

URBAN GROWTH SIMULATION THROUGH AGENT-INTEGRATED
IRREGULAR AUTOMATA (AIIA)

by

Khila R. Dahal, M.S., M.Ed.

A dissertation submitted to the Graduate Council of
Texas State University in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
with a Major in Environmental Geography
May 2014

Committee Members:

T. Edwin Chow, Chair

Ronald R. Hagelman

F. Benjamin Zhan

Robert Price

COPYRIGHT

by

Khila R. Dahal

2014

FAIR USE AND AUTHOR'S PERMISSION STATEMENT

Fair Use

This work is protected by the Copyright Laws of the United States (Public Law 94-553, section 107). Consistent with fair use as defined in the Copyright Laws, brief quotations from this material are allowed with proper acknowledgment. Use of this material for financial gain without the author's express written permission is not allowed.

Duplication Permission

As the copyright holder of this work I, Khila R. Dahal, authorize duplication of this work, in whole or in part, for educational or scholarly purposes only.

ACKNOWLEDGEMENTS

I wish to thank a number of people who spent their time and effort to help me complete my studies at the Texas State University. My sincere thanks go to the dissertation committee: My adviser T. Edwin Chow, who very closely walked me through the whole process of dissertation research plus co-authorship in multiple research articles, always guiding, supporting, correcting and inspiring to improve; Ronald Hagelman, who played an invaluable role in bringing me to San Marcos, getting adjusted to the academic environment, encouraging to publish in research journals, and helping me develop as a person through constant inspiration and guidance; Benjamin F. Zhan, who provided important feedback and suggestions that were vital to the improvement of the dissertation as a research document of significance; and Bob Price, who helped me bring up a social and holistic perspective into the dissertation research.

I would like to extend my thanks also to several members of the staff at the Geography Department at Texas State University-San Marcos: Allison Glass-Smith, Angelika Wahl, Charles Robinson and Pat Hell-Jones whose precious cooperation and support on administrative matters was essential during my stay in the program. Similarly, thank you much to my classmates including Stephen Tsikalas and Clayton Whitesides, and my office-mates: Bernie Fang, Naizhou Zhou and Nathaniel Dede-Bamfo for all the informative and humorous conversations and socials we had had together. My special thanks are due to Dr. Walter Rast, who was always supportive to my career and kindly let

me be a part of his World Lakes Project, thereby providing me the opportunity to learn further.

Last but not the least, my indebtedness goes to my wife Sunila and son Sarik for their constant cooperation, inspiration and support. I heartily appreciate that Sunila was with me to stand even during some of the frustrated moments. Finally, I would like to dedicate this dissertation to my parents who instilled in me that patience is the best remedy of all difficult moments in life.

TABLE OF CONTENTS

	Page
ACKNOWLEDGEMENTS	iv
LIST OF TABLES	viii
LIST OF FIGURES	ix
ABSTRACT	xii
 CHAPTER	
1. INTRODUCTION.....	1
1.1 Urban growth modeling	1
1.2 Cellular automata and agent-based modeling of urban growth	3
1.3 Structure of the dissertation	9
1.3.1 The AIIA model	11
1.3.2 The parcel subdivision toolset	13
1.3.3 Neighborhood sensitivity of the irregular CA models	13
1.3.4 Conclusion	14
2. AN AGENT-INTEGRATED IRREGUAR AUTOMATA MODEL OF URBAN LAND USE DYNAMICS.....	15
2.1 Introduction.....	15
2.2 Related work	17
2.3 Study site, data and software	20
2.4 Conceptual framework.....	22
2.4.1 Agents of land development	26
2.4.1.1 City planner.....	26
2.4.1.2 Land developers	28
2.4.1.3 Household agents	32
2.4.1.4 Commercial, industrial and institutional agents.....	34
2.5 Results and validation	36
2.6 Discussion and conclusion.....	48
3. PARCEL-DIVIDER: A GIS TOOLSET FOR AUTOMATED PARTITIONING OF URBAN LANDS	52

3.1 Introduction.....	52
3.2 Related work	54
3.3 Parcel subdivision toolset	57
3.3.1 Generalized parcel divider 1	60
3.3.2 Generalized parcel divider 2	65
3.3.3 Divider with inner roads	67
3.3.4 Cul-de-sac creator	70
3.3.5 L-shaped parcel divider.....	71
3.3.6 T-shaped parcel divider.....	72
3.3.7 Multi-family residential parcel divider	73
3.4 Results and validation	74
3.5 Discussion and conclusion.....	80
 4. CHARACTERIZATION OF NEIGHBORHOOD SENSITIVITY IN THE IRREGULAR CELLULAR AUTOMATA MODEL OF URBAN GROWTH.....	 87
4.1 Introduction.....	87
4.2 Literature review	89
4.3 Neighborhood types	92
4.3.1 Adjacency neighborhood	93
4.3.2 Proximity neighborhood	94
4.3.3 Intercepted buffer neighborhood.....	97
4.3.4 Extent-wide neighborhood.....	98
4.4 Urban growth simulation	101
4.5 Results and model evaluation	105
4.6 Discussion and conclusion.....	112
 5. SUMMARY AND CONCLUSION OF THE DISSERTATION.....	 115
5.1 Significance of the study.....	118
5.2 Limitations and the ways forward.....	120
 REFERENCES	 124

LIST OF TABLES

Table	Page
2-1. Weights or preferences of various agents on the influence factors (driving variables) of urban expansion applicable to the City of San Marcos, Texas.....	25
2-2. Site-specific accuracy assessment of the AIIA model outputs.....	40
2-3. Prediction accuracy and errors for the results of AIIA and baseline model.....	42
2-4. Spatial metrics for the comparison of scenario results	47
3-1. Descriptive statistics of the tool results.	78
4-1. The eight neighborhood types defined in this study.....	93
4-2. Urban growth factors included in the present study and corresponding weights for three different categories of land use development	104
4-3. Evaluation of 26 simulation scenarios characterized by neighborhood types and sizes using different spatial pattern and accuracy metrics	109

LIST OF FIGURES

Figure	Page
1.1. Cellular space and the moves of alive cells across time iterations (adapted from Gardner 1970).....	5
1.2. A flowchart showing the basic steps of an urban growth simulation modeling using vector data	10
2.1. A diagram providing a synopsis of the modeling process	20
2.2. Location of San Marcos in the Austin-San Antonio corridor of Texas, USA	21
2.3. Sequential components of the conceptual framework of AIIA and interactions among agents.....	23
2.4. Sites allocated by the planner agent into different future land use classes.....	28
2.5. Pseudo-code of the subdivision component of the AIIA model	31
2.6. Urban growth in terms of development status, i.e. developed or undeveloped lands.	37
2.7. Urban growth in terms of land use zones of developed sites.....	38
2.8. Illustration of error estimation using the technique of spatial overlay of polygons ...	39
2.9. Visualization of prediction accuracy as an error surface highlighting commission and omission errors of the AIIA.....	41
2.10. Visualization of spatial agreement of developed areas for all land use categories for the output of a) AIIA model, and b) baseline model	44
2.11. Comparison of the model results in terms of match classes.	45
2.12. Future growth scenarios for 2020: a) current growth scenario; and b) compact growth scenario	46
2.13. Map depicting spatial distribution of the four types of households in the study area	

for AIIA output.....	48
3.1. Interface of ESRI's ArcMap 10.x with a dialogue window of one of the script tools contained in Parcel-Divider	58
3.2. Pseudo-code underlying the functioning of Generalized Parcel Divider 1.....	61
3.3. Visual illustration of the process of subdivision mentioned in the pseudo-code.	62
3.4. Outputs of Generalized parcel divider 1.	64
3.5. Subdivision results of Generalized parcel divider 2.	67
3.6. Subdivision of a parcel with street looping inside.	69
3.7. Four types of configuration of the subdivided parcel with cul-de-sacs based on orientation of the parcel and position of existing roads	71
3.8. Subdivision layout generated for L- a) and opposite L-shaped b) parcels.....	72
3.9. Subdivision results of T-shaped Parcel Divider: a) parent parcel with darker shade; b) and c) two subdivision styles offered by the tool	73
3.10. Partitioning of a larger parcel into smaller multi-family residential lots.....	74
3.11. Comparison of the tool-generated subdivision of an irregular parcel.....	75
3.12. Validation of the simulation result.....	76
3.13. Output of Cul-de-sac Creator compared.	76
3.14. Modeled and observed lots compared.....	76
3.15. Results of Parcel-Divider tools in conjunction with an irregular cellular automata model of urban growth.	83
4.1. Adjacency neighborhood.	94
4.2. Proximity neighborhood.	96

4.3. Intercepted buffer neighborhood.....	98
4.4. Extent-wise neighborhood showing the core polygons exerting high impact on central polygon I and decreasing amount of impact along the distance from the focal polygon... ..	100
4.5. Distribution function of impact scores for polygons across the entire study area	101
4.6. Geographic location of San Marcos, Texas, the study site, and six land use categories used in the model	103
4.7. Results for scenarios defined by three selected neighborhood configurations.	106

ABSTRACT

The research goal of this dissertation was to build a model for simulating urban growth and producing different future scenarios. This study proposed the Agent-Integrated Irregular Automata (AIIA) model – a hybrid between the cellular automata and agent-based modeling. The model uses irregular geometries (i.e. vector data format) as the unit of operation.

This dissertation is comprised of three interrelated research projects that are summarized into distinct chapters. The first project focuses on the development and implementation of the AIIA model. The model was deployed to model the urban growth of San Marcos, Texas. By validating against empirical data, the results demonstrated that the AIIA model produces accurate future growth scenarios. The model contributes to the advancement of existing methodologies on urban modeling and its simulations provide useful insights for urban planning and policy making.

The second project delves into the neighborhood subdivision component of the AIIA model. It is important to focus on automation of subdivision of developable lands into residential parcels because meaningful execution of behavioral rules in the AIIA demands interaction at the household level. The *Parcel-Divider*, a GIS toolset was created to automate parcel subdivision and generation of urban layouts. By comparing with many real-world subdivisions, the resulting toolset generates uniform and regularly-shaped lots while maintaining egress and continuity of road networks to the newly developed areas.

The third study examines the impact of neighborhood configuration to the proposed AIIA model. It concludes that the AIIA simulations are highly sensitive to the type and size of the neighborhood and recommends that sensitivity analysis should be an integral part of calibration to urban growth simulation. This dissertation research contributes to the advancement of urban growth modeling.

1. INTRODUCTION

1.1 Urban growth modeling

Urban growth is a physical process of spatial expansion of an existing urban area. It occurs when undeveloped lands in and around an urban area are converted into built-up surfaces of varying density due to population growth and urban migration (United Nations 2005; Pacione 1995). Urban growth has been a common phenomenon in recent times. United Nations (UN) estimated that 3.5 billion people (a little more than 50% of the world population) were living in cities in 2010, while only about 25% of the world population was living in cities in 1950 (UN 2010). In the United States, urbanized land increased by 47% between 1982 and 1997, while the population went up by only 17% during the same period (Oguz et al. 2007). There were approximately 261 million Americans living in cities in 2010, and an increase of 40% is projected for the year 2050 (UN 2010).

Unregulated low density urban growth (especially in the form of sprawl) produces various environmental and social problems, such as the loss of natural vegetation and open space, degradation of natural habitats, increase of impervious surfaces, and production of waste materials and harmful chemicals (Duany et al. 2001). This type of growth is often criticized as energy inefficient and attributed as a major cause for resource depletion and environmental degradation (Nechyba and Walsh 2004). As a hindrance to sustainability, it has emerged as an issue to policymakers and the public alike. In order to devise effective control policies and management plans for urban sprawl, the concerned authorities must understand the spatio-temporal dimensions of

urban growth, including the physical, socioeconomic and cultural factors affecting the urban growth. Urban growth modeling helps shedding new insights on these dimensions, as well as providing information on data structures, predictions and perspectives (Bhatta 2010). Urban growth models are thus useful tools to understand the process of urban expansion and assess environmental impacts. They are vital to the formulation of sound policies and future plans.

There is a long tradition of modeling urban growth and the form of cities. Various urban models have been proposed in the literature, including bid-rent theory of agricultural land use (von Thunen 1826), concentric ecological rings model (Burgess 1925); central place theory (Christellar 1933; Losch 1945), ecological sectoral model (Hoyt 1939), and multiple nuclei model (Harris and Ullman 1945). Although these models paved a solid path toward exploring the macro patterns of urban structure, they were not without shortcomings.

The two major criticisms of these models are: a) they hold unrealistic assumptions and therefore are far from real-world cities and b) they are static and therefore unable to model the urban dynamics. Traditional urban models typically use a top-down approach and assume cities as static, linear and in a state of equilibrium (Cheng and Masser 2003). Thus, these assumptions fail to reflect the dynamics of complex systems, which have many components that interact in a nonlinear fashion whereby the whole is more than the sum of the components (Simon 1962). On the other hand, many researchers found cities have been exhibiting the major characteristics of complex systems: intricacies, emergence of patterns and forms, self-organization over time, self-similarity over scale, fractal dimensionality, and non-equilibrium state (Jacobs 1961; Portugali 2000; Batty and

Longley 1994; Batty 2005; Wu 2000). Given the complex nature of urban systems, new techniques have been developed to better model urban dynamics. Two emerging techniques in urban growth modeling are Cellular Automata (CA) and Agent-Based Modeling (ABM). Due to the proliferation of artificial intelligence techniques and object-oriented programming, these cell- and agent-based models have gained popularity in recent years.

1.2 Cellular automata and agent-based modeling of urban growth

Urban growth modeling entails the simulation of spatial expansion of a city over time. In addition to providing scenarios of future development, it reveals patterns of urban land use and land cover changes. Cellular Automata (CA) and Agent-Based Modeling (ABM) are two widely employed techniques of urban dynamic modeling.

Cellular Automata (CA) are generally defined as a multi-dimensional grid of identical cells. These automata are self-acting cells that process information, and their actions are guided by the specific transitional rules in an iterative fashion. Thus, the formalization of a CA model typically consists of five major components: cell space or grid, cell states, neighborhood influence, transitional rules, and time. The cell space consists of regularly tessellated cells (often squares) of same size representing discrete spatial extent. The cell states are described to be discrete (often binary e.g. live or dead). The cell states can also be stochastically defined using probabilities of a specific state. Operating based on the notion of neighborhood influence, the CA dynamic modeling utilizes the idea that state of a cell changes based on the state of surrounding cells within the neighborhood. The state transformation of the cell occurs based on certain specified rules called transitional rules. The change of state of an automaton (i.e. cell state) in the

simulation of a typical system such as urban expansion can be put in a formal notation as:

$$S_{t+1} \approx f(S_t, N, T_{(s)})$$

Where S is a set of cell states i.e. $S = \{s_1, s_2, \dots, s_n\}$, N represents a defined neighborhood, and $T_{(s)}$ means one or more transitional rule(s) applied. The function f defines the operation of transitional rules to transform the cell state from time t to $t+1$.

John Conway's "Game of Life" (Gardner 1970) is regarded as the first application of CA. Conway, in an attempt to explore micro-level spatial dynamics of population, devised a set of simple transitional rules capable of producing complex patterns. He used Moore's neighborhood for each cell representing two states of life – 'dead' or 'alive' – and used the following transitional rules:

1. Survivals: Every counter with two or three neighboring counters survives for the next generation
2. Deaths: Each counter with four or more neighbors dies (is removed) from overpopulation. Every counter with one neighbor or none dies from isolation.
3. Births: Each empty cell adjacent to exactly three neighbors – no more, no fewer – is a birth cell. A counter is placed on it at the next move.

Figure 1 illustrates that the black cells are alive and white are dead. When the abovementioned rules are applied each iteration/move, the cells die as shown in the blocks of cells.

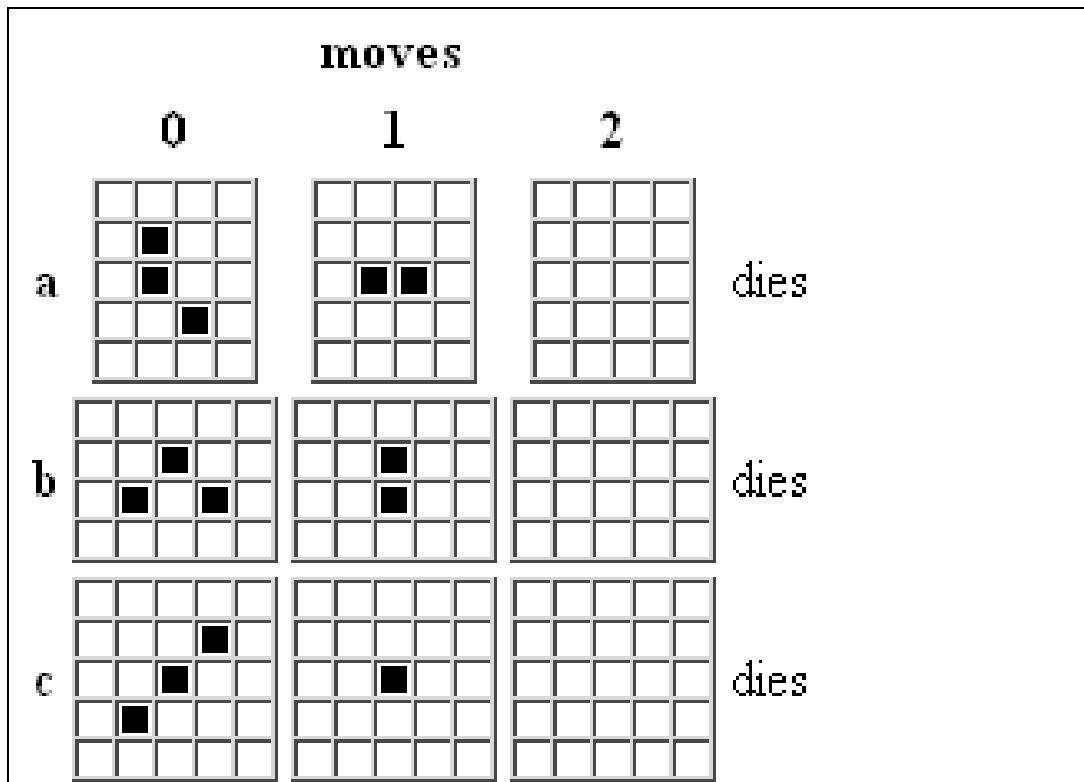


Figure 1.1. Cellular space and the moves of alive cells across time iterations (adapted from Gardner 1970).

First developed by von Neumann (1951), CA were introduced to the field of Geography in general and urban modeling in particular by Tobler (1979), and further espoused by other scholars (e.g. Couclelis 1985, 1997; Itami 1988; Batty and Xie 1997). A concept of constrained CA, whereby local impediments on homogenous spread of urban areas due to environmental factors and policies are captured, was pioneered by Roger White and his colleagues (White and Engelen 1997; Engelen et al. 1995). Various works adopted this framework to develop and apply CA models (e.g. Barredo et al. 2004; Vliet et al. 2009; Thapa and Mariyana 2011). A widely-used predictive CA model is SLEUTH, an acronym for Slope, Land-use, Exclusion, Urban, Transportation, and Hillshade that are the primary inputs of the model (Clarke et al. 1997; Clarke and Gaydos

1998). This model has been applied to a wide range of real-world cities (e.g. Yang and Lo 2003; Dietzel and Clarke 2006; Jantz et al. 2010). Researchers continued to improve existing urban CA models by integrating other techniques such as Multi-Criteria Evaluation (MCE) (e.g. Wu and Webster 1998), fuzzy logic (e.g. Wu 1998; Liu and Phinn 2003; Al-Ahmadi et al. 2009), Bayesian technique (e.g. Almeida et al. 2005), ant intelligence (e.g. Liu et al. 2008), support vector machine (e.g. Yang et al. 2008) and Artificial Neural Network (e.g. Yeh and Li 2003).

CA is an efficient geocomputation technique of simulating urban systems that have the characteristics of complexity, emergence and self-organization. However, CA modeling is ineffective to capture the interactions of mobile components of the urban systems such as migration and traffic (Torrens 2001). Moreover, the socioeconomic dimensions of the urban dynamics, such as human decisions and social behaviors, can only be implicitly represented by the transitional rules. They are more adapted to incorporate the influence of physical variables in the expense of social influence variables, for which the technique of Agent-Based Modeling (ABM) is more appropriate.

On the other hand, ABMs are computer representations of phenomena that consist of interacting virtual entities or agents (Brown 2006). The agents, the virtual entities that are capable of autonomous actions, represent actors in the real world. Residents, land developers and government officers (e.g. planners) are common examples of agents in the field of urban dynamic modeling. The agents are discrete, goal-oriented, adaptive and interactive (Macal and North 2006). They interact with each other or their environment, and perform some task based on pre-defined behavioral rules. This results in change either to the agents themselves or their environment. For example, feedbacks from and

interactions among the planner, farmer, housing developer and households result in land use change (i.e. conversion of agricultural land to residential development). A complex system such as a city has many components that are interdependent, heterogeneous and hierarchically nested (Parker et al. 2003). Representing these system components as agents, an ABM provides the flexibility needed to capture a wide range of system behaviors.

However, ABMs are computationally intensive and often problematic during their operationalization (Castle and Crooks 2006). Another demerit of ABMs is the difficulty they pose for their calibration, verification and validation (Crooks et al. 2008). Defining the agent's attributes and decision behaviors by using empirical data is often problematic (Benenson and Torrens 2004). Moreover, they are not as effective as CA models in capturing the spatial influences of neighborhoods and diffusion across the contiguous land units (i.e. the process of urban spread). Therefore, integrating cellular automata, a spatially-explicit model, with agent-based models is a natural extension of continual investigation on the topic of urban growth simulation (Wu and Silva 2010; Torrens and Benenson 2005).

One of the early works in this tradition is Torrens (2001), which introduces a conceptual framework for the hybridization of multi-agent systems and cellular automata. Ligtenberg et al. (2001) designed a technique of integrating cells and agents. Loibl and Toetzer (2003) developed an integrated software program that has basic functionalities characteristic of a GIS, cellular automata and multi agent systems. Drawing on the concept of what they called 'geographic automata systems,' Itzhak Benenson and his colleagues developed a prototype CA-ABM framework called Object-Based

Environment for Urban Simulation (OBEUS) as an urban geocomputation technique (Benenson et al. 2006). Sudhira et al. (2005) developed a tightly coupled CA-ABM model by using simple if-then rules. Using Netlogo platform, Torrens and Nara (2007) simulated inner-city gentrification. Li and his colleagues developed CA-ABM models based on the concepts of utility function and joint-probability (Li and Liu 2008; Li et al. 2011; Zhang et al. 2010).

These hybrid models have been developed in a cellular surface, i.e. a lattice of square cells. Although raster data model is more computationally efficient and convenient (i.e. many existing tools/functions are available) than the vector counterparts, the raster approach has certain limitations. The geographic and geometric details of the urban objects are captured more accurately using vector format (Crooks 2010). The cellular grids of equal size and regular shape are ineffective to represent irregular land features such as city blocks, subdivisions and parcels (Stevens et al. 2007). Similarly, the land use and land cover represented in raster data model appears to be more homogenous than it is in the real world. This problem is worsened as the spatial resolution increases (Kocabas and Dragicevic 2009). In order to address these limitations, researchers developed irregular CA models by modifying the basic CA assumption of squared-grid space (O'Sullivan 2001; Stevens and Dragicevic 2007). Irregular CA models developed using vector polygons to represent landscape features such as land use patches and cadastral parcels produce better results than the raster-based CA models (Moreno et al. 2009; Yumba and Dragecivic 2012).

Thus, an overview of the literature confirms that there are CA-ABM hybridized models employing cellular tessellation (i.e. the raster data format), and CA models

employing irregular tessellation (i.e. the vector data format), but not CA-ABM models employing irregular tessellation. Apparently, this warrants the development of a model that would integrate ABM into CA at the same time employing irregular tessellation to represent the geographic space. The present study aimed to fill the research gap by building a CA-ABM hybridized model, which is capable of capturing both the individual behaviors of actors involved in urban land use dynamics and simulating the influence of land use states of contiguous geographic units to land development. The overall research goal of the dissertation was to build a prototype simulation model with the capability to produce more accurate results than existing urban growth simulation models. Other generic research objectives were to (a) apply the developed model to simulate urban expansion of a real-world city and empirically validate the model outcomes, (b) test the model sensitivity to neighborhood configurations, and (c) empirically assess the accuracy of the simulated subdivision configurations.

