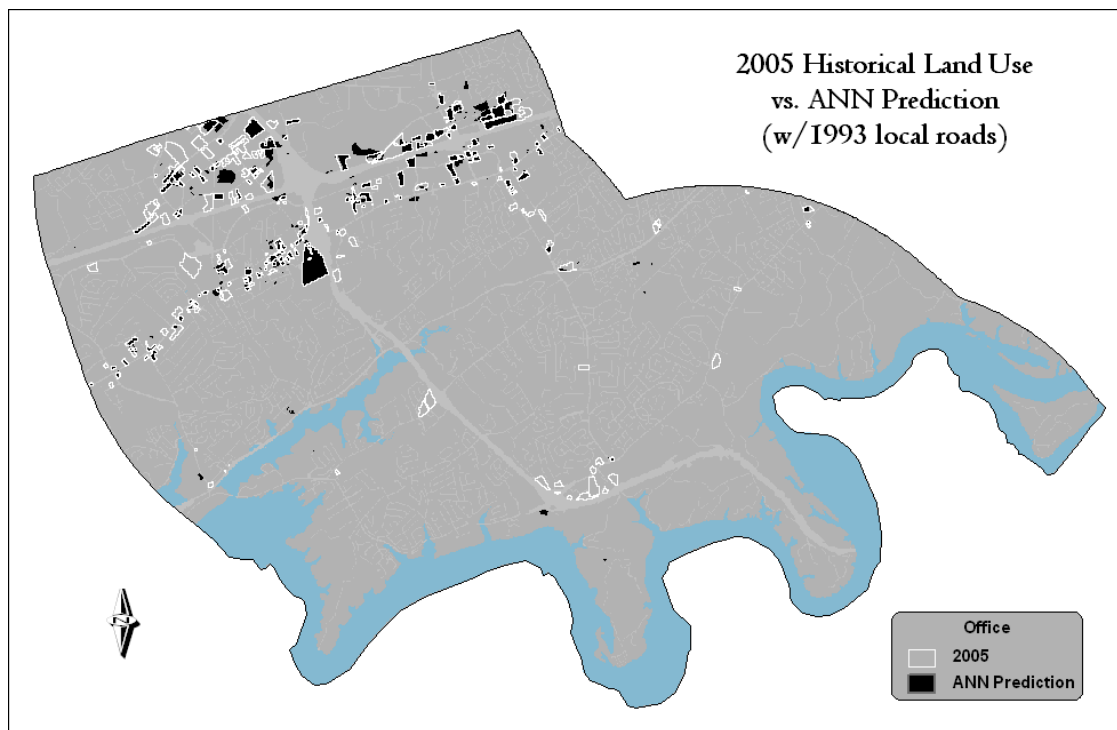


Figure 21 – Office prediction with 1993 local roads

11. Parks/Public Land/Recreation

Although a major statistical difference exists in the performance on this class, there is little difference observable in the patterns produced. For this reason, only the output of the first model configuration will be shown (see figure 22). The statistical difference was that the model configuration that superimposes the 2005 roads as the initial condition exhibited a large shift to Type 2 error, indicating that the historically inaccurate initial condition produced largely eliminates whatever statistical basis for prediction that exists. Major developments in this class located in the center and northwest corner were barely represented, if at all. The inescapable conclusion is that the model parameters used in this project poorly indicate development potential for this class. The central area is near the railroad, and adoption of a railroad distance as a model parameter as suggested in the single-family residential section may improve performance. Both areas are near clusters of residential land uses. Implementation of a cluster analysis on the residential land classes would allow the distances from these clusters to be included as an additional model parameter, likely improving the performance on this land use class.

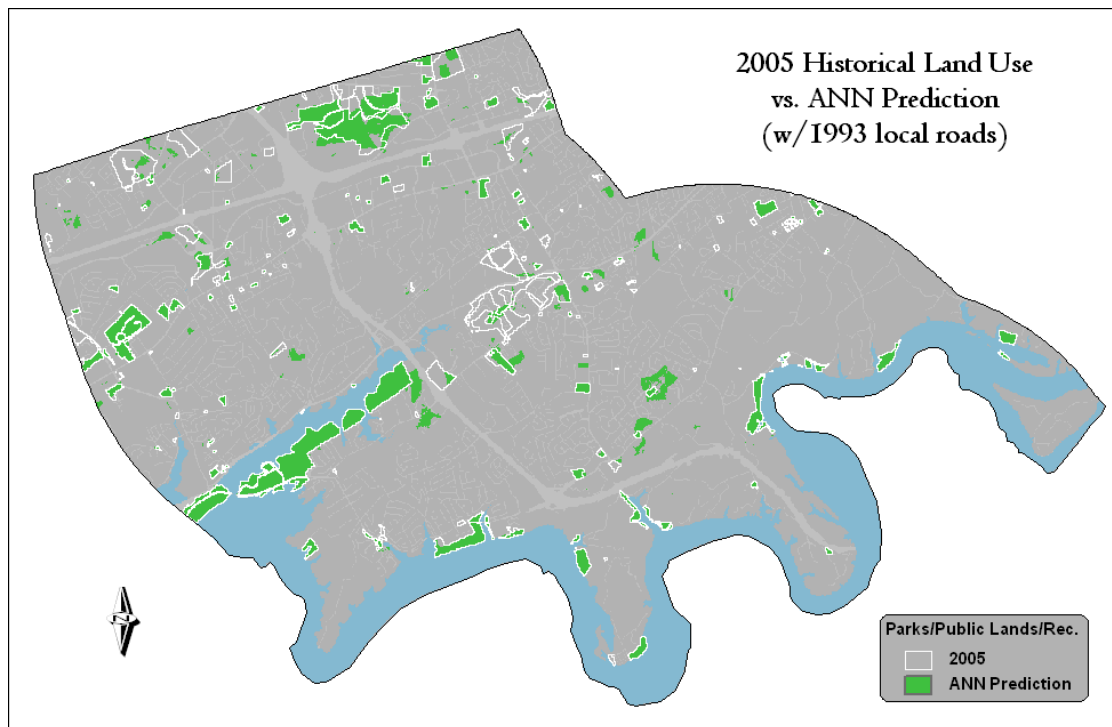


Figure 22 – Parks/Public Land/Recreation prediction with 1993 local roads

CHAPTER VI

CONCLUSIONS AND FURTHER RESEARCH

An experimental CA LUT model was created using an ANN to process the CA transitions for the I-140 study area. The model constructed considers slope, the distances to interstates, local, and major roads, as well as the current land use at a location as well as land uses present in a quarter-mile radius from that location. ANNs of various complexity levels were tried, with a second hidden layer and the *tanh* activation function providing the best simulation results. Simpler ANNs such as the commonly used single layer MLP were unable to discern the relationships present in the historical data. Once trained, the ANN was used to determine the successive land uses for each iteration of the CA, and graphical land use predictions were constructed from these. The predictions obtained without iteration were compared to the historical patterns for the six selected land use classes.

The most obvious shortcoming of the model developed in this project is its inability to extend the road network in a realistic manner. While it is possible to convert cells to the road corridor land use, there is no means to do this as an addition to the existing road network. While it is possible to have the ANN output estimates for new road distances, it is impossible to do this in an incremental fashion or have those estimates correspond to a valid geographic configuration. In this CA implementation the input parameters include distances to road features, but there exists no representation of the geometry or flow inherent in a network. It is possible to create an input parameter for adjacency, but representations of the other considerations inherent in road planning and construction pose severe difficulties within the framework of the CA itself. A sub-model with a vector road network could better model this process, and its output could be seamlessly integrated into the existing CA as its role would be to grow the network from which the CA road-distance parameters are calculated.