1.3 Structure of the dissertation

The primary focus of urban growth modeling (UGM) is to realistically simulate the process of spatial expansion of cities and land use land cover change over time, apart from producing scenarios of future development (White et al. 1997; Li and Liu 2007). As illustrated in Figure 2, a typical UGM pipeline includes (a) estimation of demand for new development based on input data and parameters, (b) assessment of suitability score or change probability for each site, b) site selection for development in the next time step based on suitability or other criteria, and c) state conversion of the selected site (e.g. from undeveloped to developed or from forest use to residential use) (White et al. 2000; Sante et al. 2010). Sometimes, the pipeline consists of an additional step for the creation of

urban layouts such as roads, city blocks and cadastral lots in the selected site(s), specifically if the model employs vector data to represent the landscape at a finer spatial resolution (Benenson and Torrens 2004; Venegas et al. 2009a).

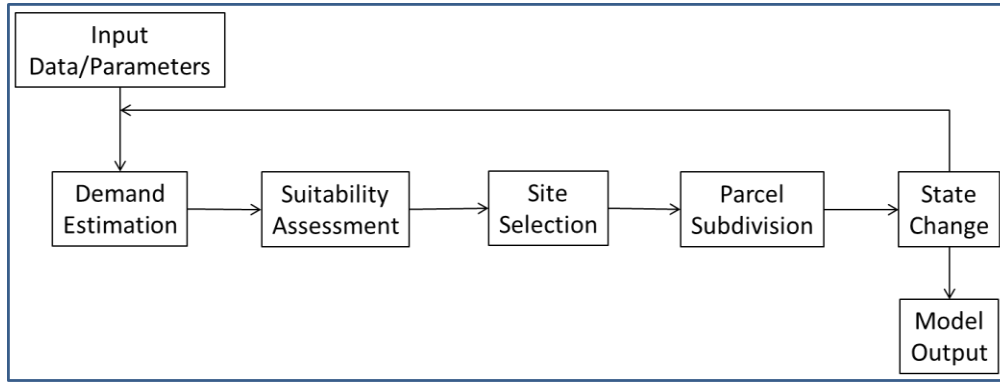


Figure 1.2. A flowchart showing the basic steps of an urban growth simulation modeling using vector data.

This dissertation is comprised of three interrelated research projects. The first project (Chapter 2) addresses the major research goal of developing a prototype software package for urban growth simulation by integrating CA and ABM. It focuses on building a model called Agents Integrated Irregular Automata (AIIA), its implementation, and validation of the model outcomes. The chapter describes the model in its entirety going through all the components of a typical UGM chain shown in Figure 2.

The second project (Chapter 3) elaborates in detail the fifth component (i.e. parcel subdivision) of the UGM chain. Partitioning of a selected parcel (site) into housing lots with a proper placement of road networks is an integral part of the UGM. Inclusion of such a subdivision mechanism in the model is important because meaningful execution of behavioral rules in the AIIA demands interaction of development agents at the household level. The inclusion also enhances the realism of the results in addition to enabling the model to take into account the bottom-up feedbacks from the micro-level components,

which ultimately determines the macro-level decisions.

The third project (Chapter 4) reports the results of a sensitivity analysis of the irregular cellular automata modeling to neighborhood configuration. This part of research does not implement the AIIA model in its entire form for conducting the sensitivity analysis, but uses a model that comprises only the irregular CA part with the agent-based component dropped. Yet, the research implications are directly applicable to the AIIA model as both projects employ irregular (vector data) tessellation to build the model and take into account the influence of neighborhood as part of modeling. Moreover, the same conceptual model described in Chapter 2 and the same code logic and algorithms for subdivision from Chapter 4 are used in the model described in Chapter 3. The investigation of neighborhood types and their impacts on model outcome is also relevant to the AIIA model because the AIIA model is also sensitive to neighborhood configuration, and neighborhood sensitivity is part of its model calibration.

Chapter 5 discusses the significance and limitations of the study before highlighting the future directions. The research objectives and questions corresponding to the three research projects of the dissertation are mentioned separately in the following subsections.

1.3.1 The AIIA model

This chapter details the development and implementation of the AIIA model. The main research goal was to build a prototype simulation model of urban growth by integrating agent-based modeling into irregular cellular automata at the finer spatial scale of cadastral parcels, and test the model by applying to a real-world city. The specific research objectives were to:

- Develop a hybrid model known as Agent Integrated Irregular Automata (AIIA) by using vector operations and parcels data at cadastral level,
- Advance existing understanding of urban growth modeling by examining:
 - The roles of various agents to formulate a holistic framework of urban growth modeling,
 - The incorporation of locational choice theories to derive meaningful behavioral and transitional rules, and
 - The impacts of urban growth policy to shed useful insights for planning and decision making.
- Model the urban land use maps for the City of San Marcos for years 2010 and 2020.

Based on the mentioned research goal and objectives, the project attempted to answer the following research questions:

RQ 1: Does incorporating commercial, industrial and institutional agents improve accuracy of modeling urban development?

RQ 2: How does the categorization of household agents into four different types and residential developer into two types improve the accuracy and precision of urban growth models?

RQ 3: What are the impacts of urban growth policy (e.g. introduction of growth boundary to encourage compact development) in modeled development?

RQ 4: What are the spatial patterns of modeled land development in San Marcos in years 2010 & 2020?

1.3.2 The parcel subdivision toolset

This chapter documents the development and implementation of software tools for automation of parcel subdivision. The main research goal was to build software tools for creating various subdivision styles so that the tools could be either integrated into urban growth simulation models such as the AIIA or be used as stand-alone application. The specific research objectives were to:

- Develop *Parcel-Divider*, a GIS toolset for automated subdivision of land parcels, containing various tools for creating a variety of subdivision styles to be applied to parcels of different geometric attributes, and
- Demonstrate the tools as a stand-alone application or to be used in conjunction with the urban growth simulation modeling.

Based on the mentioned research goal and objectives, the research attempted to answer the following questions:

RQ 5: What are the errors of the total number and size of simulated lots?

RQ 6: How closely do the modeled subdivisions resemble the observed subdivisions?

1.3.3 Neighborhood sensitivity of the irregular CA models

This chapter summarizes the results from a sensitivity analysis of neighborhood configurations on irregular CA/ABM models of urban growth. The main research goal here was to examine the impacts of neighborhood configuration on irregular CA models. The specific research objectives were to:

- Conduct a sensitivity analysis of the neighborhood size and type to the outcome of an irregular CA model,

- Explore various neighborhood types of irregular CA/ABM of urban land use dynamics, thereby expanding the existing scholarship on neighborhood typology, and
- Investigate variation in the urban development pattern due to variation in the neighborhood type and size.

Based on the mentioned research goal and objectives, the research attempted to answer the following questions:

RQ 7: What are the possible neighborhood definitions (types) in irregular CA/ABM models of urban growth?

RQ 8: How does neighborhood configuration affect the outcome of irregular CA/ABM modeling?

RQ 9: Which neighborhood type yields the most accurate results?

RQ 10: What kind of development patterns emerge as a result of variation in neighborhood types?

1.3.4 Conclusion

This chapter discusses the findings of all three research chapters from a holistic perspective. I further shed light on the overall significance and limitations of the dissertation research before highlighting the future research avenues. Limitations and future improvements specific to each of the three research chapters are also presented.

2. AN AGENT-INTEGRATED IRREGULAR AUTOMATA MODEL OF URBAN LAND USE DYNAMICS

ABSTRACT

Urban growth models are useful tools to understand the patterns and processes of urbanization. In recent years, the bottom-up approach of geo-computation, such as Cellular Automata (CA) and Agent-Based Modeling (ABM), is commonly used to simulate urban land use dynamics. This study has developed an integrated model of urban growth called Agent-Integrated Irregular Automata (AIIA) by using vector GIS environment (i.e. both the data model and operations). The model was tested for the city of San Marcos, Texas to simulate two scenarios of urban growth. Specifically, the study aimed to answer whether incorporating commercial, industrial and institutional agents in the model and using social theories (e.g. utility functions) improves the conventional urban growth modeling. By validating against empirical land use data, the results suggest that a holistic framework such as AIIA performs better than the existing irregular-automata based urban growth modeling.

2.1 Introduction

Cellular Automata (CA) and Agent-Based Modeling (ABM) are two commonly-employed techniques of urban growth modeling. Conventionally, CA formalism comprises of a regular tessellation of square cells, where the state of each cell (or automaton) is represented by its cell value and updated through iterations. CA use transitional rules to each cell based on its local neighborhood configuration to evolve complex global patterns (Batty 2005). On the other hand, an ABM typically consists of virtual entities or agents to represent real world actors such as residents who interact with other agents or geographic objects (Brown 2006). Similar to the automaton, each agent has its own states or attributes, which are updated in every time step based on behavioral rules. The primary distinction between these two simulation systems however is that the

agents are mobile and therefore able to move freely over the space as defined by certain movement rules (Torrens and Nara 2007). ABM provides an efficient framework to represent multi-actors and socioeconomic factors of land use dynamics (Parker et al. 2003; Castle and Crooks 2006) whereas CA are more effective in capturing spatial neighborhood influence in addition to being well-adapted to simulate diffusion across the contiguous land units (i.e. the process of urban spread) (Itami 1994; Couclelis 1997; O'Sullivan and Torrens 2000).

To capitalize the advantages of both systems, hybridization can result in symbiotic effectiveness (Benenson and Torrens 2004) and produce a more realistic simulation of urban dynamics (Li and Liu 2007). In a typical CA-ABM integrated model, cellular grid represents urban landscape whereas the actors or players of urban dynamics are represented as agents. However, robustness of the modeling can be enhanced by representing the urban geometries by vector polygons (Moreno et al. 2009; Jjumba and Dragicevic 2012).

The present study aimed to develop an Agent-Integrated Irregular Automata (AIIA) model capable of capturing both the individual behaviors of actors involved in urban land use dynamics and simulating the influence of land use states of contiguous geographic units to land development. The proposed model, which employs vector GIS (both data model and operations) at cadastral level, was tested for a mid-sized U.S. city. Emerging urban growth patterns of the city for the year 2020 were explored under different scenarios. In order to assess the model performance, the outcomes were compared to the outcomes of a baseline model, and validated against empirical land-use data.

2.2 Related work

Typically, CA and ABM models of urban growth are developed in a lattice of square cells, i.e. rasterized surface. Although the raster data model is computationally more efficient than the vector counterpart, it is less effective in capturing geographic and geometric details of the urban objects (Benenson and Torrens 2004). The grid cells of equal size and regular shape are incongruous to irregular land features such as city blocks, subdivisions and parcels (Kocabas and Dragicevic 2009; Crooks 2010). In order to address the limitation, researchers developed irregular CA models by modifying the basic CA assumption of squared-grid space to irregular geometries (e.g. Shi and Pang 2000; O’Sullivan 2001; Flache and Hegselmann 2001). These early works used Voronoi polygons and graph structures to represent the modeling landscape. However, they still fell short on realistic representation of cadastral structures. Stevens and his colleagues proposed land parcels as the basic spatial unit of operation in their CA model called *i-City* (short for irregular city) (Stevens and Dragicevic 2007; Stevens et al. 2007). Moreno et al. (2009) developed a vector-based cellular automata (VecGCA) model using irregular geographic objects (i.e. polygons) with the capability to dynamically update the shape, size and neighborhood of the objects. Obviously, these models had the same limitations of the CA modeling, and were unable to mimic the interactions and behaviors of actors such as households and planners involved in the process of urban land use dynamics.

Later, vector-based ABMs were developed by various scholars. Vanegas et al. (2009a; 2009b) attempted an integration of procedural generation of urban geometries into behavioral modeling. They devised a method of automatically updating urban layouts in the form of parcels and streets for every time step of the simulation. Waddell et

al. (2010) developed a micro-simulation model of land use dynamics at parcel-level by integrating an activity-based travel model. These studies focused on the economic analysis of discrete spatial choices made by resident agents, overlooking other geographic factors and neighborhood influences. Augustijn-Beckers et al. (2011) used vector-GIS layers to develop an agent-based housing model for simulating the growth of informal settlements. In order to capture the individual behaviors of concerned stakeholders, Jjumba and Dragicevic (2012) developed *Agent i-City* – an agent-based model for simulating urban land use change. The model uses heuristic rules to simulate the decision behaviors of urban planner, housing developer and household agents.

It was common to these ABM models to primarily focus on behaviors of resident (or household) agent, ignoring the roles and interactions of other players of land development such as retailers and industrial agents. As firms and industries are the major activity centers in a city and have their own locational preferences, their distribution over space directly affects the spatial decision of residents (Blair and Premus 1987; Levinson and Krizek 2008). Likewise, single family housing developers have different shape, size, proximity and other criteria of site selection from their multifamily counterparts (Kone 2006). For a realistic simulation of the urban growth process, it is important to take into account the interactions and preferences of all these actors. However, existing ABM- or CA-based simulation does not consider the interactions of different classes of residential agents with commercial, industrial and planner agents in a holistic manner. Thus, there is a lack of understanding in assessing their roles in urban dynamics.

Drawing upon the mentioned literature, the present study has developed an Agent-Integrated Irregular Automata (AIIA) model that uses vector polygons in the form of land

parcels at cadastral level. In addition to reflecting neighborhood influences, the model provides a mechanism for incorporating decisions of and interactions among different actors participating in the process of urban development. The specific research questions answered in this project include: a) Does incorporating commercial, industrial and institutional agents improve accuracy of modeling urban development? b) How much does the categorization of household agents into different types enhance robustness of the urban growth models? c) What are the impacts of urban growth policy (e.g. introduction of growth boundary to encourage compact development) in modeled development? And d) what are the spatial patterns of modeled land development in San Marcos in years 2010 & 2020?

The holistic approach adopted in this study has extended the existing framework by taking into consideration the behaviors of commercial firms, industrial entities, public institutions, and both single family and multifamily households. The decision behaviors of the agents in the model are defined by using utility maximization function. In order to assess the performance of the developed model, the outcomes were tested against the results of a baseline model that assumes invariant behaviors of single-family residential and multi-family residential agents, while randomly updating a proportional amount of lands for industrial, commercial and public/institutional development. The development and implementation of models in this chapter and their comparison is summarized in Figure 2.1. The AIIA and a baseline model were developed separately, and then run using the datasets in the year 2000 (base year) to produce predicted urban growth for the year 2010. The simulated outputs for both the models were validated against the observed data for the year 2010. Then, using only the AIIA model, two future growth scenarios of the

study site were produced for the year 2020, this time 2010 being the base year.

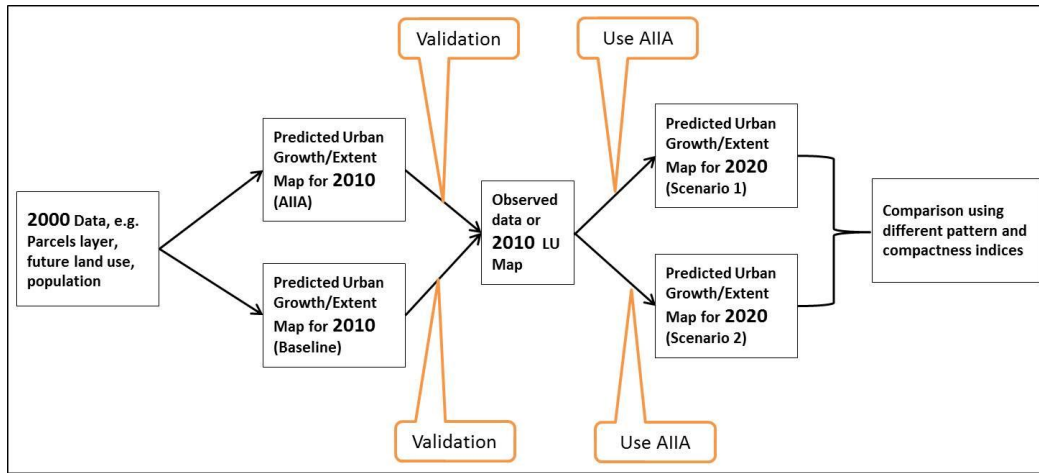


Figure 2.1. A diagram providing a synopsis of the modeling process.

2.3 Study site, data and software

The study area is San Marcos, a mid-sized city located in the Austin-San Antonio Corridor, centering along the Interstate Highway 35 (Figure 2.2). This corridor region is one of the fast growing regions in the United States, with an expansion of urban area by more than 200% in the last two decades (Voughan 2008). The centralized location of the city in this corridor has contributed much to its speedy growth. Several companies have chosen a site in San Marcos due to its proximity to the large urban markets of Austin and San Antonio. Furthermore, the size of the city is appropriate for applying the developed model. Since AIIA aims to simulate micro-level geographic objects operating at the fine scale of cadastral parcels, modeling bigger cities such as Austin or San Antonio is comparatively less efficient in terms of computation and tracking of the complex intermixes of model components.

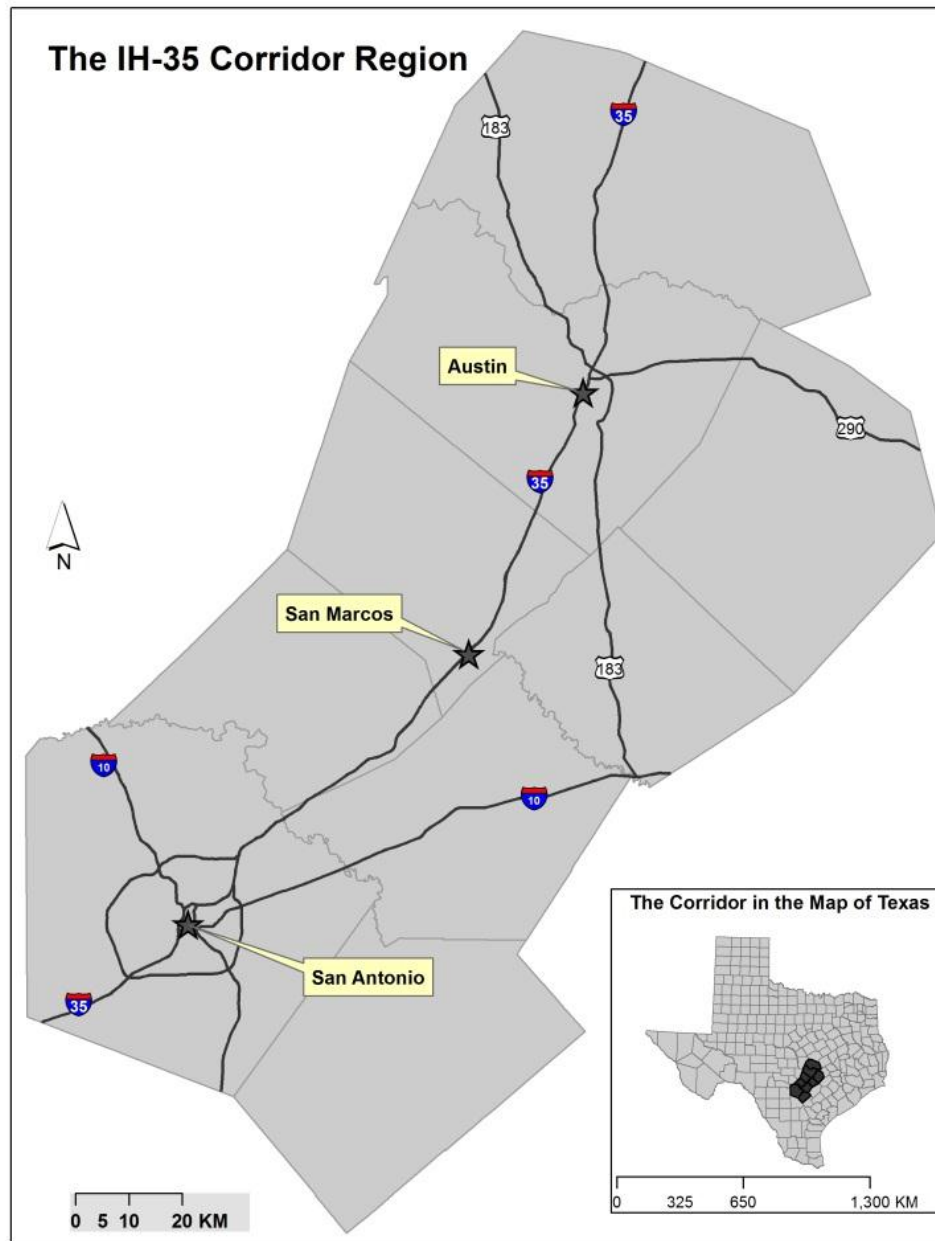


Figure 2.2. Location of San Marcos in the Austin-San Antonio corridor of Texas, USA.

Two major datasets required for this research are demographic and cadastral records of the study area. Demographic data including total population for the years 2000 and 2010 and annual growth rate were obtained from the US Census Bureau. Cadastral data were obtained from the City of San Marcos. The projection of developed

lands for future years was supplemented by the analyses of land use changes based on satellite-derived classified maps and National Agriculture Imagery Program (NAIP) imageries for the years 2000 and 2010. Other GIS layers, including elevation, geologic formations, Federal Emergency Management Agency (FEMA) Q3 floodplain, and parks were obtained from the city source as well. The research utilized Environmental Systems Research Institute's (ESRI) ArcGIS suite to prepare, analyze and visualize datasets. The AIIA model has been built using Python programming language.

2.4 Conceptual framework

The conceptual framework expounded here explores how the real-world process of urban land use dynamics is mimicked through the action and interaction of various agents. In the proposed model, cadastral polygons function as irregular (cellular) automata, which have their own states and properties. The automata states and agent characteristics are encoded as the attributes of the vector polygons. Change in land use occurs as a result of change in the state of automata, which is directly associated with the behavior of agents. Changes are updated each iteration, which is equivalent to a year in this study. The whole process of urban growth is illustrated in the framework shown in Figure 2.3.

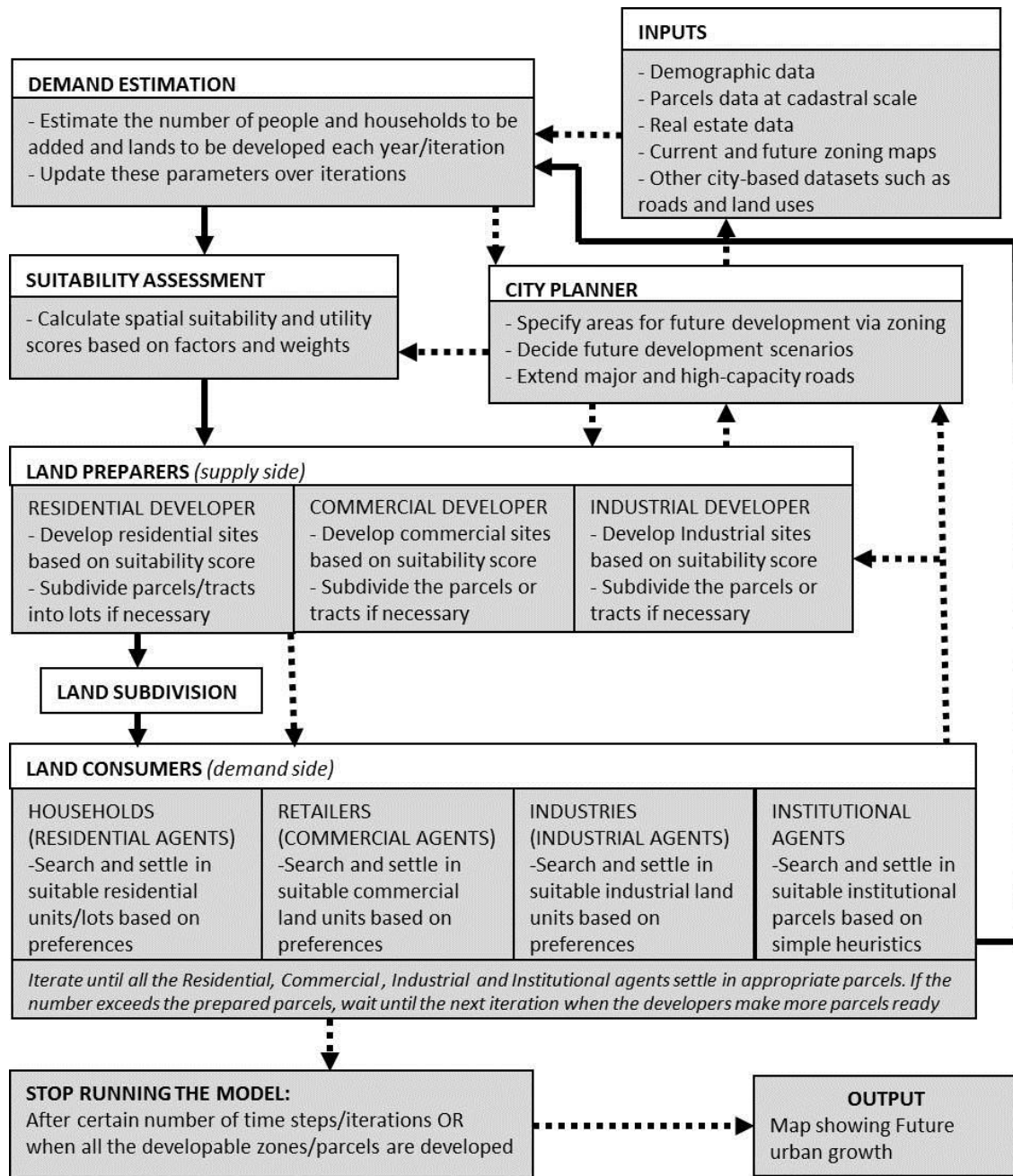


Figure 2.3. Sequential components of the conceptual framework of AIIA and interactions among agents.

The first step in the model is to determine the demand of land for different uses to be developed at time $t+1$ by estimating the net population growth through natural birth and immigration. Based on the user-specified values for total population, annual growth

rate and average household size, the number of new households coming into the city for each time step is calculated as:

$$HH_{t+1} = (P_t * G / 100) / AH \quad (2.1)$$

where HH_{t+1} is the number of new households at time $t+1$, P_t is the population at time t , G is the annual growth rate in %, and AH is the average household size. The base year (i.e. 2000) values of G , AH , and P_t for the study area was 2.6%, 2.8 and 47,500 respectively (City of San Marcos 2006). Next, total area required for development at a time step in residential, commercial, industrial and institutional land use zones is estimated based on ratios of developed lands in each category for the base year.

Having received information from the demand estimation component (Figure 2.3), land developer agents prepare lands equal to the amount demanded for a particular time step. They select the land based on the suitability of available lands. The suitability assessment involves computation of suitability score for every developable land site (parcel) in the study area by identifying local influence factors and respective weights assigned by the developers to these factors. The influence factors or variables included in this study, (i.e. applicable to the case of San Marcos) and the weights are shown in Table 2-1.

Table 2-1. Weights or preferences of various agents on the influence factors (driving variables) of urban expansion applicable to the City of San Marcos, Texas.

Factors or variables		Single family developer	Multi-family developer	Commercial developer	Industrial developer	Low income household with kids	High income household with kids	Low income household without kids	High income household without kids
Environmental	Elevation	1.8	0.0	-3.0	0.0	0.0	1.3	-0.9	1.1
	Slope	-1.2	0.0	-7.8	-7.7	0.0	-1.1	1.0	-1.1
	Distance to rivers	0.4	0.0	1.3	0.0	1.3	-0.5	1.1	0.0
Market	Population density	0.0	4.0	-2.4	-17.6	0.9	-0.8	1.7	0.4
	Land value	-4.4	4.2	12.5	0.0	-3.0	-3.1	0.0	-2.5
Transportation accessibility	Distance to IH-35	-5.6	-4.0	-3.9	0.0	-1.5	1.8	-1.9	3.4
	Distance to railroads	0.0	0.0	3.9	-1.7	1.6	0.0	1.0	-2.3
	Distance to major roads	-4.9	-4.1	-4.1	-2.4	2.6	1.0	1.2	0.0
	Distance to airport	-3.8	-2.9	2.0	-2.9	-6.4	-3.1	-2.6	-0.6
Centrality influence	Distance to Texas State University	1.2	-14.1	-31.4	0.0	10.4	10.2	-12.9	6.2
	Distance to City hospital	-1.2	3.5	-2.4	1.7	0.0	2.5	0.0	3.4
	Distance to San Marcos outlet mall	-0.5	-3.9	-2.0	-2.8	-3.5	-4.6	-2.1	-3.7
	Distance to CBD	-0.9	7.2	29.4	0.0	-8.4	-10.9	12.4	-9.6
Development influence	Residential neighborhood	8.8	-4.0	-10.7	-15.6	2.2	3.0	1.4	2.0
	Commercial neighborhood	-5.5	-5.6	17.3	-14.1	-6.7	-10.8	-5.5	-7.8
	Industrial neighborhood	-6.0	0.0	-4.4	13.1	0.0	0.0	0.0	0.0

Here, the term accessibility refers to the ease of reaching to a location such as the outlet mall from a parcel. A high accessibility score indicates a short Euclidean distance from the parcel to a specific destination. In this study, neighborhood of a parcel is defined as the buffer of a specific distance from its boundary, which was calibrated to be 200 feet on the basis of trial and error. Development influence is parameterized as the number of developed cadastral units within the neighborhood. For example, the variable ‘residential

neighborhood' indicates total number of developed residential lots in the neighborhood. Development suitability score for each parcel is calculated based on the Weighted Linear Combination (WLC) function. Different agents assign different weights to each factor. The weights were estimated by using logistic regression based on existing land use and census household characteristics in 2000 (Table 2-1). The suitability score for each parcel is derived using the following equation after Wu (2000) and Barredo et al. (2003):

$$S_i = \sum_i^n (W_i * F_i) + \varepsilon \quad (2.2)$$

where, S_i is the suitability score of parcel i , F_i is the value of a driving factor F for parcel i , W_i is the weight of factor F , n is the number of factors included (16 in this case) and a stochastic term ε . This composite suitability score for each parcel is calculated with different preferences (weights) for different agents as mentioned in the below section.

2.4.1 Agents of land development

Different actors contribute to the production of a built environment. In free-market societies, a typical land development industry comprises of a series of actors such as speculators, developers, builders, contractors, real estate agents and engineers (Pacione 2005). Following Zellner et al. (2009), and Jjumba and Dragecivic (2012), the present model simulates the behaviors of major actors of a typical U.S. city (i.e. San Marcos), including city planner, land developers, households, retailers (business firms), and industrialists.

2.4.1.1 City planner

In the model, city planner is used as an umbrella term that includes all governmental and non-governmental representatives who have direct or indirect role(s) in the making of urban plans. There is only one planning body (i.e. agent) for the city.

Although the planner agent oversees all over the study area, it does not possess any spatial location in the modeled landscape. The planner agent performs the task of extending major transport networks such as highways. It is assumed that other infrastructures such as utility lines are automatically extended with the roads, and all the applications lodged for development permit are approved. The city planner is also responsible for making decisions about various future development scenarios. This study has chosen to simulate two growth scenarios: the first scenario with the current growth trends and the second scenario in which compact growth is encouraged. The planner agent plays a vital role in formulating zoning plans and allocating available lands into appropriate land uses. The proposed AIIA model utilizes future land use map of the study area, in which all the parcels have been assigned one of the seven zoning classes: single family residential, multi-family residential, commercial, industrial, public and institutional, future roads, and restricted or open space (Figure 2.4). The parcels allocated for future development by the planner agent have ‘undeveloped’ state, which is later converted to ‘prepared’ by developers. Different agents convert the ‘prepared’ state into ‘developed’ in the course of simulation.

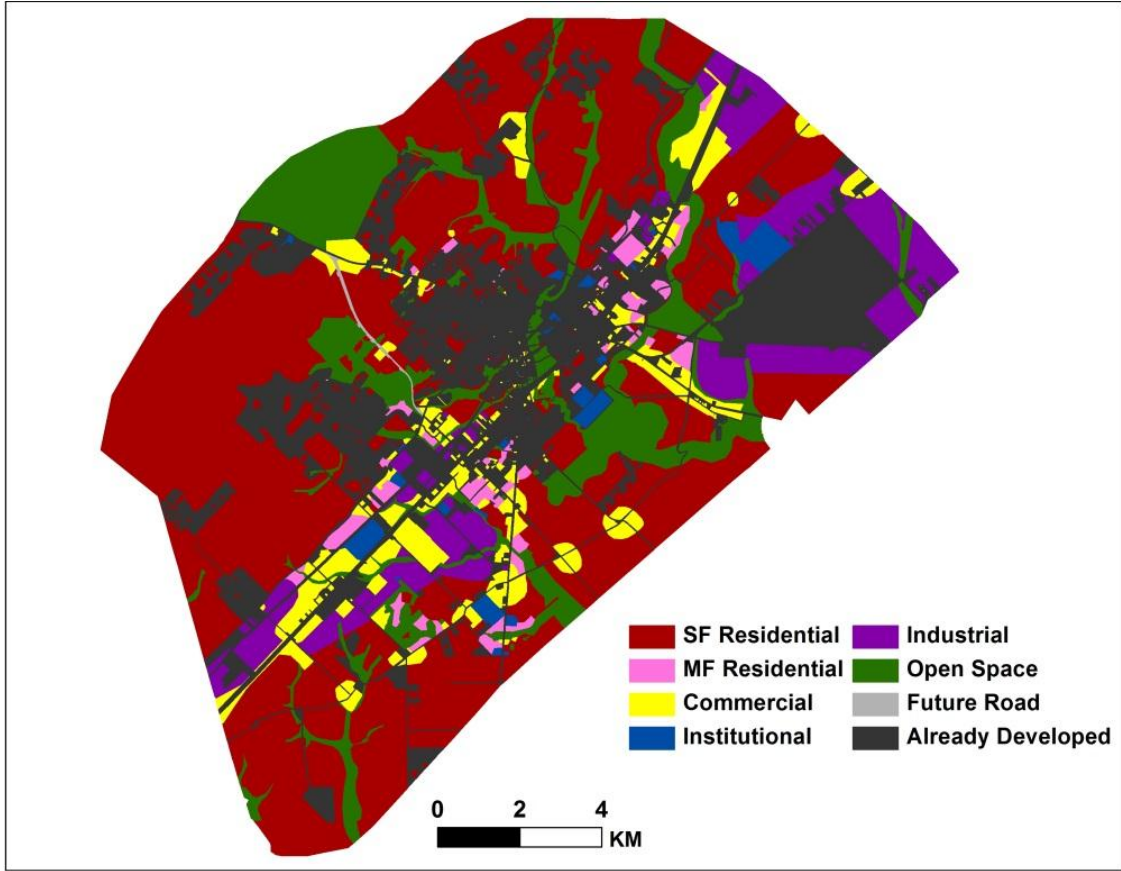


Figure 2.4. Sites allocated by the planner agent into different future land use classes. The ‘Already Developed’ class indicates developed area in base year 2000.

2.4.1.2 Land developers

The modeling assumes three types of land developer: residential, commercial and industrial. Residential developers are further categorized as single family (SF) housing developer, and multi-family (MF) or apartment developer. A general objective of the developer agent is to find (i.e. buy) ‘undeveloped’ parcels within its corresponding land use zone and subdivide them into smaller lots before turning the state to ‘prepared.’ In order to avoid a confusing usage of the terms ‘parcel’ and ‘lot’, I use the word ‘parcel’ to mean an undeveloped land tract to be subdivided into smaller areal units called ‘lots’ of ‘prepared’ state. In this sense, a lot is undividable anymore and has either ‘prepared’ or

‘developed’ state. Selection of the development site is based on the concept of profit maximization. The developer chooses a parcel that offers a higher amount of profit. In this model, profitable parcels are those that have the higher suitability score derived using Equation 2.2. The four types of developers assign separate weights to the factors included in the equation (Table 2-1). The weights were derived using the technique of logistic regression by feeding land use data in 2000. The aforementioned 16 factors were used as independent variables and the development status (i.e. developed or undeveloped state) of the parcels of corresponding land use as the dependent variable. For example, to derive weights of SF residential developer to the factors, development status of each polygon in terms of whether it is SF developed or not was used as the dependent variable. The regression coefficients were derived using backward regression method, and hence the significant variables would have a non-zero weight in Table 2-1.

Thus, the developer agents build up the lands. First, the amount of lands required to accommodate the incoming households for each time step in the SF and MF residential categories is computed. SF residential area is calculated as:

$$SFRA_{t+1} = SFHH_{t+1} * ALS \quad (2.3)$$

where $SFRA_{t+1}$ is the total SF residential area to be developed at time $t+1$, $SFHH_{t+1}$ is the number of incoming SF households at time $t+1$, and ALS stands for average lot size, a user-defined value. Similarly, MF residential area is calculated as:

$$MFRA_{t+1} = MFHH_{t+1} * AMFUS \quad (2.4)$$

where $MFRA_{t+1}$ is total MF residential area to be developed at time $t+1$, $MFHH_{t+1}$ is the number of incoming MF households at time $t+1$, and $AMFUS$ stands for average multi-family lot size, which is defined by the user. If there are not enough prepared lots for the

incoming households in their respective land use, undeveloped parcels with the highest suitability score in descending order are selected for subdivision until the total area of selected parcels is more, or equal to the required land area. Subdivision occurs if the selected parcel is larger than a user-specified size, which varies among the developer types. The pseudo-code of the subdivision procedure is presented in Figure 2.5.

Algorithms used in the subdivision component of the model were used as mentioned in Chapter 2.

Let P be a candidate parcel for partition, L lot, B block, W lot width, MF multi-family residential, SF single family residential, R existing road, S street to be created, MBR minimum rectangle boundary of P , COM commercial, and IND industrial:

```

for  $P$  in  $P_{list}$ :
    if  $P$  is disjoint to  $R$ :
        create  $S$  to connect the nearest point of  $R$ 
    else:
        pass
    if  $P = SF$ :
        if  $P_{size} > 500$  acres (i.e. very large tract):
            split  $P$  into two  $P$ 's (i.e. binary division)
            mark  $P$ 's as 'Undeveloped' and put them as  $P$  in  $P_{list}$  for next iteration
        elif  $P_{size} \leq 500$  acres and  $P_{size} \geq 20$  acres:
            split recursively into  $B$ s until  $B_{size} < \text{user-assigned block size}$ 
            create streets around each  $B$ 
            divide each  $B$  into two rows of  $L$ s of user-assigned  $W$ 
            assign 'prepared' state to the resultant  $L$ s
        elif  $P_{size} < 20$  acres:
            if short axis of  $P_{MBR} < \text{long axis of } P_{MBR} * 0.5$ :
                split  $P$  into a single row of  $L$ s
                assign 'prepared' state to the resultant  $L$ s
            else:
                create a cul-de-sac road and divide  $P$  into a row of  $L$  on either side
                assign 'prepared' state to the resultant  $L$ s
    elif  $P = MF$ :
        if  $P$  Not adjacent to  $R$ :
            create  $S$ s around  $P$ 
        else:
            pass
        if  $P_{size} > \text{user-assigned average } MF_{size}$ :
            binary split into  $P'_{MF}$  until  $P'_{MFsize} \leq \text{user-assigned average } MF_{size}$ 
            create an  $S$  street along the dividing line
            mark  $P'_{MF}$  as 'prepared'
    elif  $P = COM$ :
        if  $P$  Not adjacent to  $R$ :
            create  $S$ s around  $P$ 
        else:
            pass
        if  $P_{size} > \text{user-assigned average } COM_{size}$ :
            split recursively into  $P'_{COM}$  until  $P'_{COMsize} \leq \text{average } COM_{size}$ 
            create an  $S$  along the dividing line
            mark  $P'_{COM}$  as 'prepared'
    elif  $P = IND$ :
        if  $P$  Not adjacent to  $R$ :
            create  $S$ s around  $P$ 
        else:
            pass
        if  $P_{size} > \text{user-assigned average } IND_{size}$ :
            split recursively into  $P'_{IND}$  until  $P'_{INDsize} \leq \text{average } IND_{size}$ 
            create an  $S$  along the dividing line
            mark  $P'_{IND}$  as 'prepared'

```

Figure 2.5. Pseudo-code of the subdivision component of the AIIA model

The commercial and industrial developers prepare lands similarly to the MF residential developer. The amount of commercial land to be developed in time $t+1$ is

calculated as:

$$CA_{t+1} = TRA_{t+1} * CR_t \quad (2.5)$$

where, CA_{t+1} is commercial area to be developed at time $t+1$, TRA_{t+1} is the residential area to be developed at time $t+1$ (i.e. computed by adding $SFRA_{t+1}$ and $SFRA_{t+1}$ from the previous paragraph), and CR_t is the ratio of developed commercial area to the developed residential area at time t . Undeveloped parcels with the top suitability scores within the commercial land use zone are selected for development until the total area reaches the estimated amount. The parcels are assigned the development status of ‘prepared’ if they are smaller than the user-assigned value of average commercial lot size. Larger parcels are subdivided and updated to the status of ‘prepared’. The lots developed in this way tend to be of uniform size. Industrial development occurs in exactly the same way as that of commercial development, the only difference being that land for development is selected from within industrial land use instead of commercial use.

2.4.1.3 Household agents

A household is the basic residential unit that includes all of the persons occupying a housing unit. In the model, an SF residential lot is occupied by a household, while an MF lot accommodates more than one household. At every time step, incoming households equal to the number estimated by Equation 2.1 are introduced into the modeling landscape, where they search for the most suitable locations based on preferences. When they find the lot of their desire, they settle and change the state of the ‘prepared’ lots to ‘developed’ state.

Households have a multitude of attributes. For example, they are recognized by different characteristics such as race/ethnicity (Black, White, or Asian). Similar to Li and

Liu (2007), this study only considered two major attributes to simulate the households' spatial choices: income and household size. Thus, the households can be either high-income or low-income agents based on the economic criterion. An annual median income of \$35,000 was adopted as the threshold for categorizing these two groups following the City of San Marcos (2006). Similarly, they can be either a household with children aged 18 years or less or a household with children aged 18 or less based on the demographic criterion. Thus, all households are grouped into one of the four categories: low-income with children (LC), high-income with children (HC), low-income without children (LNC) and high-income without children (HNC).

The number of incoming households for each of the four groups is estimated based on population proportions. Using Census 2000 data, I found the proportions to be 11%, 24%, 46% and 19% for LC, HC, LNC and HNC respectively. The model automatically computes the proportions and the number of incoming households for each of the categories over iterations. At every time step, the estimated number of agents for each category enters the modeling landscape and search through 'prepared' lots. The agent finds and settles in a lot that maximizes its utility function, thereafter updating the state to 'developed'. The module runs until all household agents for the specific time step find, and settle in appropriate lots. But if the number of agents is larger than the available prepared lots, the agents that are unable to find a lot will wait for the next round of iteration. The information about the remaining households is updated to the housing developer so that it prepares additional lots in the coming time step. In the model, agent's attributes do not change. For example, an LC agent remains LC throughout the simulation period, despite the real world fact that a low income household could turn into

a high-income status after a few years or vice versa. Similarly, a lot once developed will remain developed forever. Competition among agents for finding a suitable geographic unit is realized through utility scores. The “prepared” lot with the highest utility score gets developed first.

Obviously, the household agents have different preferences and they assign different weights to different factors. In the selection decision, the housing agents optimize the utility function. The following utility function was used after Brown et al. (2008):

$$u_{(i,k)} = \prod_{f=1}^n (\gamma_{(i,f)})^{\alpha_{(k,f)}} + \partial_i \quad (2.6)$$

where $u_{(i,k)}$ is the utility of polygon i for resident type k , $\gamma_{(i,f)}$ is the value of factor f for polygon i , $\alpha_{(k,f)}$ is the weight resident k places on factor f , n is the number of factors evaluated (i.e.16), and ∂_i is the uncertainty term created by a random number generator and accounts for the uncertainties inherent to the decision making of the individual agents. Similarly, the weights were derived by using logistic regression as explained earlier (Table 2-1).

2.4.1.4 Commercial, industrial and institutional agents

What defines a commercial entity is broad and often vague. For instance, commercial entities range from a small one-room hair salon to big regional supermarkets. For convenience, all of the commercial entities have been categorized as a single type. The model assumes that one commercial agent occupies a commercial lot, although more than one commercial entity can be found in a single lot in reality. The model computes the number of incoming commercial agents for time $t+1$ (i.e. for every time step) as:

$$CUT_{t+1} = HHT_{t+1} * RCUT_t + WCUT_t \quad (2.7)$$

where CUT_{t+1} is the number of new commercial units entering the landscape at $t+1$; HHT_{t+1} is the total number of incoming households; $RCUT$ is the ratio of commercial units to residential units at time t which is calculated by dividing the number of commercial units by total number of residential units at time t ; and $WCUT_t$ stands for waiting commercial units at time t . The commercial agents enter the modeling landscape to search for appropriate ‘prepared’ lots anywhere within the commercial zone based on their preferences. In the AIIA model, the preference of commercial agents is taken to be the same as the suitability score assigned by commercial developer. Lots with the highest suitability score are chosen by the agents in descending order. Once the agent settles in the selected ‘prepared’ unit, it converts the unit’s state to ‘developed’. If the number of commercial agents is larger than the number of available ‘prepared’ lots, the remaining number of the agents will wait for the next iteration. The information about the deficit is received by the developer agent as a feedback so that it can prepare enough lands in the next iteration.

Industrial agents function the same way as the commercial agents, except that the incoming industrial agents settle in the lots prepared by industrial developer. The number of incoming industrial agents is computed in a similar way using Equation 2.7, but by replacing the term ‘commercial’ by ‘industrial’. Likewise, the number of public institutions such as schools and amenities to be set up annually in the city is estimated using equation 2.7 except that the ‘commercial’ is replaced by ‘institutional’ land use. Total area for new development due to incoming institutional entities is computed by multiplying the number by user-defined average institutional lot size. Developable parcels within the land use zone are selected randomly until the total area of the selected

units equals or exceeds the estimated amount. Then, the state of the selected parcels is converted to ‘developed.’

2.5 Results and validation

Major outputs of the AIIA model include maps that delineate predicted urban growth of the city of San Marcos for a user specified time interval such as the years 2010 and 2020. In order to assess the performance of the model, the output was compared with the empirical land use map of the study area in 2010. The model results were also compared with the outputs of a baseline model that is in par with the existing irregular urban simulation modeling. The baseline model is same as the AIIA except that it excludes simulation of decision-making of commercial and industrial agents. Simply the amount of land to be developed for each time step is calculated and the undeveloped parcels are selected randomly from among their respective land-use categories. The baseline model does not categorize the household agents into four categories; all the households are taken to be just one category. There is also no differentiation between SF and MF residential units, i.e. both SF and MF developer agents are lumped as residential developer. The suitability score for residential developer is computed using Equation 2.2, but with different weights from the ones derived for the AIIA model. Similarly, all the four agent types are lumped into one household category and their utility function is computed using Equation 2.6.

The urban growth in 2010 and 2020 was simulated in terms of development status (Figure 2.6), and land use categories (Figure 2.7) over time. The reference data shows a steady growth of the city between 2000 and 2010, with a total area of 1615 hectares developed within the period. Results of non-site-specific thematic accuracy assessment

(Congalton and Green 2009), i.e. comparing only the observed and modeled total area, reveal an overall accuracy of 96% for AIIA and 98% for the baseline model.