Another direct consequence of the over-simplification of the model structure is the poor performance regarding land uses such as 'commercial', 'office', and 'parks/public land/recreation' classes. While these land uses may have strong relationships to road distance factors (as noted in the 'commercial' class), they are also heavily influenced by socio-economic factors not modeled in this study. Possible sources of relief are the adoption of sub-models such as a Tietenberg socio-economic model to gauge demand for urban space. It may also be useful to employ equilibrium concepts in simulating competition for developable locations to better resolve locations that would be consumed by single-family residential usage and those that would fall to multi-family. It is also clear that input parameters need improvement. The addition of a railroad distance factor holds promise, and it may be advantageous to perform a cluster analysis on the existing residential land use to establish community centers to which distance metrics can be calculated for use as additional input parameters.

The simplicity of the model structure used in this study has consequence beyond a reduction in simulation performance, in that the ANN required to produce useful simulation performance is very complex. While the literature discourages complex networks, little explanation exists regarding the consequences of violating this edict. This model's generally poor iterative performance in combination with the apparent

relationship between the stopping point of the ANN training procedure and the occurrence of linear glitches in the model's iterative output indicates a low level of model generality. Although Li and Yeh (2001) recommend simple ANNs for iterative models, it is unclear whether this recommendation corresponds with the effects noted in this study or is motivated by computational concerns. And while it is true that more descriptive input parameters make for better models in the absolute sense, there is an apparent threshold level for input parameter quality that must be reached in order for the relationships present in an ANN-CA LUT model's operation to be quantified by a simple ANN and thus iterate effectively.

Although computational concerns were considered prior to this study, their exact manifestation was surprising. Given a modern PC platform, the processing involved in training the ANN and using it to process each cell's transition proved nearly negligible. The reliance on GIS spatial operations to sample the various grids (slope, land use, and neighborhood data) during parameter generation proved crippling, though. Likely relief from this factor lies in the integration of the ANN into the GIS environment. The NeuroDimensions NeuroSolutions ANN development environment has an optional component that can output an ANN to a dynamic link library (dll) file in a variety of languages, including Microsoft's Visual Basic. The creation of a custom ArcGIS application that calls such a dll file is possible, and with astute program design the spatial operations might be avoided. This would improve simulation run-time by hours, in addition to and far in excess of the time saved via the elimination of the manual data exchange between the GIS and the ANN.

The degree of success with which this model simulates several important land use classes even in its current form is encouraging. The input parameters utilized in this study—slope, current land use, road distance, and neighborhood counts—do indeed affect land use conversion as evidenced by this success. But the relationships are complex, as evidenced by the ANN complexity needed to resolve them, and in the case of several land use classes, far from definitive. The contribution of road distance, as evidenced by the loss of model effectiveness noted from the removal of the local roads, is substantial. LUT models that seek to resolve to the local level need to model the local-level roads, although this raises the data and computational needs of the model.

Short term prediction, in this case that stemming from the first iteration of the model, is usefully accurate for several land use classes, and may extend into the iterative operation once road growth is achieved. Additionally, the model does demonstrate the ability to produce iterative growth, as opposed to mere cross-sectional slices. This encourages the belief that with resolution of the difficulties encountered during iteration the overall results could realize CA's potential to effectively simulate emergent growth phenomenon as encountered in LUT scenarios. With the adoption of an appropriate sub-model for the extension of the road network the dynamic update of all model parameters would be possible between each iteration, and the required ANN complexity would lessen due to a reduction in the complexity of the relationships it is tasked with quantifying. A sub-model to gauge the demand for urban land and/or a mechanism for arbitrating competition for developable land and the introduction of a cluster analysis on residential land use to define urban centers should improve simulation performance in several classes. There is no reason that local scale simulation and limited forecast

functionality cannot be effectively achieved with a development of this model, given an effective vector road growth sub-model combined and the other possible improvements discussed here.

BIBLIOGRAPHY

BIBLIOGRAPHY

- Abdi, 1994. A Neural Network Primer. *Journal of Biological Systems* 2:3, pp. 247-283.
- Alonso, W., 1964. *Location and Land Use*. Harvard University Press, Cambridge, MA.
- Almeida, et al., 2005. GIS and remote sensing as tools for the simulation of urban land-use change. *International Journal of Remote Sensing* 26:4, pp. 759-774.
- Ameida, J., 2002. Predictive non-linear modeling of complex data by artificial neural networks. *Current Opinion in Biotechnology* 13, pp. 72-26.
- Anas, A., 1982. *Residential Location Models and Urban Transportation: Economic Theory, Econometrics, and Policy Analysis with Discrete Choice Models*. Academic Press, New York.
- Anas, A., Duann, L.S., 1986. Dynamic forecasting of travel demand, residence location, and land development: policy simulations with the Chicago area transportation/land use analysis system. In: Hutchinson, B.G., Batty, M. (Eds.), *Advances in Urban Systems Modelling*. North-Holland, Amsterdam, The Netherlands, pp. 299–322.
- Arai, T., Akiyama, T., 2004. Empirical analysis for estimating land use transition potential functions—case in the Tokyo metropolitan region. *Computers, Environment and Urban Systems* 28, pp. 65–84.
- Batty M., and Xie Y., 1994. From cells to cities. *Environment and Planning B* 21, pp. 31–48.
- Batty, M., 1994. A chronicle of scientific planning. The Anglo-American modeling experience. *Journal of the American Planning Association* 60:1, pp. 7-17.
- Batty, M., 1997. Cellular automata and urban form: a primer. *Journal of American Planning Association*, 63, pp. 266–274.
- Batty, M., Xie, Y., 1997. Possible urban automata. *Environment and Planning B*. Volume 24, pp. 175–192.
- Benensen, I., Torrens P., 2004. Editorial: Geosimulation: object-based modeling of urban phenomena. *Computers, Environment and Urban Systems* 28, pp. 1–8.
- Black, W., 1997. North American transportation: perspectives on research needs and sustainable transportation. *Journal of Transport Geography* 5:1, pp. 12-19.

Blecic, I., et al., 2004. Modelling Urban Dynamics with Cellular Automata: A Model of the City of Heraklion. Papers from: "7th AGILE Conference on Geographic Information Science" 29 April-1 May 2004, Heraklion, Greece Parallel Session 4.1- "Geographic Knowledge Discovery", pp. 313-323.

CAGE and the free software at: <http://lamp.sigis.net->

Candau, J., Rasmussen, S. and Clarke, K., 2000. A coupled cellular automaton model for land use/land cover dynamics. 4th International Conference on Integrating GIS and Environmental Modeling (GIS/EM4): Problems, Prospects and Research Needs, Banff, Alberta, Canada, September 2-8.

Citizens Against the Pellissippi Parkway Extension (C.A.P.P.E.)
<http://www.discoveret.org/cappe/index.htm>

Clarke, K., 1997. Land transition modeling with deltatrons. Department of Geography, University of California, Santa Barbara.

Clarke, K., Guan, Q. and Wang, L., 2005. An Artificial-Neural-Network-based, Constrained CA Model for Simulating Urban Growth. Cartography and Geographic Information Science 32:4, pp. 369-380.