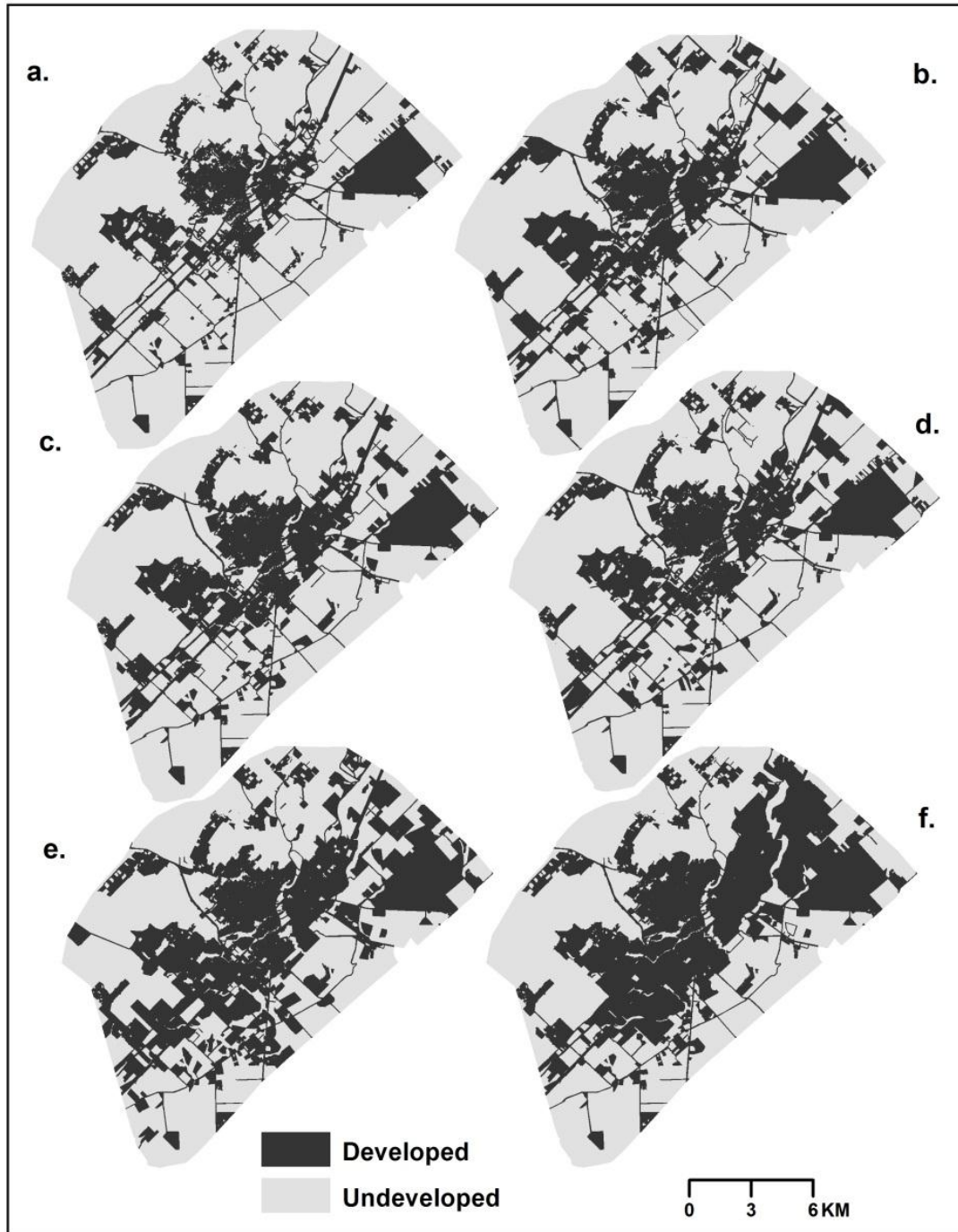


Figure 2.6. Urban growth in terms of development status, i.e. developed or undeveloped lands. a) Empirical map of the study area in 2000; b) reference map for 2010; c) output of AIIA for 2010; d) output of the baseline model for 2010; e) growth prediction of current growth scenario 2020;, and f) prediction of compact growth scenario 2020.

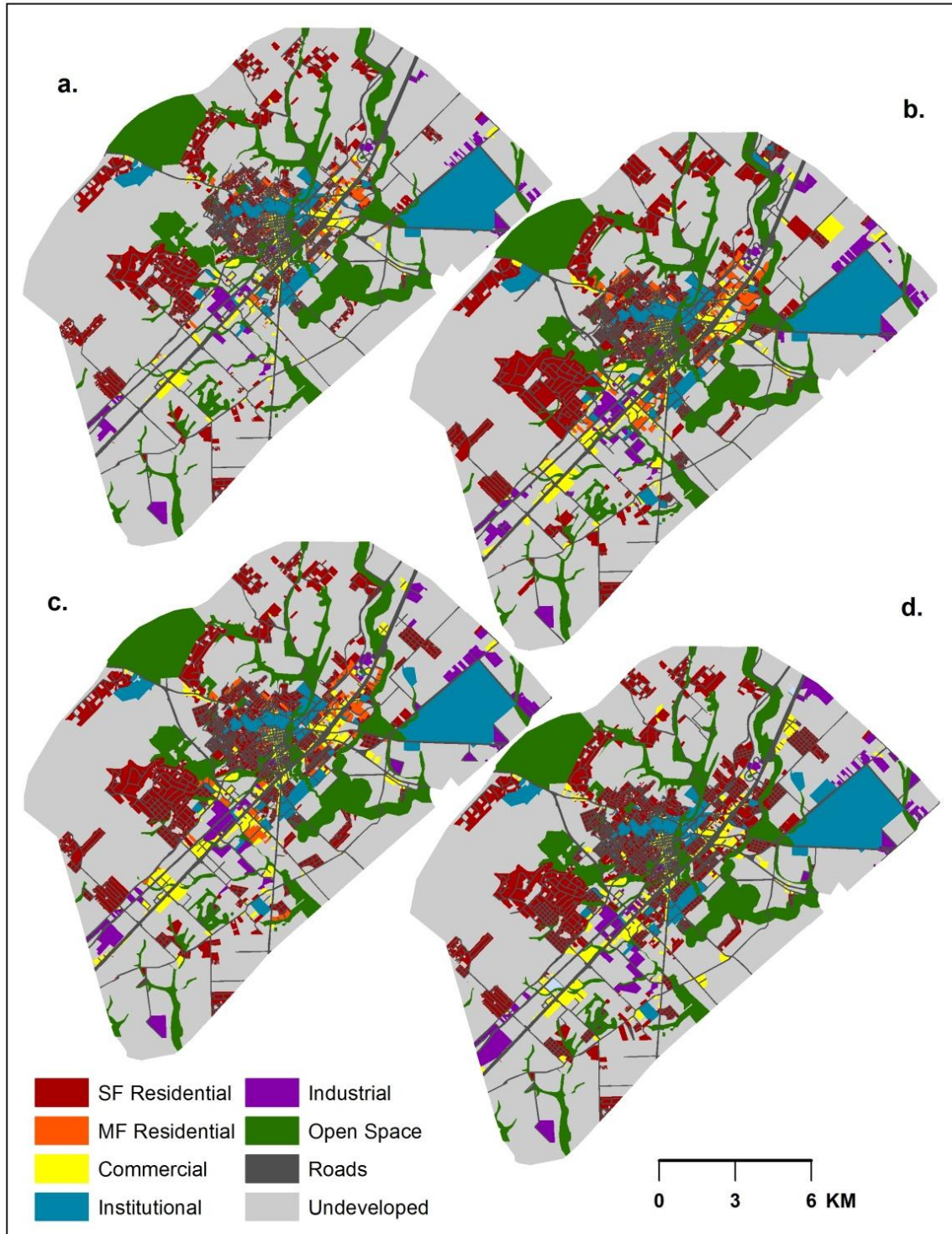


Figure 2.7. Urban growth in terms of land use zones of developed sites. a) empirical map of the study area in 2000; b) reference map for 2010; c) AIIA output for 2010; d) output of the baseline model output for 2010.

As the spatially-invariant accuracy index does not reveal anything about spatial accuracy of the modeled results, a site-specific assessment, in which matched locations for individual categories between the modeled and reference maps are tracked (*ibid*, p. 15), was carried out. An error matrix comprising a square array of cells holding areal values of matched and unmatched sites between the maps was created. Since the study employs vector data model, this was accomplished by applying the “union” overlay technique, where the modeled (or predicted) map and the reference map are overlaid in GIS. The areas that are intersected are delineated as the matched area in terms of specific states (e.g. developed or undeveloped). As shown in Figure 2.8, the newly created polygon 3 (in Figure 2.8-C) is the matched area in both the reference map (Figure 2.8-A) and the predicted map (Figure 2.8-B). It shows the accuracy of prediction. Polygon 2 and Polygon 4 in Figure 2.8-C represent the error of omission and error of commission respectively. Based on the confusion matrix (Table 2-2), indices of producers’ accuracy, users’ accuracy and overall accuracy were computed for both the AIIA and the baseline models.

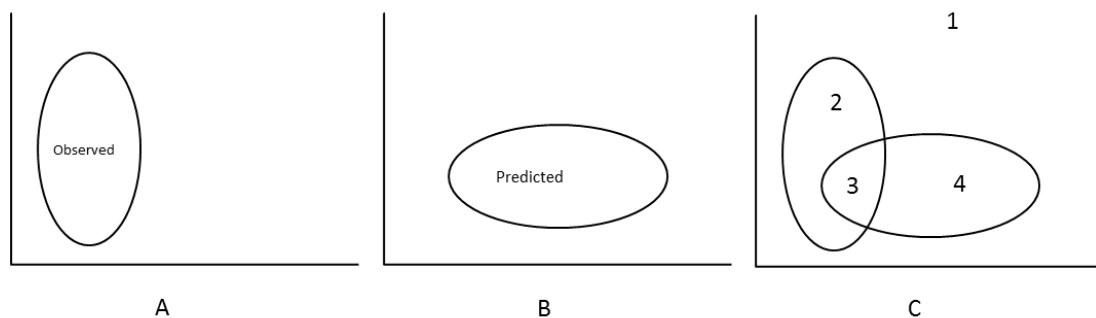


Figure 2.8. Illustration of error estimation using the technique of spatial overlay of polygons.

Table 2-2. Site-specific accuracy assessment of the AIIA model outputs. The values in parentheses are for the results of the baseline model (area in hectare).

Simulation Results	Reference Data								
		SF Residential	MF Residential	Commercial	Institutional	Industrial	Undeveloped	Row Total	User's Accuracy (%)
	SF Residential	497 (546)	7	0 (0)	0 (1)	0 (0)	308 (405)	813 (952)	61 (57)
	MF Residential	3	92	0	0	0	41	137	68
	Commercial	0 (2)	0	149 (105)	0 (0)	0 (0)	76 (156)	225 (262)	66 (40)
	Institutional	0 (0)	0	0 (0)	59 (38)	2 (0)	51 (87)	111 (125)	53 (31)
	Industrial	1 (0)	0	3 (0)	0 (0)	129 (74)	117 (175)	249 (249)	52 (30)
	Undeveloped	292 (419)	67	174 (226)	59 (64)	85 (146)	15277 (15020)	15952 (15890)	96 (95)
	Column Total	792 (967)	166	325 (331)	117 (117)	215 (220)	15870 (15842)	17486 (17478)	
	Producer's Accuracy (%)	63 (57)	56	46 (32)	50 (33)	60 (35)	96 (95)		93 (90)

The table shows that overall accuracy (the circled values) for AIIA and the baseline model is 93% and 90% respectively, suggesting that AIIA performs better spatially. In fact, the overall accuracy for both the models has been inflated by a proportionally huge amount of undeveloped land within the study area. If the amount of undeveloped land (or the developable area) that matches for both the predicted and reference map is excluded from error matrix or put at zero, the overall accuracy for AIIA and baseline models would fall to 43% and 31% respectively. A look at the indices of individual categories reveals a superior performance of the AIIA over the baseline model. Although the baseline model has commission and omission errors for residential land use only a bit higher than that of the AIIA, it fares poorly for other categories. This is because

the state of parcels in these categories in baseline model is updated randomly without having selected based on the suitability scores. This is an important difference between the two models. Totals of predicted accuracy, commission error and omission error for all categories are visualized for the output of the AIIA model as an error surface map (Figure 2.9). The area totals with corresponding percentages for both the models are given in Table 2-3.

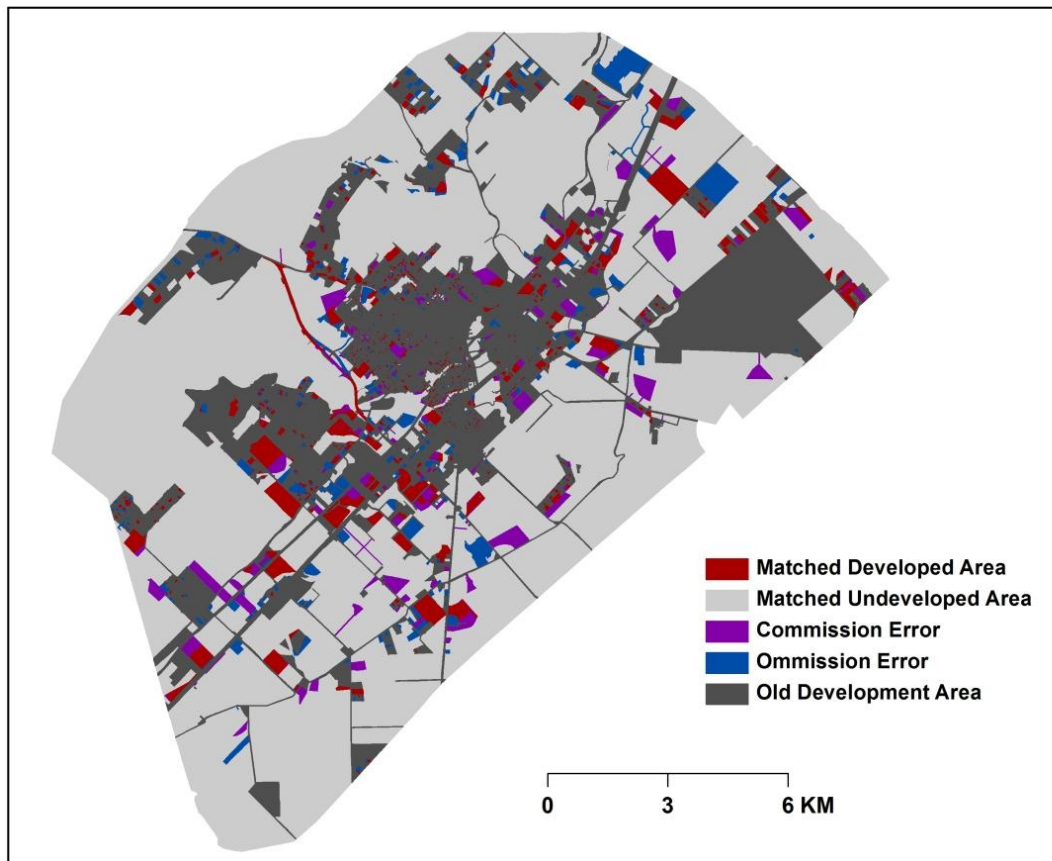


Figure 2.9. Visualization of prediction accuracy as an error surface highlighting commission and omission errors of the AIIA.

Table 2-3. Prediction accuracy and errors for the results of AIIA and baseline model.

	AIIA model (area in hectare)	Baseline model (area in hectare)
Predicted accuracy	926(43%)	763(31%)
Omission error	690(30%)	824(34%)
Commission error	609(27%)	872(35%)

The matrix-derived indices commonly used to assess classification accuracy of satellite images often fail to truly reflect the predictive accuracy of models that simulate complex systems such as cities (White et al. 2000). Even slight displacements result in classification errors, which in fact are considered accurate from a modeler's perspective (Vliet et al. 2009). This further worsens if the model employs vector data format. For example, if a model predicts a parcel to be developed at time $t+1$ that does not exactly overlap but just falls next to a developed parcel in observed data for the same time step, the prediction is quite accurate. With vector data, it is possible to evaluate accuracy of prediction in terms of how closely the developed area in a simulated map matches the developed area in the reference map. Visualization of the spatial match as a gradual error/accuracy surface map is an effective technique of error assessment. One way to visualize the errors is to create an agreement probability surface map as illustrated by Kocabas and Dragicevic (2009). Another way is to create a surface map with categorical data in terms of agreement classes, where each and every developed site in the predicted map is assigned to a class to indicate its degree of spatial agreement with the developed site in the reference map.

In this study, I categorized the predicted sites into four groups: perfect match,

very close match, close match, and poor match (Figure 2.10). The perfect match category includes developed parcels for each category in the predicted map that spatially agree with the developed area of corresponding category in the reference map. Very close match indicates the newly developed parcels in the predicted map that are adjacent to the developed parcels in observed map for each land use class. Close match category includes developed parcels in the predicted map that are not adjacent but within a distance of 300 meters from the developed parcel of corresponding land use in reference map. Any predicted sites beyond this distance from the developed parcel in reference map are classed as poor match. The total area for each category in both models is presented in Figure 2.11. The percentages indicate the proportion to the total developed area during the simulation period. More than 90% of the newly developed area in the simulation falls within a distance of 300 meters from the developed sites in reference data (i.e. with only less than 10% falling in the ‘poor match’ category). Thus, the model performs satisfactorily despite the fact that it has comparatively smaller overall, producer’s and user’s accuracies.

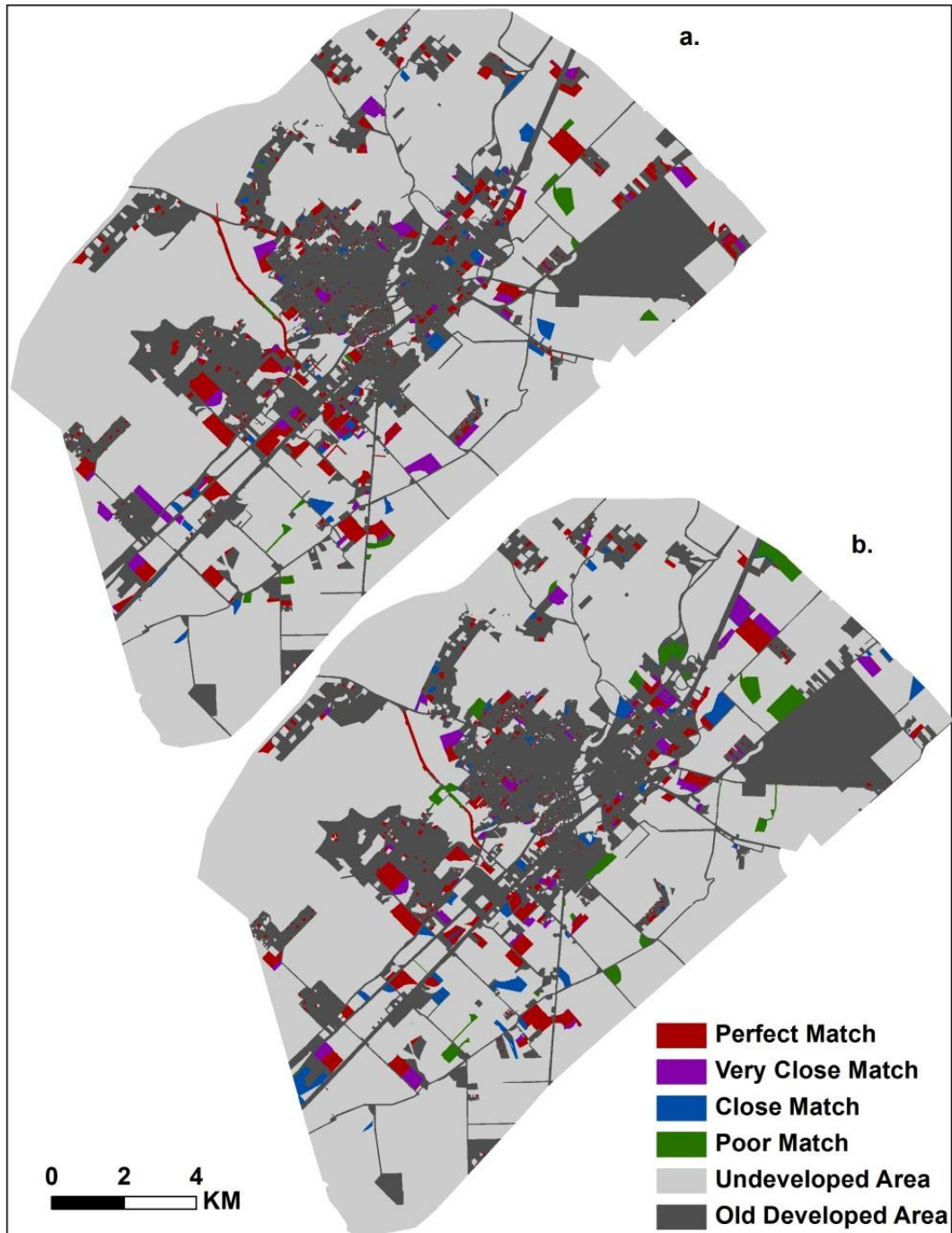


Figure 2.10. Visualization of spatial agreement of developed areas for all land use categories for the output of a) AIIA model, and b) baseline model.

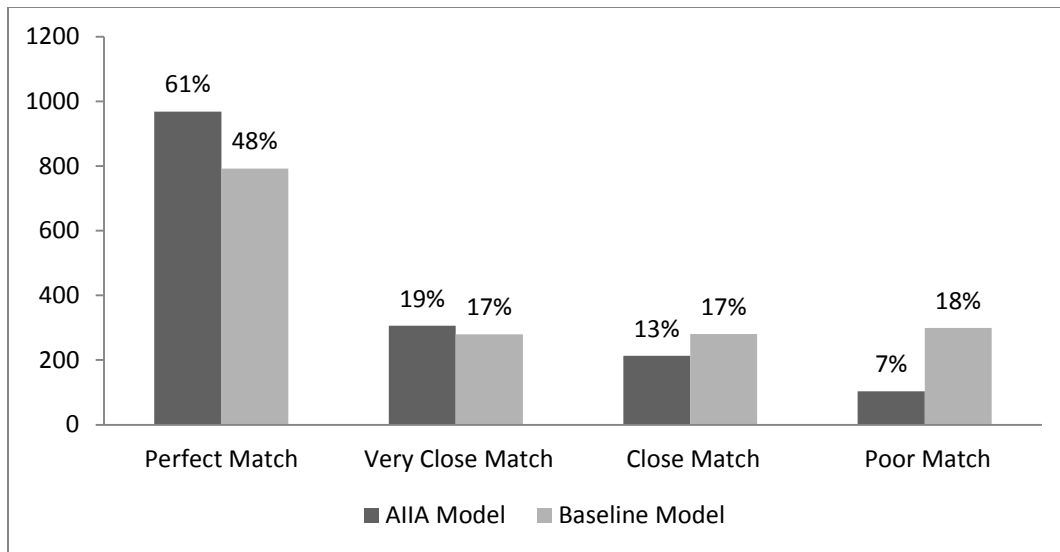


Figure 2.11. Comparison of the model results in terms of match classes. The y-axis represents newly developed area in hectare.

Two scenarios of future growth were simulated for the year 2020. A baseline scenario, which lets the development continue as it is with the current trend of growth, was simulated. This gives continuity to the spread of urban sprawl. In another scenario, compact development was encouraged by setting an urban growth boundary. Every parameter was kept the same as that of the baseline scenario except that the development was forced only around the city center. Development was allowed only within a distance of 1 mile from IH 35 on both sides and a radius of 2 miles centering on the TSU main campus. Results of the scenario simulation are presented in Figure 2.6e-f and Figure 2.12. In order to assess the results, four spatial metrics of shape index (SI), Moran's Global I, number of patches, and mean patch size (MPS) were computed (Table 2-4). The lower values of MPS, SI and Moran's I but higher number of patches for the results of baseline scenario suggest a less compact and dispersed development of the city. This verifies the fact that uncontrolled urban sprawl can be contained by the implementation of strategies

such as growth control boundary.