Colonna, A., et al., 1998. L.A.U.D.E.: Learning Automata For Urban Development Exploration The Case Study Of Rome Urban System. Paper presented at the 38th ERSA Congress, Vienna, August 28-3, 1998.

Couclelis, H., 1997. From cellular automata to urban models: new principles for urban development and implementation. Environment and Planning B, 24, pp. 165-174.

Harris, C., Ullman, E., 1945. The nature of cities. Annals of the American Academy of Political and Social Sciences 242, pp. 7-17.

Hoyt, H., 1939. Structure and Growth of Residential Neighborhoods in American Cities. US Federal Housing Administration, Washington, DC.

Krugman, P., 1996. The Self-Organizing Economy. Blackwell, Malden, MA.

Lee, D. B., 1973. Requiem for large-scale models. Journal of the American Institute of Planners 39, pp. 163-178.

Lia, X., Yeh, G., 2002. Neural-network-based cellular automata for simulating multiple land use changes using GIS. International Journal of Geographical Information Science 16:4, pp. 323-343.

Lowry, I.S., 1964. A Model of Metropolis. RM-4035-RC. Rand Corporation, Santa Monica, CA.

Miller, E., Kriger, D. and Hunt, J., 1999. Integrating urban models for simulation of transit and land use policies: guidelines for implementation and use. Transit Cooperative Research Program (TCRP) Report 48, Transportation Research Board, National Research Council. National Academy Press, Washington, DC, USA.

Miller, H., Shaw, S.L., 2001. Geographic Information Systems for Transportation: Principles and Applications. Oxford University Press, New York.

Prastacos, P., 1986. An integrated land use-transportation model for the San Francisco region: 1. Design and mathematical structure. *Environment and Planning* 18A, pp. 307–322.

Salvini, P., Miller, E., 2003. ILUTE: An Operational Prototype of a Comprehensive Microsimulation Model of Urban Systems, paper presented at the 10th International Conference on Travel Behaviour Research, Lucerne, August 2003.

Shaw, S.L., Xin, X., 2003 Integrated land use and transportation interaction: a temporal GIS exploratory data analysis approach. *Journal of Transport Geography* 11, pp. 103–115.

Silva, E., Clarke, K., 2002. Calibration of the SLEUTH urban growth model for Lisbon and Porto, Portugal. *Computers, Environment and Urban Systems* 26:6, pp. 525–552.

Silva, E., Clarke, K., 2005. Complexity, Emergence and Cellular Urban Models: Lessons Learned from Applying Sleuth to Two Portuguese Metropolitan Areas. *European Planning Studies* 13:1, pp. 93-116.

Tietenberg, T., 1992., *Environmental and Natural Resource Economics*. Harper and Collins, New York, NY.

Torrens, P., 2000a. How cellular models of urban systems work. CASA working paper 28. London: University College London, Centre for Advanced Spatial Analysis.
<http://www.casa.ucl.ac.uk/how_ca_work.pdf>

Torrens, P., 2000b. How land-use-transportation models work, University College London, Centre for Advanced Spatial Analysis, Working Paper 20. Available online at
<http://www.casa.ucl.ac.uk/how_lutms_work.pdf>

von Neumann, J., 1966. Theory of self-reproducing automata. University of Illinois Press, Urbana, IL.

Wegner, M., Feurst, F., 1999. Land use transport interaction: state of the art. Deliverable 2a of the project TRANSLAND of the 4th RTD Framework Programme of the European Commission. Institut für Raumplanung, Universität Dortmund.

Wegener, M., 2004. Overview of transport and land use models. In: Handbook of Transport Geography and Spatial Systems 5, Elsevier Science, Amsterdam, pp. 127-146.

White, R., Engelen, G., 1993, Cellular automata and fractal urban form: a cellular modelling approach to the evolution of urban land-use patterns. Environment and Planning A, 25, pp. 1175-1189.

White, R., Engelen, G., 1994. Cellular Dynamics and GIS: Modeling spatial complexity. Geographical Systems 1, pp. 237-253.

White, R. Engelen, G. 1997. Cellular automata as the basis of integrated dynamic regional modelling. Environment and Planning B 24, pp. 235-246.

White, R., Engelen, G. and Uijee, I. 1997. The use of constrained cellular automata for high resolution modeling of urban land use dynamics. Environment and Planning B, 24, pp. 323-343.

Wolfram, S., 1994. Cellular Automata and Complexity. Collected Papers Addison-Wesley Publishing Company, Reading MA.

Wu, F., 1999. A simulation approach to urban changes: experiments and observations on fluctuations in cellular automata. Sixth International Conference on Computers in Urban Planning and Urban Management, Venice, Italy.
<<http://brezza.iuav.it/stratema/cupum>>

Wu, F., 2002. Calibration of stochastic cellular automata: the application to rural-urban land conversions. International Journal of Geographical Information Science 16:8, pp. 795-818.

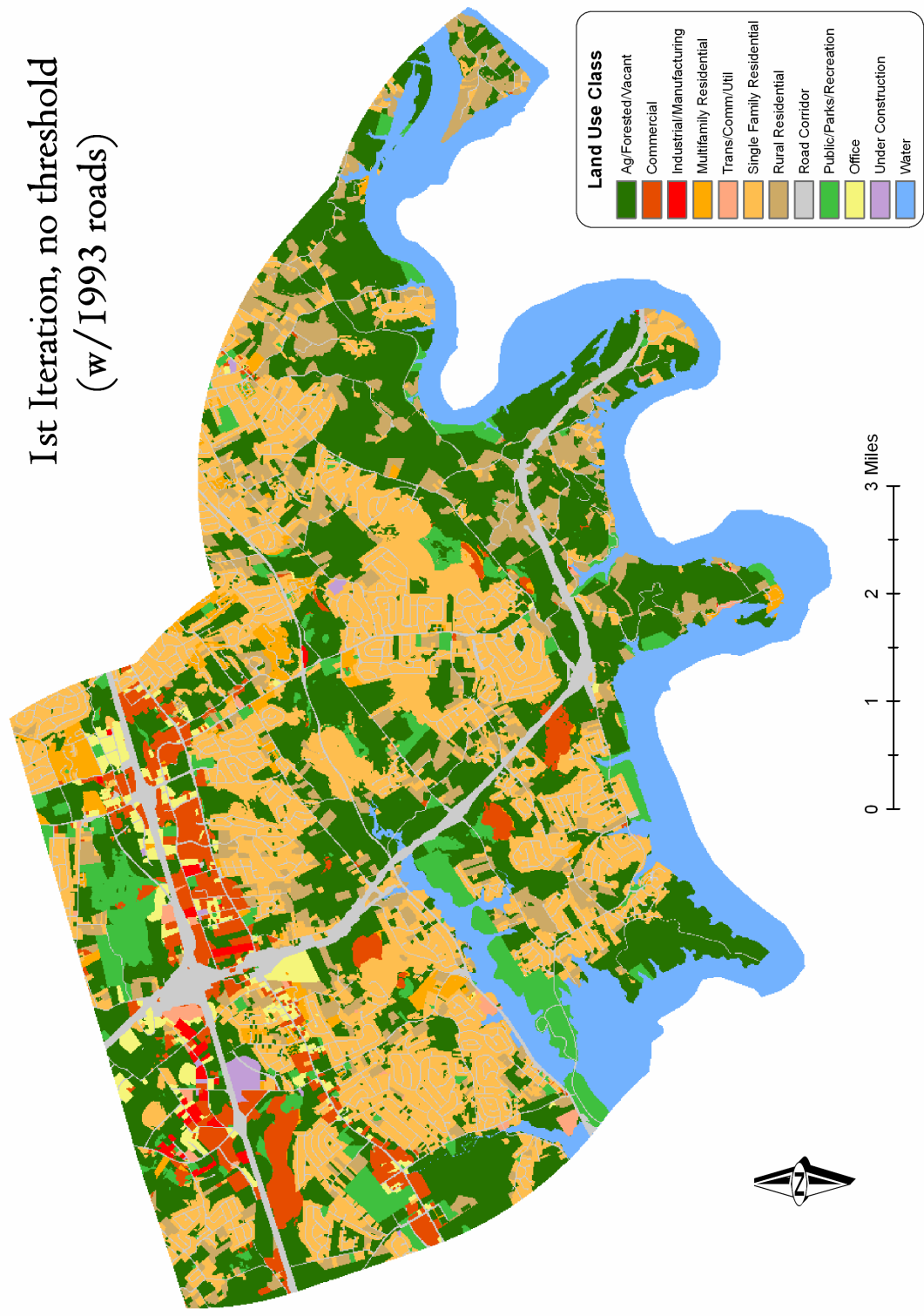
Wu, F., Webster, C., 1998. Simulation of land development through the integration of cellular automata and multi-criteria evaluation. Environment and Planning B, 25, pp. 103-126.