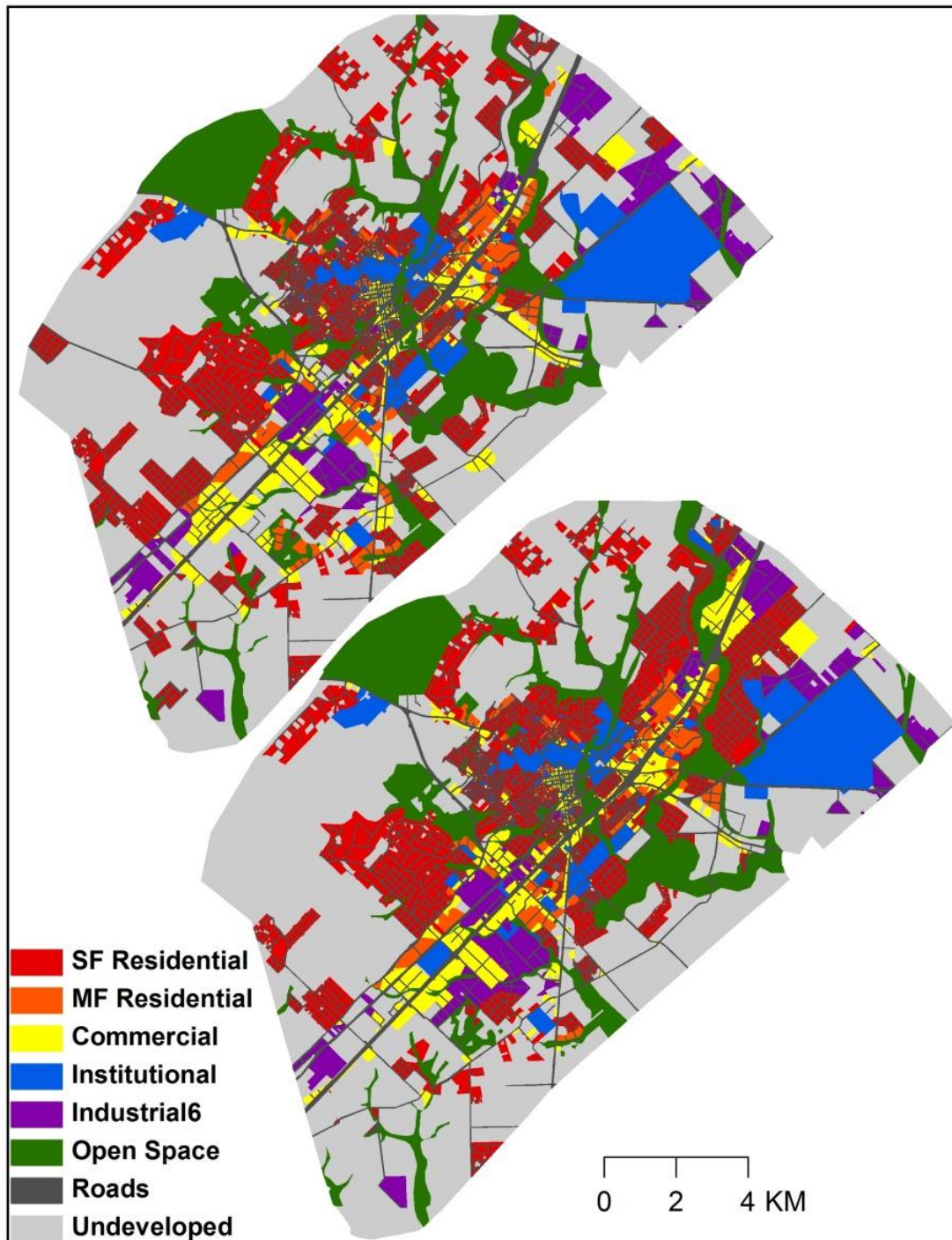


Figure 2.12. Future growth scenarios for 2020 a) current growth scenario (upper map), and b) compact growth scenario (lower map).

Table 2-4. Spatial metrics for the comparison of scenario results.

Scenario	Number of Patches	Mean Patch Size	Shape Index	Moran's I
Baseline	392	146.96	568.85	0.27
Compact Development	259	195.30	685.53	0.45

With the categorization of households into four types (subsection 2.4.1.3), AIIA is able to simulate the locational choices of households of various socioeconomic statuses at micro level. Clustering patterns and distribution of the households can be explored. Figure 2.13 shows that the multifamily units (i.e. apartments) are dominantly occupied by low income households without children, whereas the countryside spacious lots are resided by high-income households. Among the total 20,540 households in 2010, 22% were high income with children, 10% low income with children, 20% high income without children, and 48% low income without children. Furthermore, the model estimates the number and ratio of the four household types for every MF lot. For example, a newly developed apartment complex (MF lot) with an area of 11acres was occupied by 3 low income households with children, 9 high income with children, 20 high income without children, and 51 low income without children households.

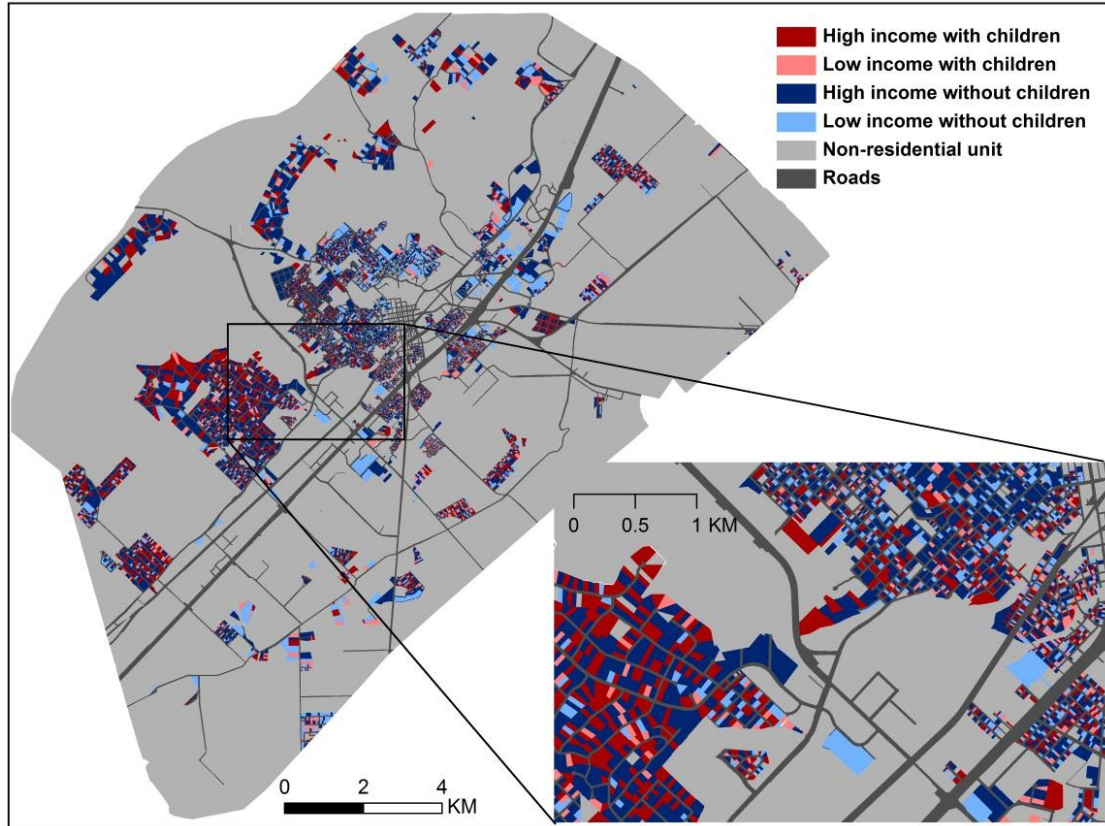


Figure 2.13. Map depicting spatial distribution of the four types of households in the study area for AIIA output.

2.6 Discussion and conclusion

This article outlines the development and implementation of the proposed AIIA model of urban land use dynamics. The AIIA model integrates agents' decision processes into irregular automata that are represented by cadastral land parcels. By including the site selection behaviors also of commercial, industrial and public/institutional agents, AIIA provides a holistic framework capable of capturing the processes of urban growth. The model's efficiency is corroborated by the higher values of overall as well as producer's and user's accuracies for the results of AIIA as opposed to the results of a baseline model that is on par with the existing irregular urban growth modeling. Table 2-2 shows that user's and producer's accuracies for commercial, industrial and institutional

categories are distinctly higher and hence the overall accuracy as well. This fact is also evident in Figure 2.11, which shows that the area in perfect match and close match categories are higher for the AIIA model. This answers the first research question of this project (i.e. RQ1) that incorporating the simulation of commercial, industrial and institutional agents' behaviors enhances the model accuracy.

Change in urban land use occurs as a result of locational decisions made by the agents based on preferences or utilities, which are derived from empirical data in contrast to the existing practice of employing heuristic rules to govern the decision of the agents. This data-driven procedure avoids subjective biases inherent in the behavioral rules, thereby enhancing applicability of the model to other cities as well.

As an improvement to the existing irregular automata-based modeling, AIIA includes classification of housing developer agent into two categories: single-family residential and multi-family residential, and household agents into four types based on socio-demographic backgrounds. This has increased the accuracy of the modeling. Table 2-2 reveals that both producer's and user's accuracies for the residential category are higher for the AIIA model despite the fact that both baseline model and the AIIA use utility functions to simulate the residential agents' behaviors. This clearly addresses the second research question (i.e. RQ2) that categorization of the household agents into four categories substantially enhances the robustness of the model. Furthermore, simulation of housing developer as two types results in more realistic results. For example, the baseline model divides the MF housing units also into smaller lots as opposed to the fact that a MF residential parcel consists of just one areal unit with multiple household units. Modeling households as units of various socio-demographic backgrounds yields

important findings about the micro-dynamics of urban processes such as the distribution of low-income versus high-income households across the city. For example, the AIIA results (e.g. Figure 2.13) show that the majority of low-income households without children through apartment residences.

The results of scenario simulation (Figure 2.12 and Table 2-4) address the third and the fourth research questions (i.e. RQ3 and RQ4) about impacts of urban growth policies and growth patterns in the study area. Urban growth policies such as growth boundary can have distinct impacts in the resultant pattern of urban development. The current trend without any control measures result in sprawled development while the growth boundary results in a compact development. Similarly, the results reveal that San Marcos has witnessed a proportional increase in commercial development and a decrease in industrial development during the simulation period, which is in par with the fact that mid-sized US cities are transforming to service economy for the last two decades (Kaplan et al. 2008). Although the study showed simulation for only two scenarios, the AIIA model can be used to simulate various other development scenarios. Thus, the model serves as a useful instrument for testing different policy choices and scenarios.

Nevertheless, the proposed model has some limitations and can benefit from further improvement. Creation of a module capable of carrying out logistic regression and then passing the coefficients (weights) to the suitability calculation module would enhance automation of the software because the present software requires that users carry out regression analysis separately in SPSS or other statistical packages and then input the coefficients into the model manually. Although the model has been developed to be generic, it fits more appropriately to small or mid-sized US cities that have enforced

exclusionary zoning policies. Further improvement of the model would make it well applicable to a wider range of cities of the world. Issues related to missing influence factors, accuracy of the weights, errors inherent in the GIS data and model uncertainties should be addressed in the future research.

3. PARCEL-DIVIDER: A GIS TOOLSET FOR AUTOMATED PARTITIONING OF URBAN LANDS

ABSTRACT

Urban growth models operating at a finer spatial scale usually incorporate a subdivision module that carries out automated partitioning of the lands selected for future development. In this paper, I describe the development and implementation of *Parcel-Divider* – a GIS toolset for automated subdivision of land parcels. In addition to automating the process of generating urban layouts such as city blocks, streets and cadastral lots, the toolset is capable of extending roads to new subdivisions. Researchers can integrate the toolset into their modeling while planners can use it as a standalone application to visualize scenarios of infrastructure arrangements in growing areas of the city. Our tool-generated subdivision configurations closely match the subdivision styles observed in real-world cities when compared visually as well as statistically.

3.1 Introduction

The primary focus of Urban Growth Modeling (UGM) is to realistically simulate the processes of spatial expansion of cities and land use dynamics over time, apart from producing scenarios of future development (White et al. 1997; Li and Liu 2007).

Commonly employed UGM techniques include micro-simulation modeling (e.g. Waddell et al. 2010), cellular automata (e.g. Clarke et al. 1997) and agent-based modeling (e.g. Tian et al. 2011) along with integration of other systems such as fuzzy logic (e.g. Al-Ahmadi et al. 2009) and artificial neural network (e.g. Yeh and Li 2003). A typical UGM pipeline includes a) estimation of suitability score or change probability for each site, b) site selection for development in next time step based on suitability or other criteria, and c) state conversion of the selected site (e.g. from undeveloped to developed or from forest use to residential use) (White et al. 2000; Sante et al. 2010). Sometimes, the pipeline

consists of an additional component for the creation of urban layouts such as roads, city blocks and cadastral lots in the selected site(s), specifically if the model employs vector data to represent the landscape at finer spatial scale sufficient to resolve the cadastral details (Benenson and Torrens 2004; Venegas et al. 2009a).

As an iterative process, the delineation of urban layouts typically represents the final blueprint of urban development that constrains the modeled development allowed in that time step. This information loops back into the macro urban environment, and would affect the global parameters (e.g. demand for suitable residence) for UGM simulation in the next time step. Hence, the delineation of the urban layouts in the newly developed area is critical to the performance of UGM. The model should be capable of visualizing the proper placement of the layouts. To accomplish this task, UGM often incorporates a subdivision module that automates the partitioning of lands selected for future development (Jjumba and Dragecivic 2012), thereby enabling the model to update development states iteratively at different time steps of simulation. This is accomplished as the last component of the UGM chain mentioned above, which is the focus of the present study. In this paper, I describe the development and implementation of a land subdivision toolset that can be used either as a module to be integrated into UGM or as a stand-alone application. The aim was to provide easy-to-use tools capable of creating accurate spatial configurations of urban residential lands.

The process of land subdivision involves splitting up of a larger land tract into streets and smaller subspaces variously called parcels and lots. In general, it is carried out by developers and land owners based on field surveys, empirical designs and legal criteria after their subdivision plan is approved by planning officials (Boucher 1993).

Constraints such as minimum lot size and egress (i.e. access to road) to each lot are maintained as per the local legislature (Schmitz 2004). Computer-assisted design (CAD) and Geographic Information Systems (GIS) are primarily used to generate and modify the subdivided lots (Demetriou et al. 2012), which is typically labor intensive and time-consuming as it involves a good amount of digitizing with frequent manual interference. A few scholarly works have focused on automation of the process (e.g. Wikramasuriya et al. 2012), but the existing solutions are limited in reflecting the diversity of subdivision layouts observed in real world urban landscape with plausible results. In the present research, I have developed *Parcel-Divider*, a GIS toolset for automated subdivision of urban lands. It offers a variety of subdivision styles to be applied to parcels of different geometric attributes. The toolset contains seven GIS tools, each is designed to carry out a specific subdivision style. Upon receiving the input of land parcels in the form of vector polygons, each tool in *Parcel-Divider* creates urban layout of specific type.

3.2 Related work

Only a few scientific studies have dealt with the automated simulation of the process of parcel subdivision using vector data format. In an early work, Wakchaure (2001) designed a stand-alone GIS tool for generating a subdivision layout at a single parcel level for the purpose of build out analysis. Although the tool could partition a parcel into lots showing a possible pattern of development, it was not automated and the accuracy of the output was always doubtful. Stevans and Dragecivic (2007) mentioned as their future work of developing an algorithm for automatically constructing small land parcels so that they could integrate it into their vector-based *i-City* model of urban growth. Later, Jjumba and Dragecivic (2012) embedded a land subdivision module into

their newly developed *Agent iCity* – an agent-based model of urban growth. Although they stated that the module would first divide the larger land parcel into city blocks and then the blocks into cadastral lots, further details about the algorithm, implementation and results were not reported.

With an orientation to the paradigm of computer graphics, Vanegas et al. (2009a) and Vanegas et al. (2009b) came up with an automated urban layout generation module as part of an attempt to integrate geometrical modeling (i.e. visualization of urban spatial configurations) into behavioral modeling (i.e. urban simulation modeling). By employing the algorithm of recursive binary division, their solution is capable of generating streets, lots and lot contents such as buildings. The outputs are a plausible match to the real-world urban layouts. However, the generated lots are not always guaranteed an access to streets, and the solution performs better only for rectangular or square land parcels. It also does not account for various subdivision styles. The validation is solely based on visual comparison. In a recent sequel by the team, Vanegas et al. (2012) gave continuity to the integration of urban simulation modeling and procedural generation of urban geometries. They introduced a new algorithm to address the problem of block subdivision inherent in previous versions, and upgraded the software by incorporating an environment for interactive editing. Supported by both visual and statistical tests, their results show high similarity to the real-world urban spatial geometries. However, this comes at the cost of model complexity. Their software is less user-friendly, inaccessible to ordinary users, requires specific data structure of graph-based networks, and does not support vector GIS data model with which almost all the real-world cadastral parcel data are produced and distributed. Their solution can only be

loosely-coupled with GIS, a common platform for UGM.

Demetriou et al. (2012) designed a GIS-integrated software module called Land Parceling System (LandParcelS) as part of their land consolidation integrated support system (LACONISS) for land planning and decision making. *LandParcelS* automates the process of parcelization, generating a set of new parcels that represent an alternative plan for land reallocation. The system generates new parcels based on optimization of shape, size, land value, and road access. However, it is not capable of creating roads, and is inappropriate for subdividing parcels into city blocks and then city blocks into housing lots. Wickramasuriya et al. (2011) developed a GIS-based subdivision tool capable of generating urban subdivision layouts including both streets and lots. The tool, which carries out the partition of both rectangular and irregularly shaped parcels, aims to optimize the output by creating the highest number of lots and the lowest number of streets possible. Yet, the model does not offer options for different subdivision styles, and performs poor in terms of shape and size of resultant lots adjacent to the boundary of irregular parent parcel. The tool cannot extend the road network if the candidate parcel is disjoint from the existing roads.

Building on these previous efforts, the present study has developed a GIS toolset for automated subdivision of parcels. The solution offers a variety of subdivision tools to choose for applying to a developable parcel according to its shape, size and orientation. Computationally simple and operational in a familiar ArcGIS environment, it is easy-to-use with tools that have only a few parameters and are partitioned into smaller tasks. The solution is expected to generate streets, city blocks and lots that are highly similar to the ones found in real-world urban landscape. The research aimed to answer these specific

questions: a) What are the errors of the total number and size of simulated lots? How closely do the modeled subdivisions resemble the observed ones? In the following section, the tools contained in *Parcel-Divider* are described in terms of functioning and their outputs.

3.3 Parcel subdivision toolset

Parcel-Divider comes as an ArcGIS toolset. It is comprised of seven script tools along with the corresponding Python modules. Users with some programming know-how can run the modules in Python environments such as IDLE as a standalone application or import the modules and integrate them into other Python programs. On the other hand, the script tools can be run in the ArcGIS interface by any ordinary GIS user. The script tools can also be used in the programming environment by calling upon as part of ESRI's ArcPy package functions. For this, the user should import the custom toolbox containing the tools. A screenshot of the interface of ArcGIS 10.x with a dialog box of one of the script tools from *Parcel-Divider* is shown in Figure 3.1. It provides a glimpse of the software platform in which the tools are operated and results visualized. The toolset with its tools is shown in the red box on the right of the figure.

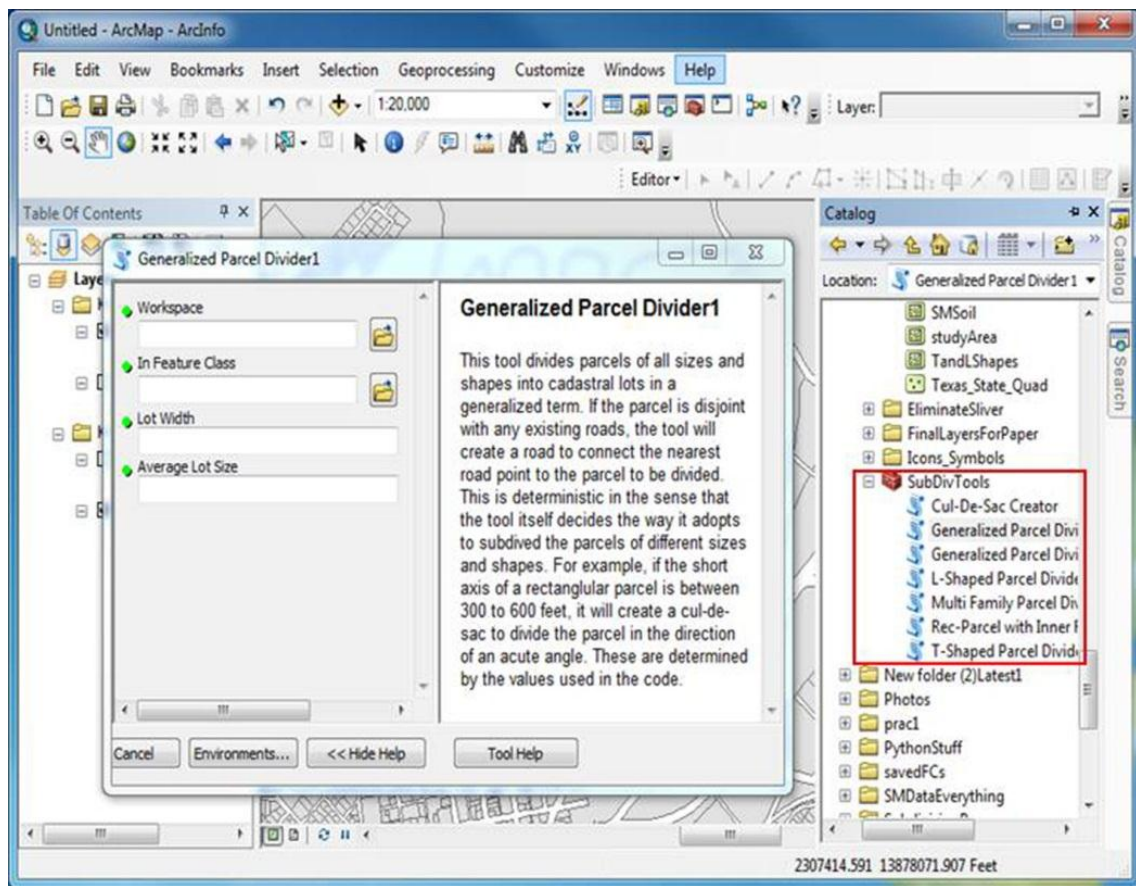


Figure 3.1. Interface of ESRI's ArcMap 10.x with a dialogue window of one of the script tools contained in Parcel-Divider.

The terms 'parcel' and 'lot' are often used synonymously despite their differing connotations in legal and taxation parlance. In this article, 'parcel' is used to describe a land tract to be subdivided into smaller polygon units called 'lots.' Thus, parcels are subject to subdivision whereas lots are undividable. However, if the parcel is of larger size, it is first subdivided into smaller areal units called 'blocks' which are surrounded by streets and consist of lots. The block size is a user-defined variable to the tools. So, any candidate parcel smaller than this size is directly split into lots. Likewise, although 'urban growth' is generally defined as spatial expansion of an existing city (Bhatta 2010), the phrase in the article is used in a bit broader sense to include any community development

regardless of the site's geographic location. In this sense, this term also includes the suburban and exurban development.

Parcel-Divider is intended to be used for generating urban geometries, a final step in the simulation of urban growth employing vector data. With input variables such as total population, annual growth rate, available lands and intended built-density, the simulation model estimates the number of incoming households and the amount of lands to be developed at each time step (Li and Liu 2007). The model then computes development or change probability for each site. This is usually accomplished by utilizing the demand estimations, various driving factors of urbanization, and transitional (or behavioral) rules (Batty 2005). The sites with higher development probability are selected for future development or change of their land use and land cover. The model also decides if a site is excluded from future development based on zoning information. Now, the toolset developed as part of this study is integrated to act upon the selected sites. For this, the tools are called as geo-processing functions and applied to subdivide the selected (i.e. candidate) parcels. In order to select an appropriate tool from the toolset, modeler can program simple rules such as “if selected parcel $P > 5$ acres, land use class == residential, and shape == irregular; then apply *Generalized Parcel Divider 1*”. Thus, all the selected parcels for a specific iteration of the simulation are subdivided by one or more of the tools depending upon the rules applied. However, the tools can be used as a stand-alone application as well. In this case, analysts are required to choose the parcel to be subdivided and the tool to be applied manually. Details on the tool parameters are provided in the documentation that accompanies each tool.

Subdivision development is a comprehensive process that engages a range of

individuals including planners, developers and field technicians. The location and design of a subdivision is driven by developers' motivation for profit and individual choices of homebuyers (Bowman et al. 2012). Decision about the shape and size of lots, width and tortuosity of streets, built density, and spaces for other amenities is made based on technical strategies often guided by heuristics and best practices, local subdivision ordinances, and homeowner preferences (Schmitz 2004). As a result, a wide variety of subdivision styles is observed in the real-world urban space. Moreover, each land parcel is unique in terms of its land use and land cover, terrain, shape, size, orientation, distance from road and other locational characteristics. Automating such a complex process using a single model is impractical. A single style or design does not fit parcels of all shapes, sizes and orientations. Different partition designs should be considered for different parcels in accordance with their unique features. *Parcel-Divider* offers options for seven different patterns of subdivision at this stage of software development. While there are different options for subdivision, all modules are designed to a) ensure egress to each lot, b) maintain shape-regularity of lots, c) reduce the number of streets while increasing the number of lots, d) create lots that closely match to the ones found in urban geographies, e) warrant continuation of existing streets into newly developed areas, and f) keep the automation computationally as simple as possible. In the following subsections, I present a detailed description of each of the subdivision tools contained in *Parcel-Divider*.