Yeh, A., Lia, X., 2000. Simulation of compact cities based on the integration of cellular automata and GIS. In *Theoretical and Practical Issues on Cellular Automata*. S. Bandini and T. Worsch. London, Springer-Verlag, pp. 170-178.

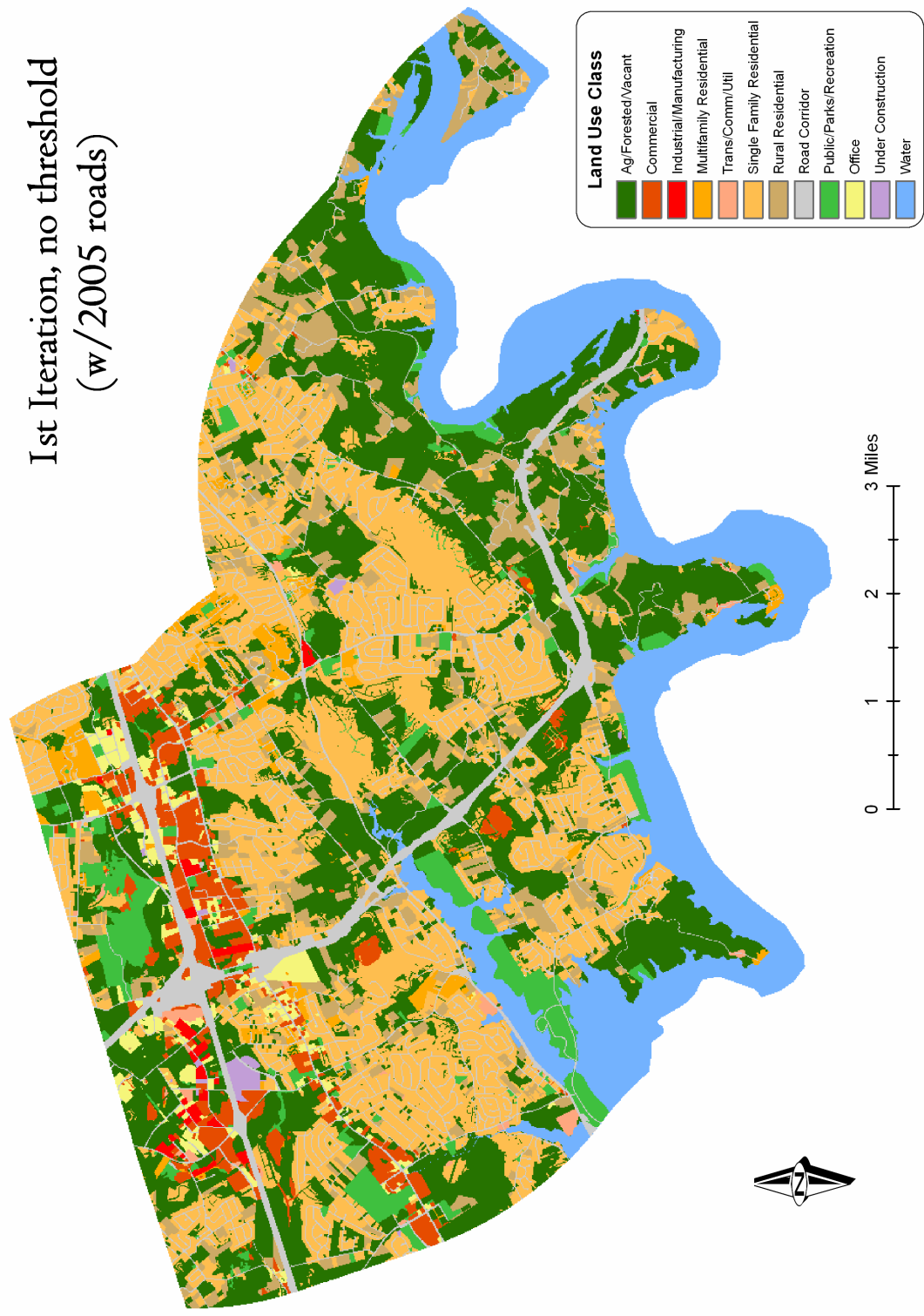
Yeh A., Li X., 2001. A constrained CA model for the simulation and planning of sustainable urban forms using GIS. Environment and Planning B: Planning and Design 28, pp. 733-753.