3.3.1 Generalized parcel divider 1

As the name suggested, this tool is designed to be generalized for parcels of all shapes, sizes and orientations by recursive subdivision. In addition to the parcel data, users are required to input values for lot width, lot length and average lot size. There

should be a specific field in the attribute table where users can indicate whether the feature (i.e. polygon) is a candidate for partition or not. It should be a string field type with the attribute value of 'Undiv' to indicate that the feature is undivided yet. When the tool is integrated in a simulation model, the field attribute is automatically updated for candidate parcels. But the users should update it manually if the tool is run on its own. Detailed and specific instructions for running the tool are provided in the tool documentation. Major rules and code structure used to develop the tool are given in the pseudo-code shown in [Figure 3.2].

```

Let P = candidate parcel, R = roads feature, S1 = average block
size, W = lot width, L = lot length, S2 = average lot size, MBR =
Minimum Bounding Rectangle, Min = short axis of the MBR, Max = long
axis of the MBR, and P' = divided parcel (but larger than S1), Max'
= long axis of MBR of P', and Min' = short axis of the MBR of P'
for P in Plist:
    if Pgeometry is disjoint to Rgeometry == False:
        pass
    else:
        create R connecting P to the nearest point of existing R
        create R around P
        take MBR of P
        if Psize < (S1 + S1*0.5):
            if Min < (L + L*0.5):
                divide the MBR into a single row of lots
                divide Max by W to get no. of lots
            else:
                divide the MBR into two rows of lots
                divide Min by 2 and Max by W to get no. of lots
        elif Psize >= (S1 + S1*5):
            undergo binary split of the MBR across Max and produce two P'
            create road along the dividing line
            while P'size < (S1 + S1*0.5):
                undergo binary division of MBR across Max' and produce two P'
                create road along the dividing line
            for all P' in P'list:
                divide the MBR of P' into two rows of lots
                divide Min' by 2 and Max' by W to get no. of lots
            clip the MBR with subdivisions with P as mask
            select lots < (S2 + S2*0.5) after clipping
            merge selected lots to their respective adjacent lots

```

Figure 3.2. Pseudo-code underlying the functioning of Generalized Parcel Divider 1

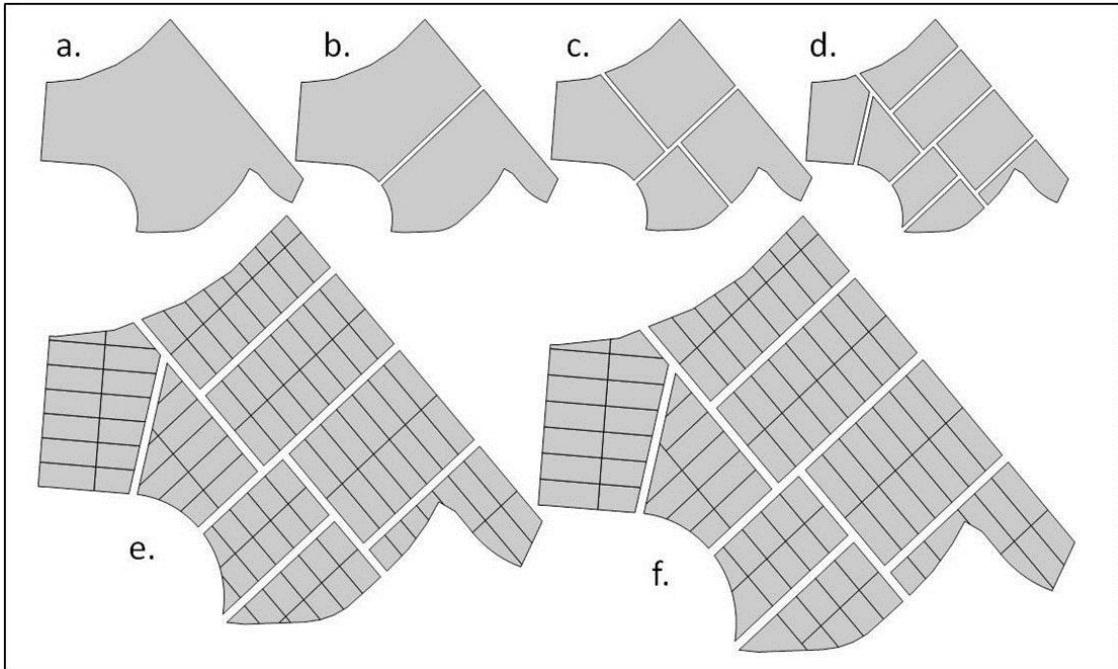


Figure 3.3. Visual illustration of the process of subdivision mentioned in the pseudo-code. Steps a)-d) show the recursive binary split operation upon the irregular candidate parcel, e) creation of lots, and f) elimination of undersized polygons.

As the first step, the tool figures out whether the candidate parcel's geometry touches the geometry of road features. If the parcel polygon is disjoint to the road, new road is created to connect it to the existing road. Then, the tool partitions the parcel into city blocks defined as enclosed areas surrounded by roads. If the parcel is substantially larger than the user-defined average block size, the tool applies a recursive algorithm to split the parcel into two halves using Minimum Bounding Rectangle (MBR) until all the divided parcels approximate the average block size. As shown in Figure 3.3a-d, this operation results in city blocks and roads within the parent parcel. However, the tool does not divide the blocks into lots recursively. For the block subdivision, long and short axes of the block's MBR are calculated. The long axis is divided by the user-assigned value of lot width and the short axis is divided by 2 which together determine how many lots the

particular block can have. Often times, the width of the lot remains equal to the user-assigned value while length of the newly produced lots may vary from block to block. Once the splitting of the parcel is completed, the tool clips the now subdivided MBR by the original candidate parcel as a mask. This creates two rows of lots for each block (Figure 3.3e). Then, it eliminates unacceptably small polygons and slivers, resultant of clipping, by merging them into adjacent lots as shown in Figure 3.3f. The users assign a size value as one of the tool parameters to decide which polygons are undersized. When integrated into a simulation model, the parameter can be inferred from a zoning data layer in order to be consistent with the land use categories. For example, the parameter value of undersized polygon can be larger for industrial parcels than for single family residential.

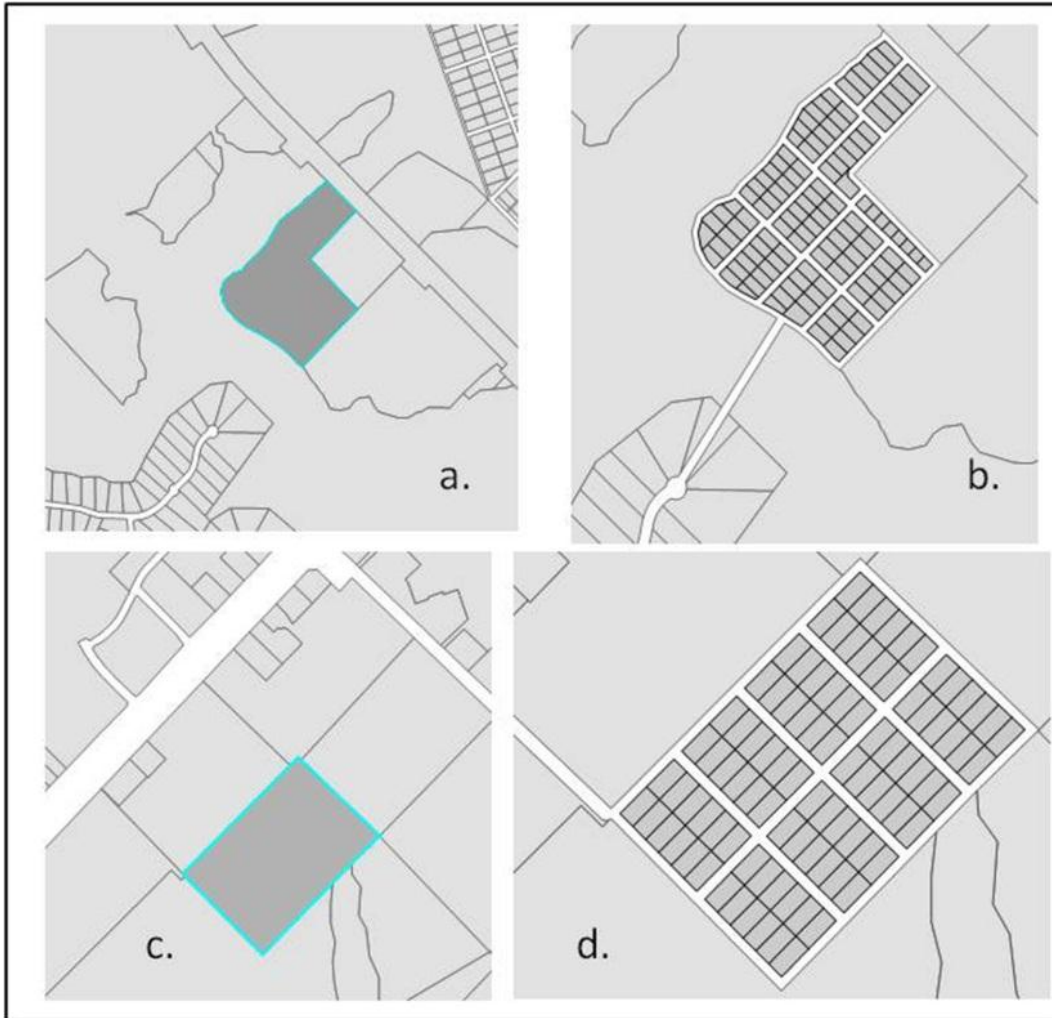


Figure 3.4. Outputs of Generalized parcel divider 1. a) Candidate parcel of irregular shape and disjoint from roads, b) extension of roads and partition of the irregular parcel, c) candidate parcel of regular shape, and d) the parcel after partitioning.

Thus, the produced lots are of same shape (rectangular) for a block not adjacent to parcel boundary [Figures 3.3e-f, 3.4]. But the size may slightly vary because the algorithm adjusts the length based on the block width. The lots adjacent to other parcels have irregular shape and their size varies considerably but within a limit. That is, some lots can be larger than the user-assigned average size if they are the result of merging smaller ones, and others can be smaller than the average if they are larger than the

unacceptable size to be eliminated. But in any case, they cannot be double the user-assigned average size and smaller than the user-assigned unacceptable size. Figure 3.4a shows an irregular parcel, which was subdivided by the tool into lots as shown in Figure 3.4b. The resultant lots are more uniform in terms of shape and size if the parent parcel is of regular shape (i.e. a rectangle or square in shape) as illustrated in figures 3.4c and 3.4d. The tool extends a road to the parcel that is aloof from the existing road by calculating the nearest point. The extended road is always straight and represents the shortest distance from the candidate parcel. The tool at the current stage of development lacks the capability of creating roads with sinuosity and other landscape attributes of real world routing. Although the continuation of road is highly plausible in Figure 3.4d, the road extends out from the dead-end of a cul-de-sac in Figure 3.4b, which is uncommon in the real-world landscape. The tool considers only the nearest road point from the parcel.

3.3.2 Generalized parcel divider 2

This is another version of the generalized parcel partitioning tool available in *Parcel-Divider*. It is similar to the *Generalized Parcel Divider 1* except that it does not divide the parcel recursively. For a given parcel, it designs all the streets, blocks and lot's shape and size in a single step. It calculates the MBR of the candidate parcel as its first step. Street lines are created along the right-hand side long axis and left-hand side short axis of the MBR if the polygon main angle is zero or more than zero (i.e. positive). The main angle is the angle of longest collection of polygon segments that have similar orientation (ESRI 2012). In other words, it is the angle of the polygon's main axis, which is taken to be the longer side of its MBR. On the other hand, street lines are created along the left-hand side long axis and right-hand side short axis of the MBR if the orientation of

the MBR is negative (i.e. less than zero). In each case, regardless of the orientation, the long axis and short axis of the MBR are divided by user-assigned values of lot width and lot length respectively. This determines the total number of lots to be generated for the parcel. For example, if we let A_1 = length of the longer axis of the polygon's MBR, A_2 = length of the shorter axis of the MBR, L_w = lot width, L_l = lot length and T_n = total number of lots; then $T_n = (A_1/L_w) * (A_2/L_l)$.

Intersecting streets that demarcate city blocks and run across the body of the MBR are also generated along both of its axes. Depending on the size of the MBR, a set of street lines is created that run parallel to the long axis. The origin points lie on the left-hand side of short axis of the MBR if MBR main angle is positive. The origin points are located incrementally based on cumulative distance from the corner point. Distance between two street lines running parallel is calculated to perfectly align the end edges of the lots. For example, if we let D = distance of street's origin from a corner point, then $D = L_l * 2$ for the first street line and $D = D + L_l * 2$ for every other street line until $D < A_2$. But, if $A_1 < L_l$, then no street running parallel to the long axis is created. As shown in Figure 3.5, the tool creates 5 streets running parallel to the long axis perfectly aligning to end edges of the lots. In this particular example shown in the figure, the first street is seen only partially in the middle because most of it fell out of the boundary of the parent parcel when the subdivided MBR was clipped. Similarly, the number of streets that run parallel to the short axis of the MBR is determined by dividing the long axis by a user-assigned integer. For example, a value of eight would generate a street every eighth lot, and the distance between the first street and the new street will be $8 * L_w$. In Figure 3.5, streets have been created after every fifth lot from the street running parallel to the short

axis. Thus, this tool always tends to generate block pattern with Manhattan-style street network.

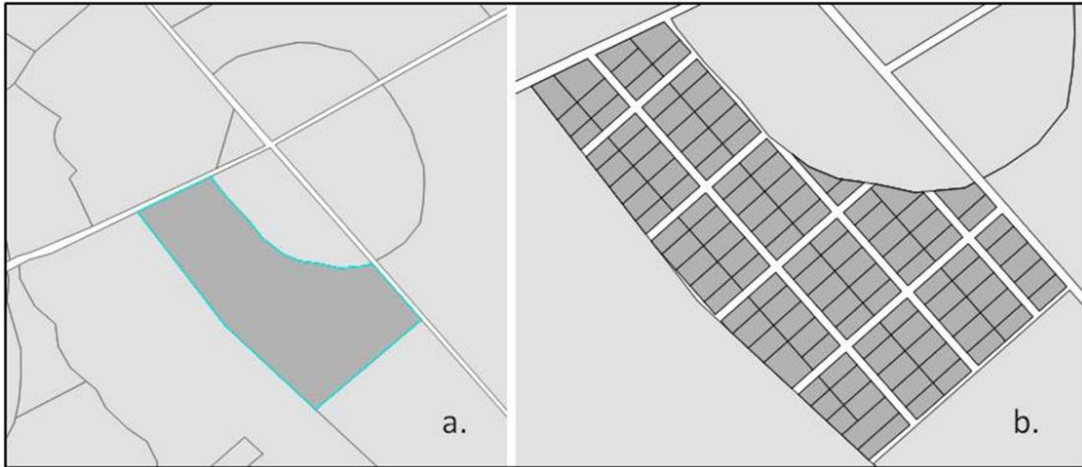


Figure 3.5. Subdivision results of Generalized parcel divider 2. a) a candidate parcel, and b) resultant lots (the darker ones).

As in the case of *Generalized Parcel Divider 1*, the partitioned MBR is clipped by the parent parcel. The resultant lots smaller than a user-assigned unacceptable size are merged with their respective adjacent lots. This tool performs faster than its counterpart as it does not have to go through iterations. If the parcel is comparatively larger and regularly shaped, this option can be a better choice. This tool also creates a road line to connect the parcel with existing roads if it was disjoint from the roads.

3.3.3 Divider with inner roads

Parcels of near-rectangular shape are observed to have a subdivision style that is defined by inner looped roads such as the one shown in Figures 3.6a and 3.6b. *Divider with Inner Roads* mimics this type of subdivision pattern. The tool fits better if the parent parcel is rectangular in shape and has the width approximately 5 times the lot width because it divides the rectangle into four rows of lots plus two roads. Like other tools in

Parcel-Divider, it takes an MBR of the parent parcel and operates subdivision before clipping the lots by original parcel, eliminating the undersized polygons, and then appending them back to the input Shapefile. The only user-input in this tool is lot width. Lot length is adjusted based on the width of the parcel's MBR. The steps below outline the basic operations in the process of subdivision using this tool.

- First, find a middle co-ordinate point on the short axis of the MBR, i.e. point E in Figure 3.6c, which is the midpoint of line PS. Midpoint of any two given points is calculated using the following simple trigonometric function.

$MP_{x,y} = (P1_x + P2_x)/2.0, (P1_y + P2_y)/2.0$, where $MP_{x,y}$ is the midpoint between two points P1 and P2. Labels P1 and P2 do not denote any specific points in the figure.

- Get Point A which is the midpoint of points E and P in the same way as above. This serves as one of the endpoints of the inner street.

- Get the angle of line PQ using the function $\theta = \text{atan} (P2_y - P1_y) / (P2_x - P1_x)$, where θ is the angle of the line defined by coordinate points P1 and P2.

- Find point B using point A as an origin point, angle θ and a distance of $L-W*2$ where L is the length of MBR and W is the lot width. The function to find a coordinate point given the origin point, an angle and a distance is stated below.

$P_{x,y} = (P1_x + \text{Dist}*\cos\theta), (P1_y + \text{Dist}*\sin\theta)$, where $P_{x,y}$ is the new point, P1 is the given coordinate point, Dist is the given distance and θ is the given angle.

- Find points C and D in the same way as above.
- Create a road of user specified width that runs through the points [A, B, C and D].
- Create line EF, and divide it by W. Then, generate lines of a number equal to the resultant quotient with a gap of distance W. These lines should extend

perpendicular to line EF on either side.

- In order to delineate the corner lot near R, compute the location of two points X and Y that have the distance W from the corner point R. Create two other points on either side of point C that have the distance of $W/6$. Then connect the lines as shown in the figure. The same should be done for corner point Q.
- To create the street section connecting the two inner streets, find midpoint N of points A and B. Then, locate point M using the same angle and distance of line AE. Find points L and K in the same way. Connect the four points to create the street section.

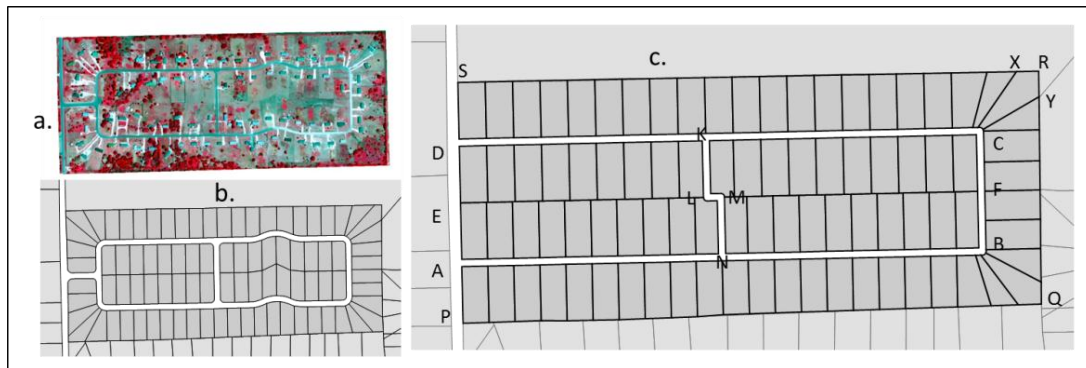


Figure 3.6. Subdivision of a parcel with street looping inside. a) Imagery of the subdivision, b) observed layout, and c) modeled layout.

The resultant subdivision of this tool is a near match to the layout configurations observed in real world urban space (City of San Marcos, Texas in Figure 3.6). The main differences seen are about the shape and orientation of the lots attached to the existing front road, and the main outlet. The observed layout has a single outlet while the modeled layout has two. The modeled streets lack the slight circularity or bending as displayed in the reference map [Figure 3.6b]. The trigonometric functions mentioned in this subsection to describe the tool's functioning are used for other tools also.

3.3.4 Cul-de-sac creator

Although a variety of cul-de-sacs exists across the world, one type commonly observed in the US suburbs extends into an elongated parcel with a straight street as shown in Figure 3.7. One end of the street is connected to the existing road while another end has a circular dead-end. *Cul-de-sac Creator* has been designed to create this type of cul-de-sac and residential lots on either side of the street. As illustrated in the figure, the tool offers four options based on whether the main angle of the parent parcel is negative, and whether the lower end (i.e. bottom short axis) is attached to the existing road. For example, Figure 3.7a shows subdivision of a parent parcel that had its upper end attached to the existing road and had a positive main angle, whereas parcel in Figure 3.7b had positive main angle with its lower short edge adjacent to existing road. On the other hand, Figure 3.7c shows a parcel with slightly negative main angle and the road on its upper short edge, whereas the parcel in Figure 3.7d had its lower end attached to the existing road with a negative main angle. For this tool to work, users will have to assign whether the existing road runs along the bottom short edge of the polygon or along the upper end although it identifies the main angle by itself.

Users assign the lot width, but the lot length is adjusted based on the length of the short axis of the parent parcel's MBR. The length of the short axis should not be more than three times the expected length of lots because it creates only two rows of lots one on either side of the street. The tool applies similar logic of coding and operational steps as the tools mentioned above. A visual inspection of the results reveals that the tool generates uniform lots in terms of size and shape.

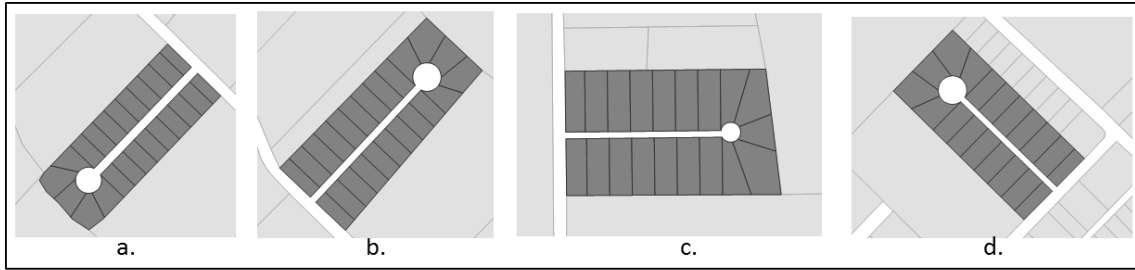


Figure 3.7. Four types of configuration of the subdivided parcel with cul-de-sacs based on orientation of the parcel and position of existing roads.

3.3.5 *L-shaped parcel divider*

Subdivision of an L-shaped parcel may be more challenging if the lots and streets have to be created only within the given parcel without encroaching to adjacent parcels. Inconsistent lots in terms of shape and size are likely to be created along with discontinued road sections if the parcel is subdivided using the logic of recursive splitting. *L-shaped Parcel Divider* offers two optimal ways of dividing a parcel that is closer to L- or opposite-L in its shape. In both the options, the goal is to maximize the number of lots with the minimum number of roads while guaranteeing egress to every lot. According to the first option, the tool creates a road that runs along the middle of the body of the candidate parcel [Figure 3.8a-b] to connect the two ends of the shape. Then, it divides the parcel into lots of a given width on either side of the road [Figure 3.8c-d]. In the other option, the tool creates a road along the outer boundary of the body of the parcel. Then, the parcel is split into two rows of lots as shown in Figure 3.8e-f. The users input the lot width while the lot length is adjusted based on the width of the leg of the L-shape. Likewise, the users have to choose one of the two options with the first being a default parameter. Operational steps and code structures are similar to that of *Cul-de-sac Creator* and other tools except that the present tool calculates convex hull instead of

MBR of the candidate parcel in order to estimate the location of different coordinate points such as the origin point, midpoint, and distance between two points.

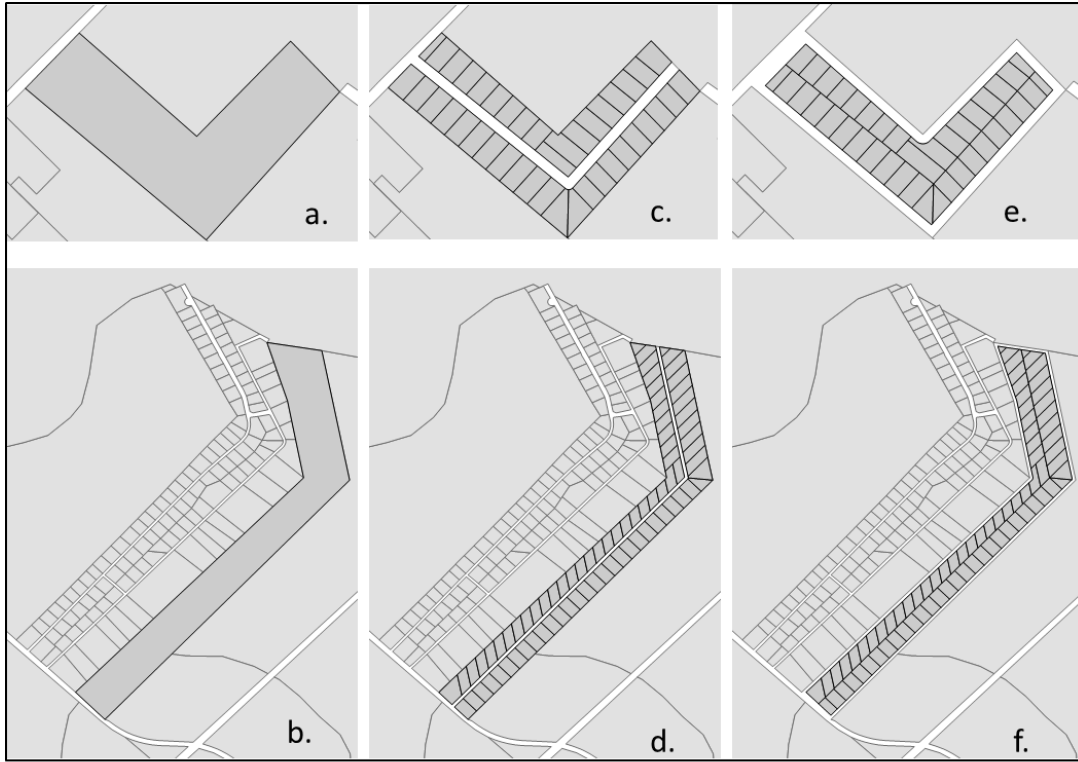


Figure 3.8. Subdivision layout generated for L- a) and opposite L-shaped b) parcels. c) and d) show the subdivision style in which street runs bisecting the parcel; e) and f) show the subdivision style in which roads surround the parcel

3.3.6 T-shaped parcel divider

Similar to the *L-shaped Parcel Divider*, this tool takes up convex hull of the T-shaped candidate parcel (Figure 3.9a) for creating roads and lots. It also offers two options: one with a road running through the body to connect three ends of the shape (Figure 3.9b), and another with roads running all along the boundary of the parcel [Figure 3.9c]. In the second option, it also creates a street line that cuts apart the leg of the T-shape to connect the boundary street. This ensures that each and every lot has an egress. Both the T-shaped and L-shaped parcel dividing tools from *Parcel-Divider* perform

better if the shape of the candidate parcel is closer to a perfect T or L shape.

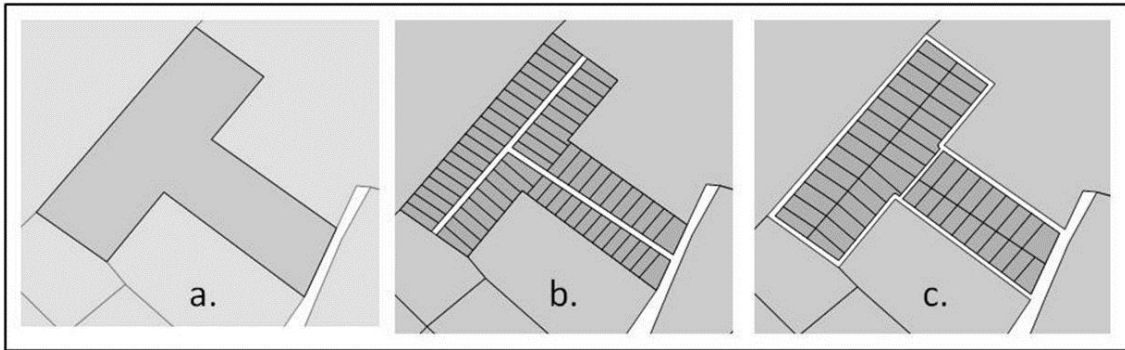


Figure 3.9. Subdivision results of T-shaped Parcel Divider: a) parent parcel with darker shade; b) and c) two subdivision styles offered by the tool.

3.3.7 Multi-family residential parcel divider

Parcels zoned for multi-family residential development are not subdivided in the same way as parcels of single-family residential land use. Multiple buildings with ample parking and service areas are constructed in a single tract of land without splitting it into smaller lots. A parcel or tract is subdivided only if it is comparatively larger and needs to be broken into parcels of smaller size but big enough to serve as the basic multi-family land unit. *Multi-family Residential Parcel Divider* is used for the purpose of dividing a large parcel into smaller ones rather than generating lots. As shown in Figure 3.10, the tool creates multiple multi-family residential units surrounded by roads. The tool mimics the *Generalized Parcel Divider 1*, except instead of generating two rows of rectangular lots within a city block, it splits the parcel into two halves recursively and creates a street along the dividing line until the size of child parcels reaches the user provided average multi-family parcel size.



Figure 3.10. Partitioning of a larger parcel into smaller multi-family residential lots

3.4 Results and validation

For any model or software to be practically useful, it is essential that the modeled output is comparable to the existing entities in the real world. The reliability of the proposed methodology is determined by conducting visual and statistical tests against certain criteria. Common criteria to evaluate the performance of a subdivision model include: a) generated objects, i.e. the ability to create streets, lots, and building; b) egress, i.e. the assurance that the generated lots have street access; c) plausible shape and size of lots, d) subdivision styles, i.e. if it offers choices for different subdivision styles and can create cul-de-sacs etc.; e) efficient use of available space, i.e. the ability to create the least possible number of streets at the same time ensuring egress to all lots; and f) model efficiency (Vanegas et al. 2012; Wikramasuriya et al. 2011).

This study also adopts these criteria to examine the performance of the proposed tools. Both visual and statistical comparisons were conducted to validate the modeled results against observed residential lots or the reference data. Figures 3.11 through 3.14 present four different subdivision styles each showing a) a parent parcel, b) results of my tools, c) cadastral parcel data obtained from the City of San Marcos, Texas for the corresponding site, and d) an orthophoto of the subdivision. Figure 3.11 depicts the action

of *Generalized Parcel Divider 1* upon an irregular parcel. The size and orientation of blocks vary between the modeled and observed subdivisions. The blocks in the observed scene are smaller with a variation in size and main angle whereas the modeled blocks are fewer in number, larger in size and uniform in orientation each having a positive main angle. One of the blocks in the observed scene comprises only of a single row of lots and a small triangular block with only three lots [Figure 3.11c]. This indicates that the tool can optimize the limited space for lot creation.



Figure 3.11. Comparison of the tool-generated subdivision of an irregular parcel. a) a candidate parcel, b) modeled lots and streets, c) observed layouts in vector data model, and d) an orthophoto of corresponding subdivision.

Figure 3.12 shows the results of *Divider with Inner Roads*, both the modeled and observed subdivisions having about the same number of lots of equal size. An apparent difference is the presence of a street section that connects two street sections running parallel to the long axis of the parcel in the modeled scene [Figure 3.12b]. Figures 3.13b and 3.14b depict the outputs of *Cul-de-sac Creator* that resemble the observed scenes [Figure 3.13c and 3.14c] in terms of number, size and shape. One noticeable difference based on the visual cue is that the observed lots adjacent to the existing road are larger and oriented to a different direction than the rest. Results of statistical tests are presented in Table 3-1.

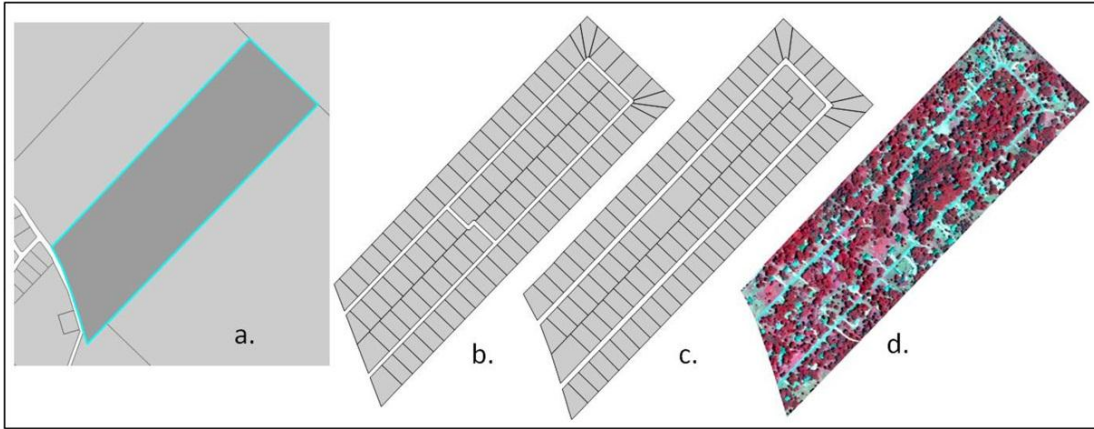


Figure 3.12. Validation of the simulation result. a) parent parcel, b) modeled lots and streets, c) observed lots and streets, and d) orthophoto of the subdivision.

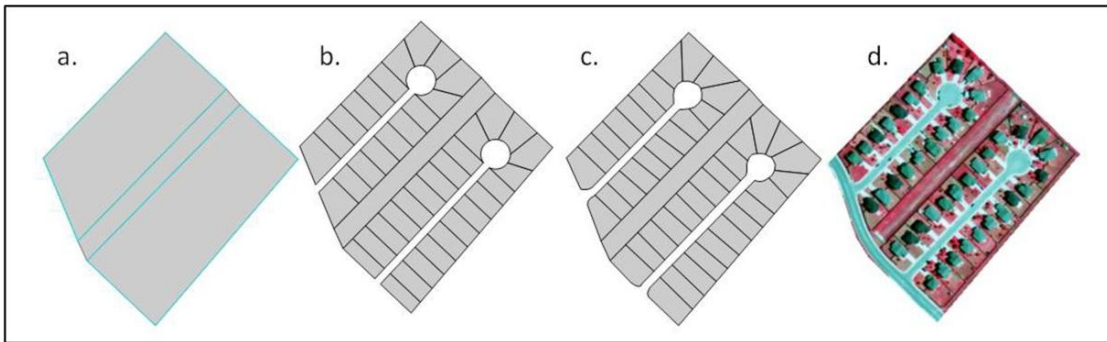


Figure 3.13. Output of Cul-de-sac Creator compared. a) parent parcel, b) modeled lots and streets, c) observed lots and streets, and d) orthophoto of the subdivision.

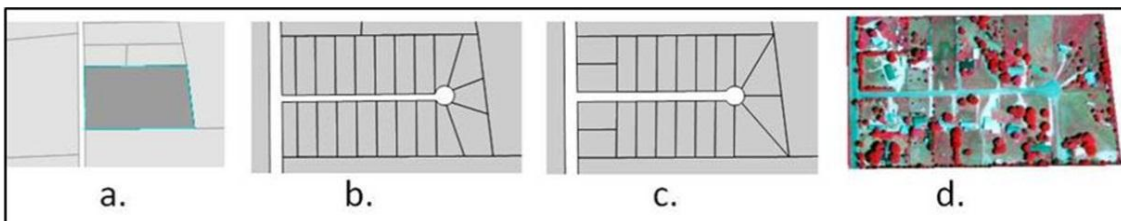


Figure 3.14. Modeled and observed lots compared. a) parent parcel, b) modeled lots and streets, c) observed lots and streets, and d) orthophoto of the subdivision.

As described in Section 3.3, the proposed tools can create both lots and streets apart from extending a new road to the parcels untouched by existing roads. Visual inspection is a general way of assessing whether egress is guaranteed to each lot. In the

study, the total number of lots that are disjoint from the street was counted [Table 3-1]. The number of lots having egress to multiple streets was also calculated for each subdivision. Although it is common to find lots that are adjacent to streets in more than one direction [e.g. Figure 3.11c], having egress to multiple streets is deemed to be a less efficient design because it violates the assumption of optimal use of available space. It increases unnecessary consumption of space by streets and may deem undesirable to commercial builders. Since the street width is a user-assigned variable, the total length of streets per subdivision was calculated instead of total area in order to see if the subdivision has redundantly allocated space for streets.

Although variation in the shape and size of lots within a block is common, lots produced by the tools tend to be relatively uniform in size and regular in shape. The mean and standard deviation (STD) of width, length and size of lots were calculated for each subdivision to assess the plausibility of lot size. For judging the shape uniformity, simple indicators called Shape Index (SI) and Regularity Index (RI) were calculated for each lot. Mean and standard deviation of these two indices are also reported for each subdivision. SI is estimated by dividing the lot perimeter by its area, and RI is calculated by dividing the lot area by the area of its bounding rectangle. An SI value of 1 or closer to 1 indicates that the parcel is a square in shape. This index shows how elongated a lot is. Similarly, an RI value of 1 indicates solidity of the lot, i. e., regularity of the shape. For example, square or rectangular lots tend to have an RI value of 1.

Table 3-1. Descriptive statistics of the tool results. For each descriptive column, results of modeled and observed subdivision are given, the observed ones being in the parenthesis. Error percentages are shown in curly brackets.

Fig No	Total lots	Mean lot size	STD lot size	Lots with out egress	Lots with multiple egress	Mean SI	STD SI	Mean RI	STD RI	Total road length	Mean Short Axis	STD Short Axis	Mean Long Axis	STD Long Axis
6	88 (89) {-1.12}	45379 (43101) {5.28}	4403 (4578)	0 (0)	8 (12)	0.020 (0.021)	0.002 (0.001)	0.96 (0.91)	0.113 (0.137)	8300 (8378)	150.63 (150.94)	12.46 (18.75)	321.43 (315.66)	22.76 (44.24)
11	194 (206) {-5.62}	9319 (8766) {4.98}	1728 (3201)	0 (0)	35 (58)	0.03 (0.05)	0.004 (0.008)	0.99 (0.93)	0.07 (0.08)	11054 (12201)	64.62 (62.92)	7.53 (22.41)	174.42 (147.49)	23.16 (19.21)
12	103 (105) {-1.90}	42348 (40518) {4.51}	5563 (6255)	0 (0)	8 (4)	0.021 (0.022)	0.002 (0.001)	0.96 (0.95)	0.11 (0.09)	8423 (7931)	157.80 (155.91)	22.00 (26.82)	282.95 (276.13)	22.47 (27.62)
13	41 (43) {-4.65}	8669 (8231) {5.32}	1263 (1315)	0 (0)	4 (4)	0.046 (0.047)	0.004 (0.004)	0.91 (0.87)	0.14 (0.13)	1015 (1094)	75.18 (74.35)	21.16 (18.91)	131.67 (131.40)	9.30 (12.27)
14	19 (20) {-4.99}	32666 (30490) {6.84}	4454 (4532)	0 (0)	2 (2)	0.025 (0.026)	0.002 (0.001)	0.91 (0.90)	0.15 (0.174)	849 (863)	126.13 (132.01)	25.92 (47.62)	294.05 (278.34)	28.19 (43.35)

Visual verification as well as statistical result reveals that all the lots generated by the tools are guaranteed an access to road. As shown in Table 3-1, the number of lots having multiple egresses is higher for the observed lots documenting the fact that the tools perform better in allocating minimum possible space for streets. This is also supported by the fact that the total length of modeled streets for each of the compared subdivisions is shorter than the length of observed counterparts. Only in Figure 3.12, the road length is longer by about 400 feet because of a street section the tool created to connect the two street sections running parallel to each other. The observed streets are wider than the simulated. This has contributed to larger size of the modeled lots. However, the street width is a user-defined value, and can be changed by the user or calibrated by real-world observation.

The total number of modeled lots in a parcel is the function of average lot size – a user-defined value. The smaller the lot size, the higher number of lots a parcel can have.

The lot size in turn is dependent on its width. When the same average lot width is used, the total number of tool-generated lots was almost equal to the total number of observed lots for a given subdivision. As shown in table 3-1, the tools tended to create slightly higher number of lots than the observed ones as the tools are designed to use space efficiently for lot creation. Each subdivision has slightly lower number of modeled lots but a slightly larger average lot size. Error percentages of total lots and mean lot size, two important parameters in determining the effectiveness of the proposed tools, were calculated. The lower values as shown in the curly brackets imply a good match between the modeled and observed subdivisions. Error Percentage was calculated using the formula: $\text{Error \%} = (Z_{\text{modeled}} - Z_{\text{reference}}) / Z_{\text{reference}} \times 100$, where Z represents a cell value for each column.

Width for both the modeled and reference lots for a given parcel is equal whereas the length varies conspicuously. The reason for this is that the algorithm of the tools adjusts the lot length by dividing the width of a block into two. The width or short axis of the blocks varies during the process of binary split of a parent parcel. The “Mean Major Axis” column of Table 3-1 shows that the modeled lots have slightly longer major axes making them a bit more elongated. This is supported also by the lower values of SI, which increases for the lots that are created in narrower blocks. The mean RI also is slightly higher for modeled subdivisions indicating that the lots are a bit more rectangular or square (i.e. higher regularity of shape). Similarly, standard deviation of the size is slightly smaller for modeled lots, implying that the lot size is uniform. This relatively higher uniformity of size and regularity of shape for the simulated lots may have been caused by the fact that real-world lots are created based on decisions made out of field

observations and often influenced by individual needs and economic abilities of homebuyers (Kone 2006) whereas lots are designed in the model by using computer coded rules maintain uniformity. Furthermore, the higher variability in lots' shape and size owes to topographical features such as small mounds and gradient land surfaces that make uniform lot sizes and shapes impractical (Boucher 1993).

3.5 Discussion and conclusion

Parcel-Divider is a vector-based urban land partitioning toolset that offers multiple subdivision styles to match the geometric and locational attributes of a land parcel selected for development. The toolset contains seven tools to generate various lot and street configurations in the developable parcel. It is distributed as a customized interface to be operated in ESRI's ArcGIS suite. The Python modules contained in the toolset can be run in Python IDE such as IDLE and PythonWin. However, it is noted that the proposed toolset uses modules from ArcPy site package, and therefore users are required to have a valid license of the ArcGIS 10.x software. The tools are executed and the results displayed in ArcMap.

The two major research questions of this project (i.e. RQ 5 and RQ 6) are well addressed by results of the project. Answer for the first question "What are the errors of the total number and size of simulated lots?" is that the errors range between 0-6%, which means the tools produce lots with an accuracy of more than 94%. Another research question "How closely do the modeled subdivisions resemble the observed subdivisions?" evokes the answer that the resemblance between observed and modeled subdivisions is higher. It is documented by the results that the total number of lots is almost same for a particular subdivision and each lot in the subdivision has an egress.

The fact that SI and RI values for both the observed and modeled lots are almost same also indicates that the resemblance is higher. Moreover, this is better corroborated through visual comparisons in Figures 3.11 – 3.14.

In general, users can run the tool by accepting or modifying the default parameters to meet their applications. Since the source code and software logic is also accessible to them at the time of software distribution, they can create new script tools by modifying the code or adding parameters into the tools to meet their specific needs. But this will require some programming knowledge on their part. The current system can be tightly coupled with *Agent-Analyst*, an ArcGIS extension for agent-based modeling and simulation (Johnston 2013). *Agent-Analyst* provides a framework for behavioral simulation but not for geometric modeling. That is, one can build a simulation model of urban land use change, for example, using vector GIS data to depict what changes would occur on the landscape across time, but cannot subdivide or modify the geometry of the land parcels where changes occurred. *Parcel-Divider* can be integrated to accomplish the later task. Similarly, the results of the system can be complemented with *CityEngine*, a software facility for procedural modeling of 3D urban environments (ESRI 2013). As it is an ArcGIS portfolio for interactive editing with a full drag-and-drop support, the results of *Parcel-Divider* can be imported into it for further editing and modification. Moreover, landscape contents such as buildings, trees, canals and bridges can be added into the subdivided lands to produce a 3D model of the newly developed areas.

The toolset can be useful for researchers as they can integrate it as a subdivision module into their urban simulation modeling. For example, Figure 3.15 shows the result of the current system in conjunction with an irregular cellular automata model of urban

growth when applied to the case of a mid-sized U.S. city (Chapter 2). A close-up view of a small section of the simulated landscape makes it obvious that undeveloped parcels of different shape, size and orientation within different land use zones (Figure 3.15a) have been developed at time $t+1$ producing different subdivision styles (Figure 3.15b). Inclusion of such a subdivision mechanism in the simulation enhances the realism of the results in addition to enabling the model to take into account the bottom-up feedbacks from the micro-level components, which ultimately determines the macro-level decisions. For example, the behaviors of individual household, which is represented by a subdivided lot in the model, influence the decisions of housing developers about new development in next iteration.

Likewise, the toolset can be used by land developers and planning officials to assess and monitor different designs of urban layouts in the developable lands. It can be useful while conducting build-out analysis to estimate residential density and allocating land areas for roads and open spaces. Surveying technicians can utilize the tools for initial warm-up work which can give them guidance to conducting their field works. As part of an exploratory exercise, they can apply the tool to the parcel they are going to partition, producing different scenarios for varying parameter values. This may provide insights on different scenarios of parcel subdivision.

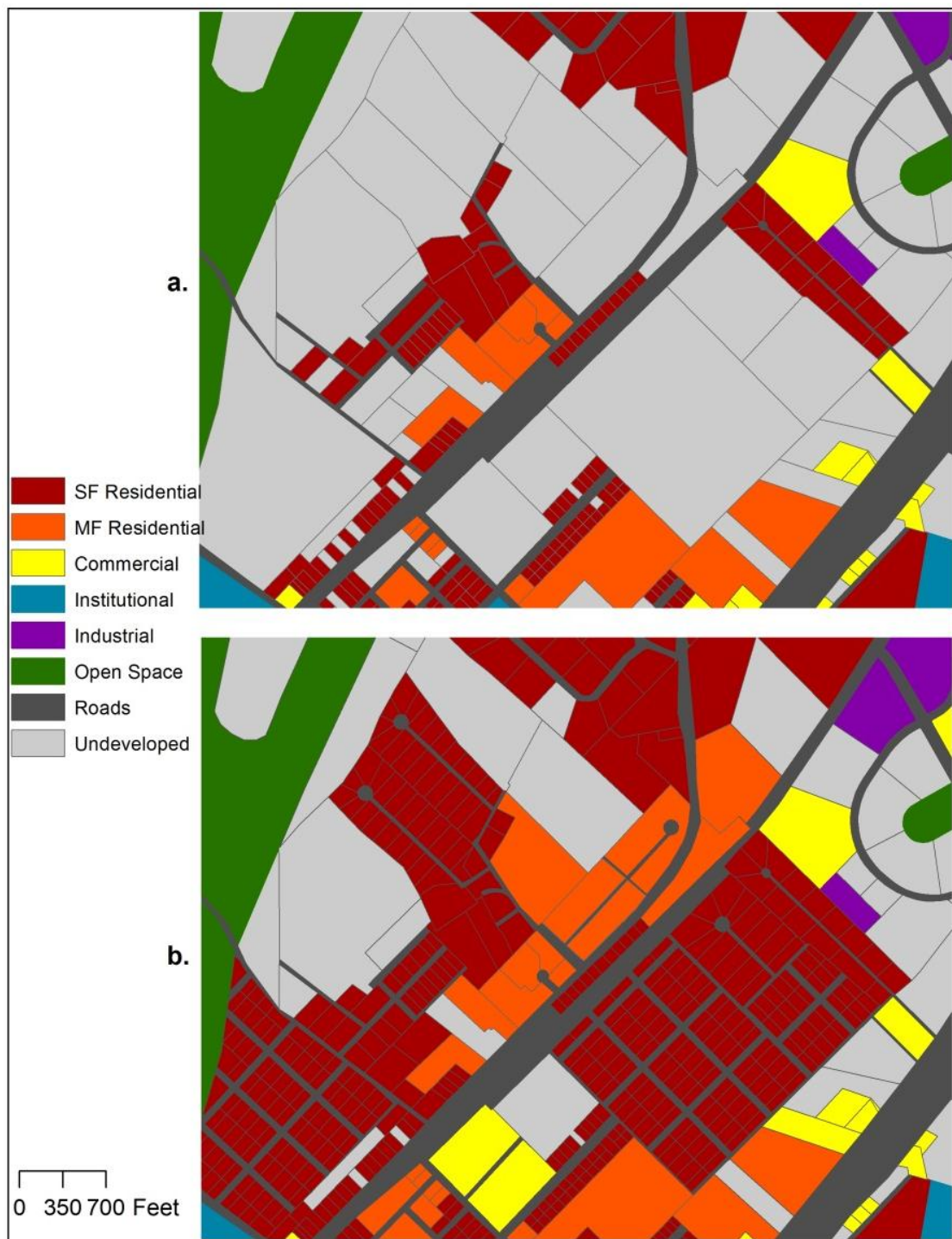


Figure 3.15. Results of Parcel-Divider tools in conjunction with an irregular cellular automata model of urban growth: a) a section of the City of San Marcos, Texas in 2000, and b) simulated urban land use changes after 10 years.