APPENDIX

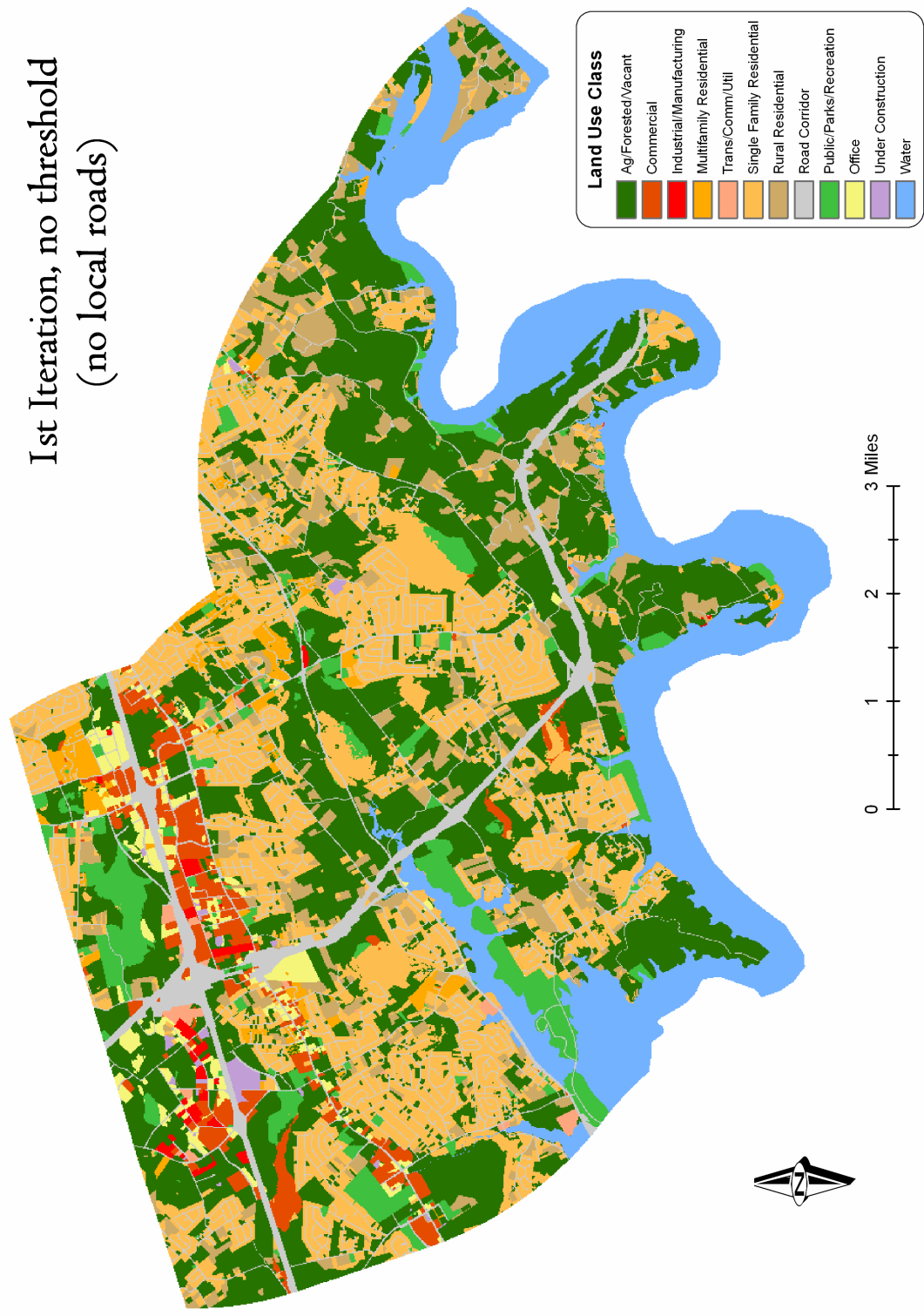
Ist Iteration, no threshold
(w/ 1993 roads)



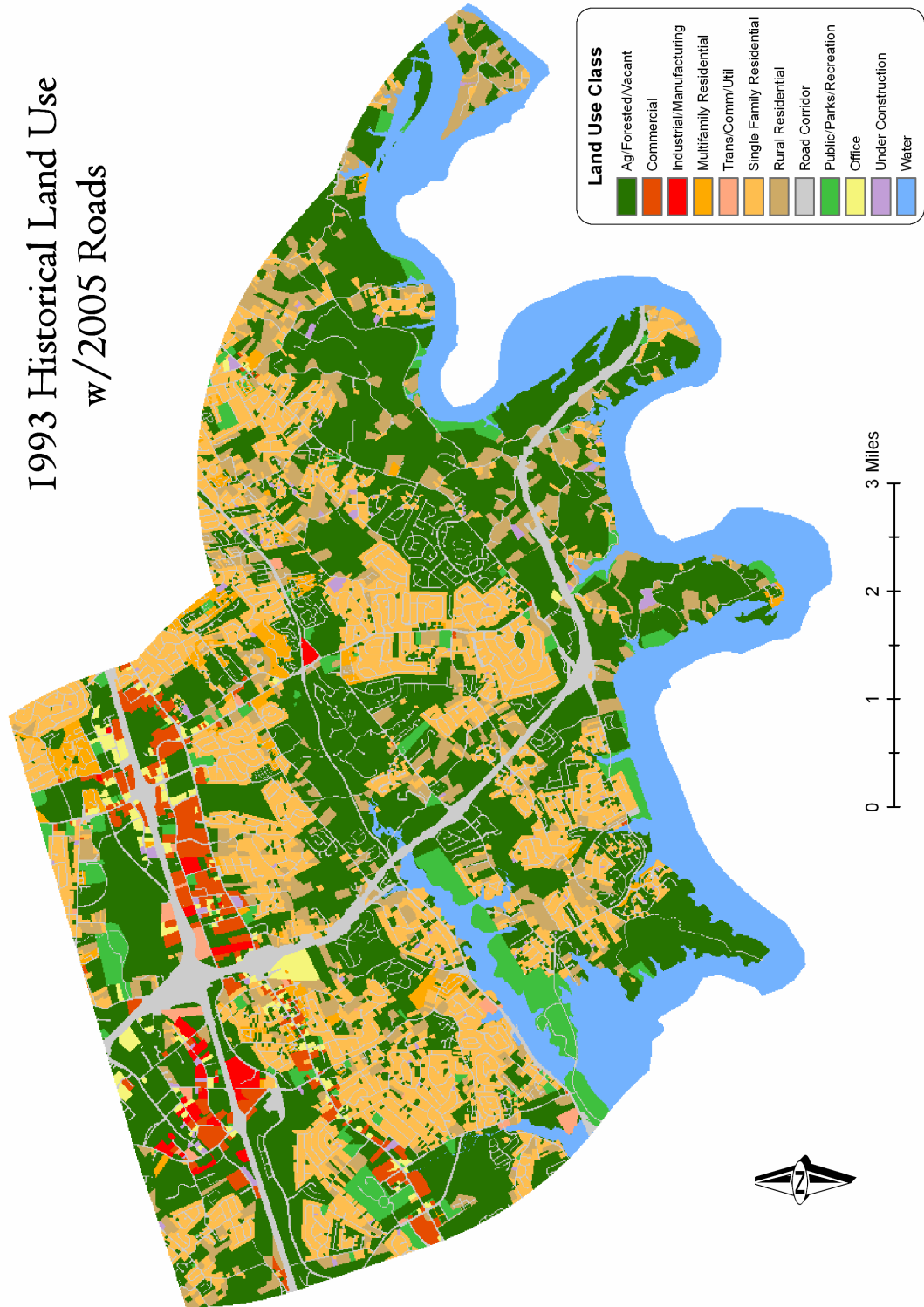
Ist Iteration, no threshold
(w/2005 roads)



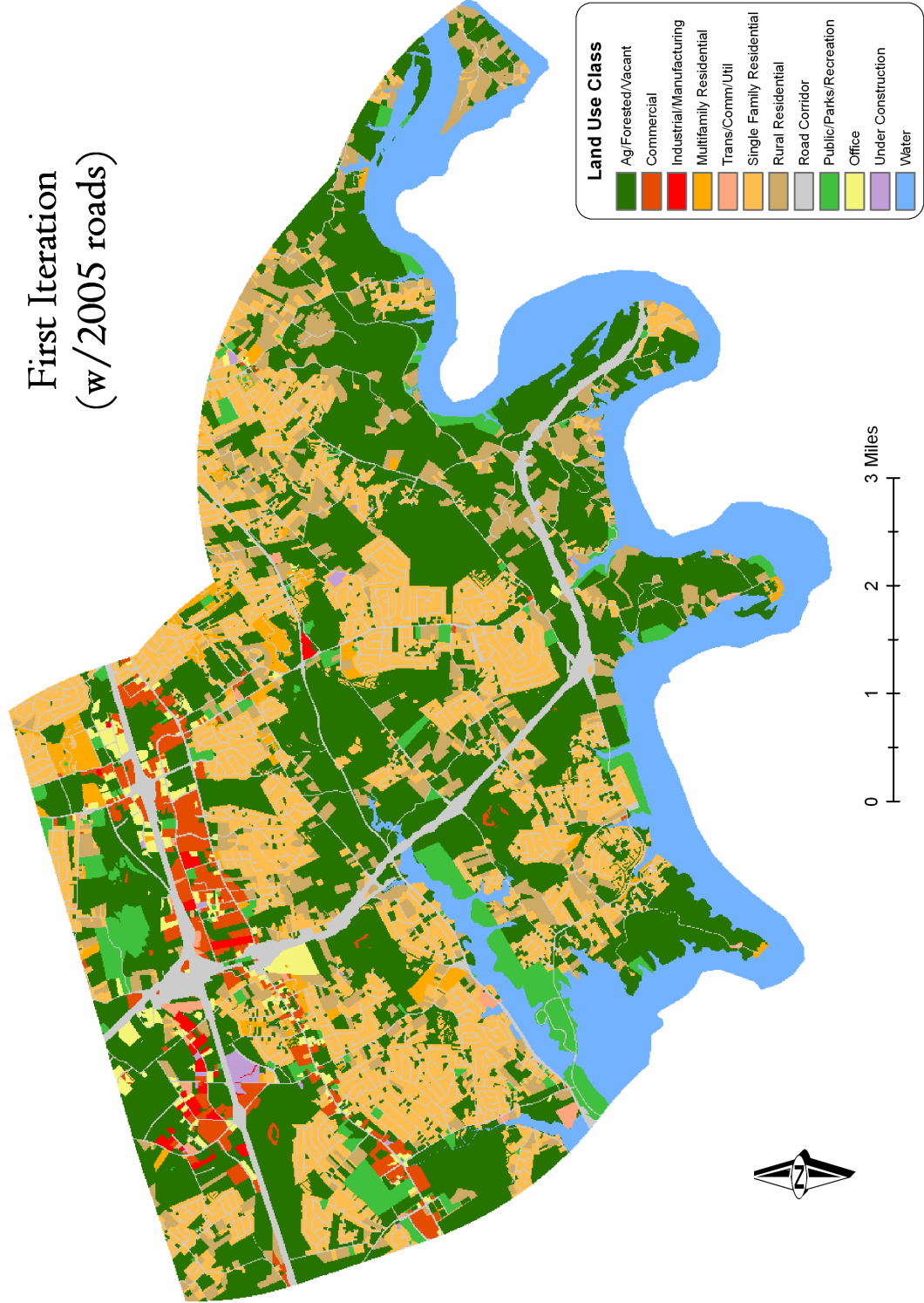
Ist Iteration, no threshold
(no local roads)



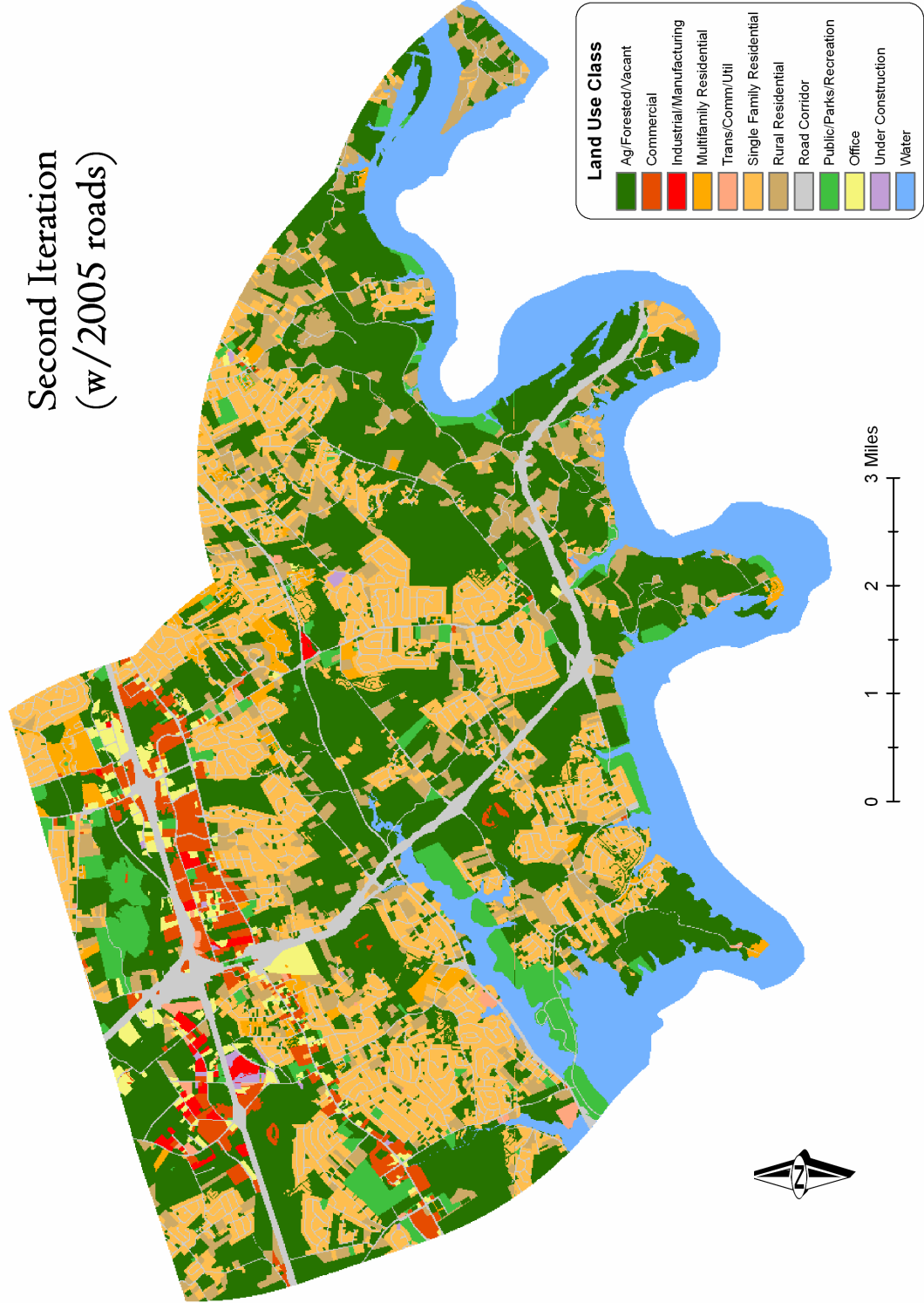
1993 Historical Land Use w/2005 Roads



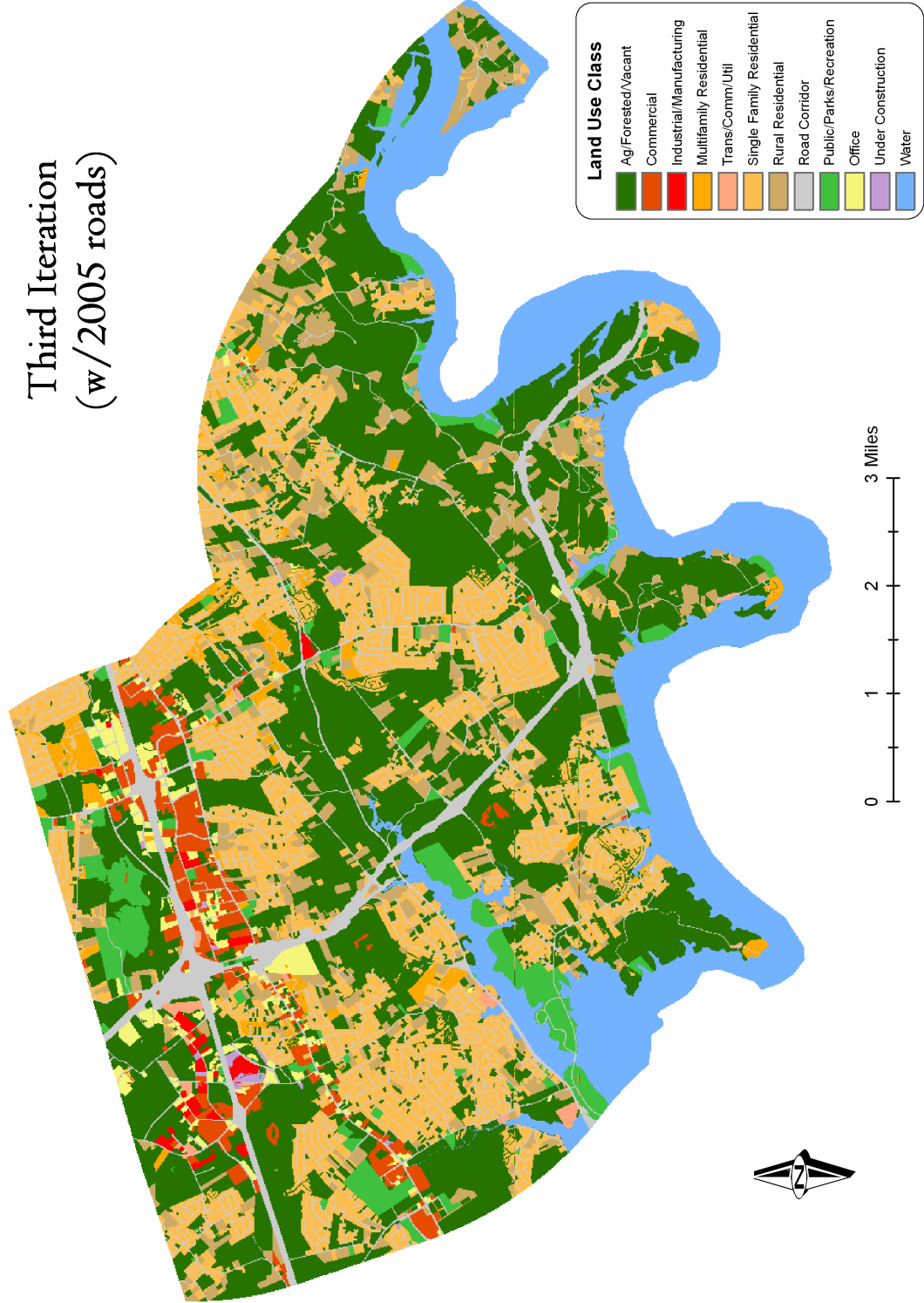
First Iteration (w/2005 roads)



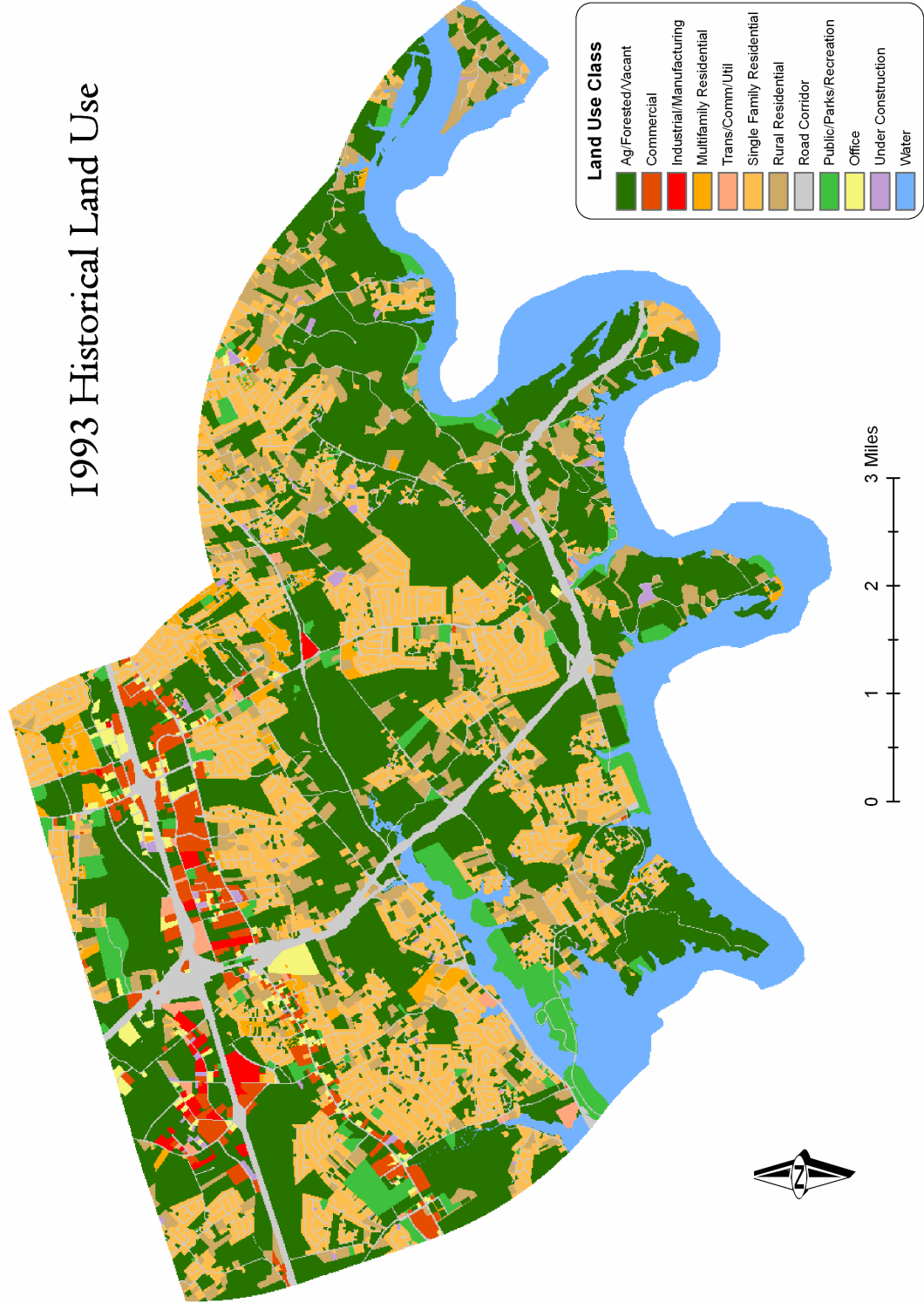
Second Iteration (w/2005 roads)



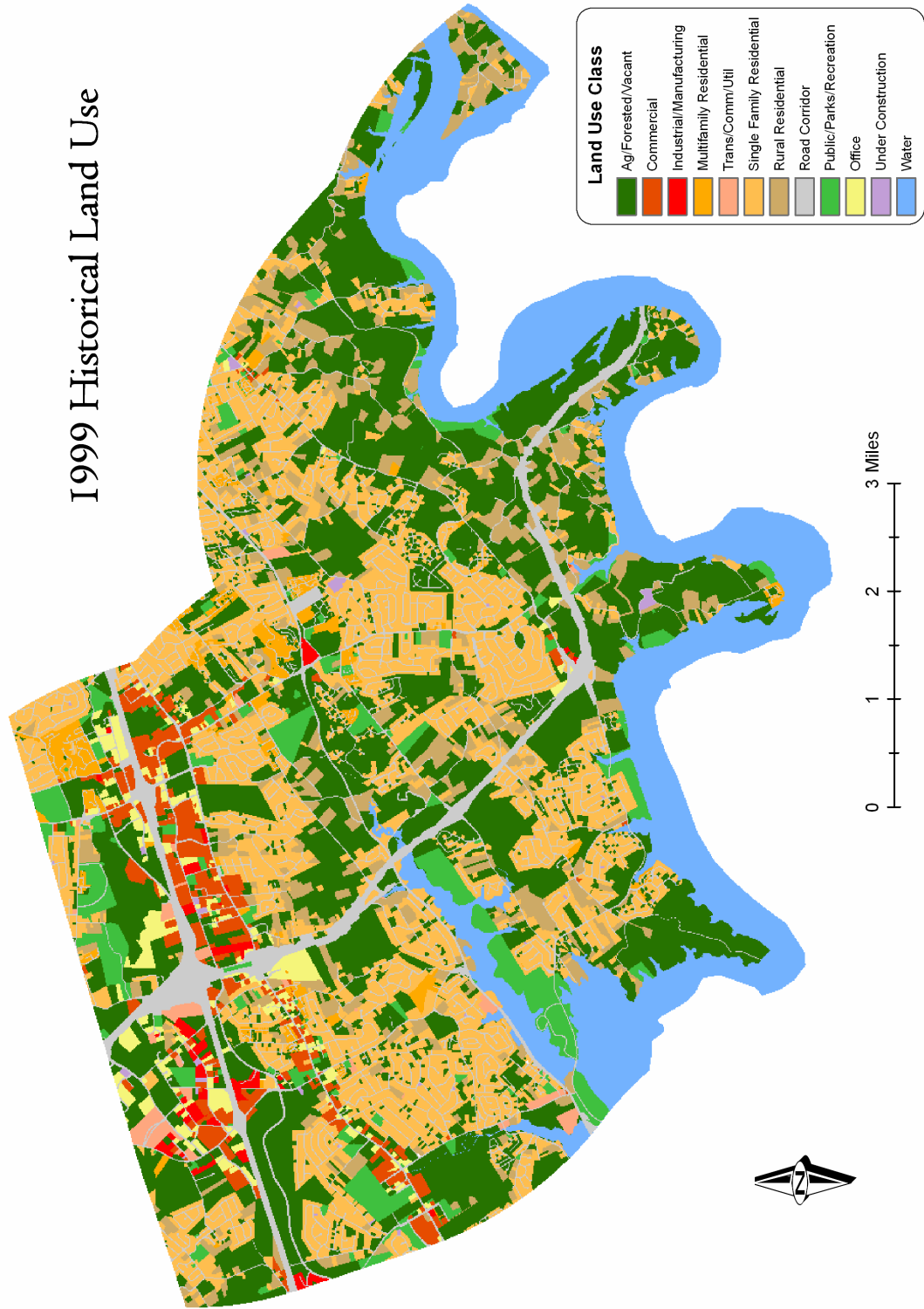
Third Iteration (w/2005 roads)



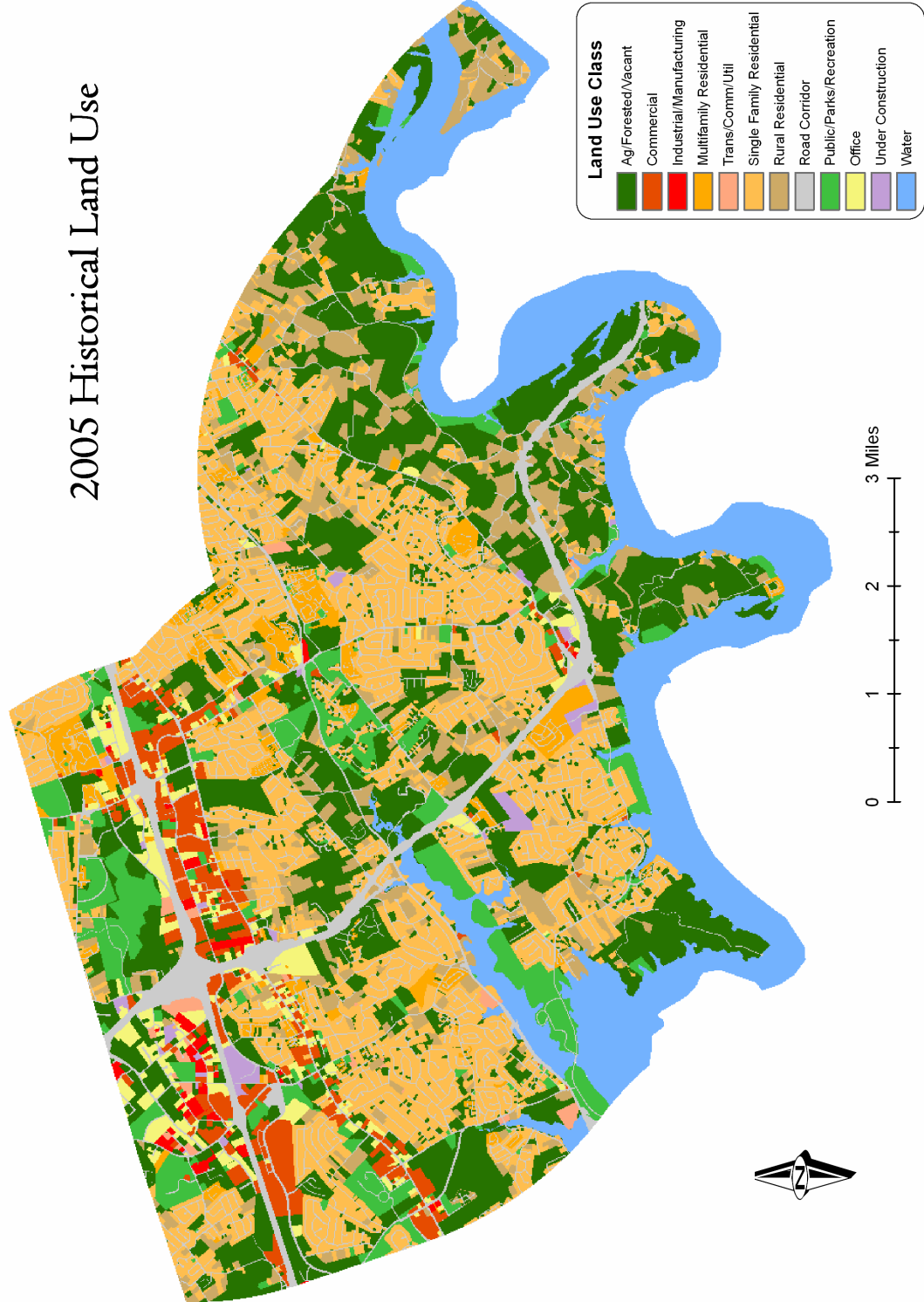
1993 Historical Land Use



1999 Historical Land Use



2005 Historical Land Use



VITA

Steve Ahrens completed high school in Tempe, AZ in 1988. He obtained an associates degree in engineering from Pellissippi Technical Community College in 1993. After spending a decade in the personal computer system design and construction field he attended the University of Tennessee, Knoxville, earning a Bachelor of Arts degree in geography in 2005 and a Master of Science degree in geography specializing in GIS and transportation in 2008. He is currently employed as a GIS analyst.