The present study builds upon the existing work on automated partitioning of urban lands, and further extends the methodology for creating urban layouts using vector data model. Specifically, it extends the previous work by Wickramasuriya et al. (2011). The proposed GIS solution automates parcel subdivision by utilizing a similar concept of using minimum bounding geometry of a parcel for partitioning. However, there are a number of additional features and capabilities in the toolset that extends previous limitations. It offers various choices of subdivision styles to fit parcels of different shape, size and orientation. Instead of using MBR solely for subdividing, the *Parcel-Divider* use convex hull as well considering the shape and orientation. The proposed toolset also provides capabilities to efficiently generate lots with cul-de-sacs apart from carrying out optimal subdivision of T- and L-shaped parcels, which pose a comparative challenge in terms of applying the logic of binary subdivision. The tools are integrated with a mechanism to merge the under-sized lots and slivers into adjacent lots – something that was mentioned as a future work in the previous work. The new tools are capable of creating a new road to connect the parcel that was disjoint to the existing roads. Moreover, the overall accuracy of the simulated output is more satisfactory in terms of visual comparison and statistical tests.

Parcel-Divider generates subdivisions of comparable realism. The solution passes the tests of maintaining egress to each generated lot, ensuring the optimum utilization of available space, and creating lots of uniform size and regular shape. In addition to residential parcels, the tools can be applied for the partitioning of lands zoned for commercial and industrial land uses. Specifically, the *Multi-family Residential Parcel Divider* tool can be run for the purpose in the ‘as-is’ form. Alternatively, the tool can be

modified and integrated with other tools of *Parcel-Divider* through minor changes of the code.

Nevertheless, the proposed toolset has some limitations and can be benefitted from further enhancement. *Parcel-Divider* still falls short on generating the variety of subdivision layouts exhibited in the real-world urban landscape. For instance, there is no tool for appropriately subdividing an upside-down T-shaped parcel. The tools cannot create the layout having sinuous streets with multiple cul-de-sacs. This warrants the inclusion of additional tools in the toolset in its future version. Appropriate sinuosity to the newly created roads should be introduced in the future work because new roads created by the current tools are always straight.

Results from a cost-path analysis can be incorporated in order to guide the path the newly created roads can follow. The higher regularity of lots and reduced road egress produced by our solution may not be desirable under all real world circumstances. With the emerging development approaches including low-impact development (Coffman 2000), conservation subdivision design (Arendt 2004), and new urbanism (Fulton 1996; Downs 2005), different subdivision designs are being practiced. Residential lots of uniform configurations and same subdivision type are not appealing from both conservation and aesthetic perspectives. More varied and organically grown neighborhoods with mixed land use units and street networks efficient for both intra- and inter-subdivision accessibility are encouraged (Bothwell et al. 1998; Duany et al. 2001). In order to be able to mimic these diverse designs, future upgrade of the toolset should include not just street sinuosity but also the variability in lot sizes, shapes and orientations.

Similarly, terrain attributes like slope should be taken into account in the future improvement as the current algorithm does not consider these factors, which in totality result in the variability of lot configurations as well as subdivision designs. Moreover, introducing appropriate shape recognition algorithm could enhance the tools' robustness. The tools recognize the size and orientation but not the shape of the parent parcel at current stage. Users should decide which of the tools to apply for a particular parcel. Integration of a module capable of detecting the parcel's shape and the side to which the existing road is adjacent could make it possible to automatically decide which of the subdivision tools to be applied for that particular parcel. While the toolset eases the procedure of parcel subdivision that is rife with labor-intensive editing, it is intended not to replace but assist in making land development decisions. The layouts generated by the current solution does not substitute the ones produced by surveyors based on field observations. The existing tools work with parcel data in ESRI Shapefile format, but future prototypes can easily expand to accept other file formats (e.g. coverage and geodatabase).

4. CHARACTERIZATION OF NEIGHBORHOOD SENSITIVITY IN THE IRREGULAR CELLULAR AUTOMATA MODEL OF URBAN GROWTH

ABSTRACT

Neighborhood definition, which determines the influence on a focal cell from its nearby cells within a localized region, plays a critical role in the performance of a cellular automata (CA) model. Raster CA models use cellular grid to represent geographic space, and are sensitive to the cell size and neighborhood configuration. However, the sensitivity of vector-based CA, an alternative to the raster counterpart, to neighborhood type and size remains uninvestigated. The present article reports the results of a detailed sensitivity analysis of an irregular CA model. The model was developed employing parcel data at the cadastral scale to represent geographic space, and implemented to simulate urban growth in central Texas, USA. Twenty-six neighborhood configurations defined by types and sizes were applied to the model in order to examine the variability in the model outcome. Results from multiple accuracy assessments and landscape metrics confirmed the model sensitivity to neighborhood configurations. Furthermore, the centroid intercepted neighborhood with a buffer of 120 meters produced the most accurate simulation result. This neighborhood produced scattered development while the centroid extent-wide neighborhood resulted in a clustered development predominantly near the city center. The adjacency neighborhood, another type defined, promoted leapfrog development.

4.1 Introduction

In recent decades, the spatial simulation technique of Cellular Automata (CA) has been widely used to predict urban growth and land use/land cover change. CA are generally defined as a multi-dimensional grid of identical cells that are self-acting and capable of processing information. Actions of the automata (i.e. cells) are guided by their surrounding cells in the local neighborhood based on specific transitional rules over a number of iterations (Gardner 1970; Wolfram 1984). Thus, the basic CA formalism consists of five components: cell space or grid, cell state, neighborhood influence, transitional rules, and time step (Batty 1997).

In a typical CA model, discrete geographic space is represented in a raster environment consisting of regularly-tessellated cells (often square) of the same size. However, the rasterized representation of the geographic space is ineffective because the uniform shape and size of cells do not match the shape and size of real-world geographic objects (O'Sullivan 2001; Stevens and Dragecivic 2007). For instance, urban landscape features such as city blocks and cadastral units are irregular, and therefore are better represented as polygons of irregular configuration (White and Engelen 2000; Benenson and Torrens 2004). This motivated the researchers to develop vector-based cellular automata, hereafter irregular CA, as an alternative framework to the raster CA, hereafter regular CA (e.g. Shi and Pang 2000; Benenson et al. 2002). Irregular CA models developed using vector polygons to represent landscape features such as land use patches and cadastral parcels produce better results than the regular CA models (Moreno et al. 2009; Yumba and Dragecivic 2012).

Neighborhood configuration (e.g. type and size) plays a critical role in CA modeling as it operates based on the notion of local influence whereby state of a cell (automaton) changes in accordance with the state of surrounding cells (O'Sullivan and Torrens 2000). The shape and size of neighborhood determines the number of cells that it encompasses. While the model outcome of regular CA has been reported to be sensitive to the neighborhood type and size (Kocabas and Dragicevic 2006), it is little known whether the same finding applies to the irregular counterpart. Moreover, raster-based neighborhoods such as von Neumann and Moore are not applicable to the framework of irregular CA due to the asymmetrical shape and non-uniform size of the irregular polygons. Therefore, alternative neighborhood definitions must be applied to the irregular

CA. The primary purpose of the present study was to conduct a sensitivity analysis of the neighborhood size and type to the outcome of irregular CA models. In addition, various neighborhood types of irregular CA of urban land use dynamics are explored, thereby expanding the existing scholarship on neighborhood typology.

4.2 Literature review

The outcomes of geographic CA are susceptible to changes in model parameters. Specifically, variation in any of the five components of the CA formalism generates a change in the model output (Sante et al. 2010). For instance, different sets of transitional rules or neighborhood types result in different simulation scenarios (Batty 2005). Various scholarly works have documented the model sensitivity of regular CA to cell size and neighborhood configuration. For example, Menard and Marceau (2005) examined five different cell sizes and six neighborhood sizes, and found that a nonlinear relationship exists between spatial scale and the simulation result. Jantz and Goetz (2005) observed a difference between the results of their CA model with changes in cell size, confirming that the SLUETH model of urban growth performs better at a finer resolution (45-90 m cell size) when measured in terms of the *Compare* metric, a ratio of matched urban pixels in the modeled and reference maps. Samat (2006) characterized up-scaling effect on the results of a land use change CA model and concluded that selection of spatial resolution should be made carefully for obtaining better results. Kocabas and Dragicevic (2006) used circular and rectangular neighborhood types with spatial resolutions of 50m, 100m, 150m, and 250m. The neighborhood size in the model varied from 2 to 10 cell radii. As expected, they concluded that CA models are sensitive to cell size, neighborhood size and neighborhood type. They also recommended making sensitivity analysis a mandatory

step in CA modeling to improve the accuracy and precision of the results.

In addition to testing the effect of cell size, and neighborhood size on the model outcome, Pan et al. (2010) confirmed the effect of spatial extent as well. They suggested a spatial resolution of 25m and rectangular neighborhood of 9x9 cells for better results. The sensitivity of regular CA modeling has been also verified in terms of asynchronous updating (Baetens et al. 2012), stochastic component (Garcia et al. 2011), and probability distribution (Kim 2013), although the majority of the works has focused on cell size and neighborhood type as the important dependency factor.

In irregular CA, however, the problem of cell size variability is eliminated with the use of vector data (Moreno et al. 2008) such as parcels, which perfectly match the shape and size of the real-world geographic objects. Yet, the use of irregular spatial units does not obviate the variability of model behavior to neighborhood configurations (*ibid*). Rather, it complicates the definition of neighborhood. Simple definitions such as Moore neighborhood are impossible because of the non-uniform size and shape of the polygons. Moreover, a neighborhood of a specific size and shape does not always have the same number of polygons, unlike the rasterized equivalent. This is likely to add to the sensitivity of irregular CA to changes in neighborhood definitions. Only a few studies have directed their focus on analyzing the sensitivity of irregular CA.

Stevens and Dragicevic (2007) defined neighborhoods based on topological functions of adjacency and proximity. They suggested using different neighborhood types for different purposes. For example, they used adjacency to find out whether or not a park parcel was next to a residential parcel, and proximity to determine if the park parcel was within a specific distance from the residential parcel. However, these definitions were not

examined with respect to their sensitivity to the modeled results. Crooks (2010) used distance neighborhood demarcated by a buffer of specific radius. According to his definition, if a geographic barrier such as river crosses the buffer, only the entities that fall within the central part of the buffer are counted as neighbors. The author verified the effect of neighborhood size on model results and concluded that clusters of segregated point-neighbors were larger and more distinctive when used with increased buffer distance. Moreno (2008) used neighborhoods defined by external buffer of 10m, 30m, 60m and 120m in his Vector-based Geographic Cellular Automata (VecGCA) model to simulate land use dynamics in an agroforested region, and found variation in the results, with smaller neighborhoods performing slightly better. Moreno et al. (2009) introduced a dynamic neighborhood to the VecGCA by incorporating the concept of distance decay, thereby neutralizing the effect of neighborhood size. However, this method computes the state-transition influence exerted by other polygons on the central polygon based on non-geometric properties (e.g. land use information), and thus is a method of computing transition probability rather than defining a neighborhood. Ballestores and Qiu (2012) used a buffer of specific distance from the outer boundary of a polygon object to define the neighborhood. They applied three different distances and observed only a slight impact on the overall accuracy of simulation results.

Only a few previous studies explored different types of neighborhood to be used in the irregular CA and even fewer examined the effect of neighborhood configuration. Moreover, the investigation was limited to varying neighborhood sizes within a single type of neighborhood (Moreno 2008; Crooks 2010, Ballestores and Qiu 2012). No scholarly work to date has investigated the bivariate effect of neighborhood type and size

in the outcomes of irregular CA. The present study aimed to fill this research gap by conducting a detailed sensitivity analysis of neighborhood configurations. First, a comprehensive inventory of neighborhood definitions (types) possible in irregular CA was proposed. Based on the eight neighborhood definitions, a total of 26 different neighborhoods were generated by varying the neighborhood size to be used in subsequent urban growth modeling. The modeled scenarios were then compared with the historic land use data of the study area for validation.

4.3 Neighborhood types

In a cellular grid, von Neumann neighborhood includes the four cells adjacent to left, right, top and bottom side of the central cell, while Moore neighborhood captures all the 8 cells touching the central cell. Given the non-regular shape and non-uniform size of polygons in irregular CA, the popular neighborhood configurations of von Neumann and Moore (or their extension with radii of specified number of cells from the focal cell) are not applicable. A neighborhood must be defined based on geospatial operations such as point-in-polygon, and topological functions such as adjacency and proximity. Unlike the raster counterpart, these neighborhoods are not symmetrical in shape, and the number of neighbors (i.e., polygons) in a defined neighborhood varies for each polygon. Eight types of irregular CA neighborhood are defined and implemented in the study as shown in the first column of Table 4-1. For each of the types (except for the adjacency neighborhood), four different distances of 120m, 240m, 360m, and 480m have been used to decide the neighborhood size, resulting in a total of 26 scenarios. The neighborhood typology is described in the following subsections.

Table 4-1. The eight neighborhood types defined in this study. The number of neighboring polygons and the total area covered by them is reported for a parcel (processing polygon). The values are referenced to Figures 4.1-4.4 and buffer distance of 240 meters was used.

Neighborhood types	Number of neighbors (for polygon I)	Total area of neighbors (in hectare)
Adjacency	16	10.73
Extended adjacency	38	12.48
Boundary proximity	287	46.24
Centroid proximity	82	16.41
Boundary intercepted buffer	280	44.11
Centroid intercepted buffer	82	16.41
Boundary extent-wide	(Impact score*) 5117.53	10649.63
Centroid extent-wide	(Impact score*) 4984.31	10649.63

4.3.1 Adjacency neighborhood

A straightforward way of defining neighborhood is based on adjacency relations. Neighbors of an irregular automaton (i.e. the polygon object) are defined as a set of all other polygons sharing a common edge or points with the polygon being considered. As shown in Figure 4.1a, this neighborhood includes comparatively fewer members. Larger polygons tend to have more neighbors than the smaller ones. In Figure 4.1a, the smaller focal parcel (Parcel II) surrounded by street on three sides has only two adjacent parcels while the larger one (Parcel I) has many parcels spatially attached to it. However, this way of computing neighbors may be less accurate, specifically if the central polygon has streets or minor dividing lines on its side(s). Very often, the neighborhood of a polygon can extend beyond the street adjacent to it and include the lots across the street, especially in the single family residential development. For example, darker lots across the street from both focal polygons I and II in Figure 4.1b are close enough to exert an influence, and hence to be considered as the neighbors. Although it is a bit more challenging computationally, this extended adjacency neighborhood represents the local influence

more realistically. Conceptually, it is an equivalent to Moore neighborhood of the regular CA.

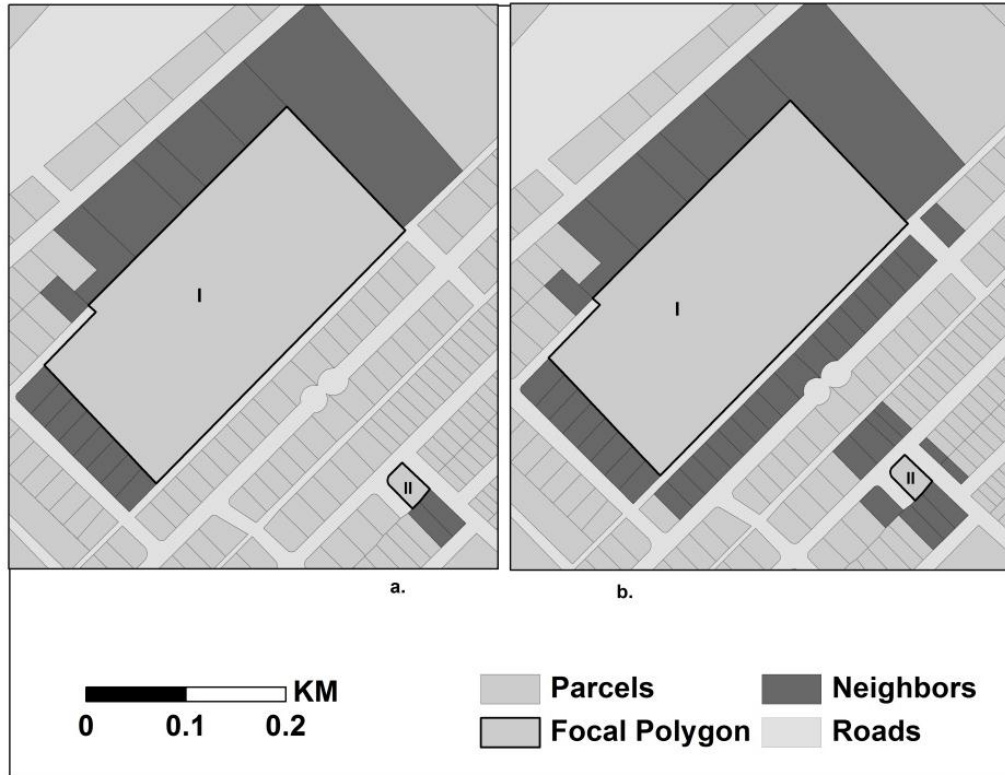


Figure 4.1. Adjacency neighborhood. a) Polygon units defined to be neighbors of focal polygon I and II based on adjacency relation; b) neighborhoods of focal polygons I and II based on extended adjacency.

4.3.2 Proximity neighborhood

Another method of computing neighbors for an irregular automaton (or polygon object) is to define a region of space based on spatial proximity that would include objects more than just the adjacent ones. This is delineated by using a specific distance from the focal object. In the boundary proximity neighborhood, a buffer of specific distance from the polygon boundary is created and all other polygons that intersect the buffer are included (Figure 4.2a-b). For the centroid proximity neighborhood, on the

other hand, a buffer of the same distance is drawn from the parcel centroid and the neighbors are selected based on the complete containment of centroids of the other polygons in the buffer (Figure 4.2c-d). It is noted that there is a conspicuous difference in the number of neighbors captured within a buffer of the same distance by the two methods (Figure 4.2). Apparently, the number of neighbors and the total area they occupy is smaller for centroid proximity neighborhood than the boundary alternative for a particular polygon (Table 4-1). For example, the neighborhoods in Figure 4.2a and Figure 4.2b are larger than that of Figure 4.2c and Figure 4.2d respectively. However, the difference or the ratio of the neighborhood size between these two types is greater if the focal polygon is larger. For instance, the difference in the number (and the total area) of neighbors for Parcel I between Figure 4.2a and Figure 4.2c is larger than the corresponding difference for Polygon II between Figure 4.2b and Figure 4.2d. Likewise, bigger polygon (e.g. Parcel I in Figure 4.2a) has more neighbors and therefore a larger neighborhood size than the smaller polygon (e.g. Parcel II in Figure 4.2b) for the boundary proximity type. But, it may not be true for the centroid proximity. In fact, the number of neighbors as well as the areal size of the neighborhood is larger for the smaller polygon (e.g. Parcel II in Figure 4.2d) than the larger polygon (e.g. Parcel I in Figure 4.2c). However, these generalizations can be safely made only if the neighboring polygons of the irregular CA lattice are of similar size for both the focal polygons.

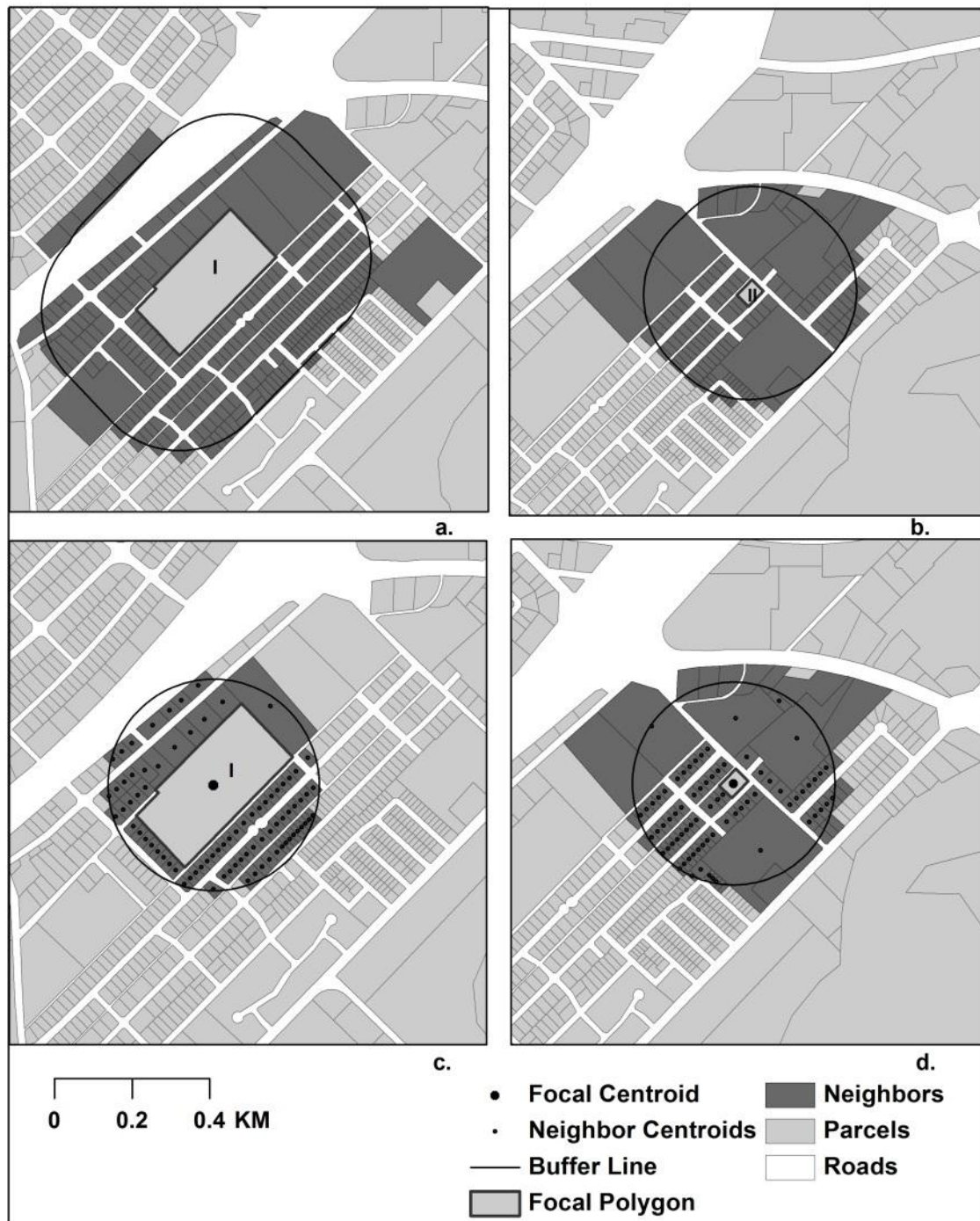


Figure 4.2. Proximity neighborhood. a) Neighborhood of polygon I defined by an external buffer of 240 m; b) Neighborhood of a smaller polygon II defined by an external buffer of 240 m; c) Neighborhood of polygon I defined by a buffer of 240 m from the centroid; d) Neighborhood of a smaller polygon II defined by a buffer of 240m from the centroid.

4.3.3 Intercepted buffer neighborhood

For some polygon objects, the neighborhood defined by a buffer of specific distance may be intersected or divided by topographic barriers such as major highways, rivers and flood zones, faults, gorges, and hills. The polygon on the other side of the barrier despite its inclusion into the defined buffer cannot exert an influence enough to be considered as a true neighbor (Crooks 2010). As shown in Figure 4.3a, eight lots (i.e. ones with slightly darker shade) lying north-west to the interstate highway are not considered as neighbors of focal parcel I although they have been intersected by the external buffer. Similarly, centroid intercepted buffer neighborhood is delineated by using a buffer from the centroid rather than the boundary of the polygon (e.g. Parcel I in Figure 4.3b). All other polygons with their centroids contained by the buffer but not separated by a barrier are counted as neighbors of the focal parcel.

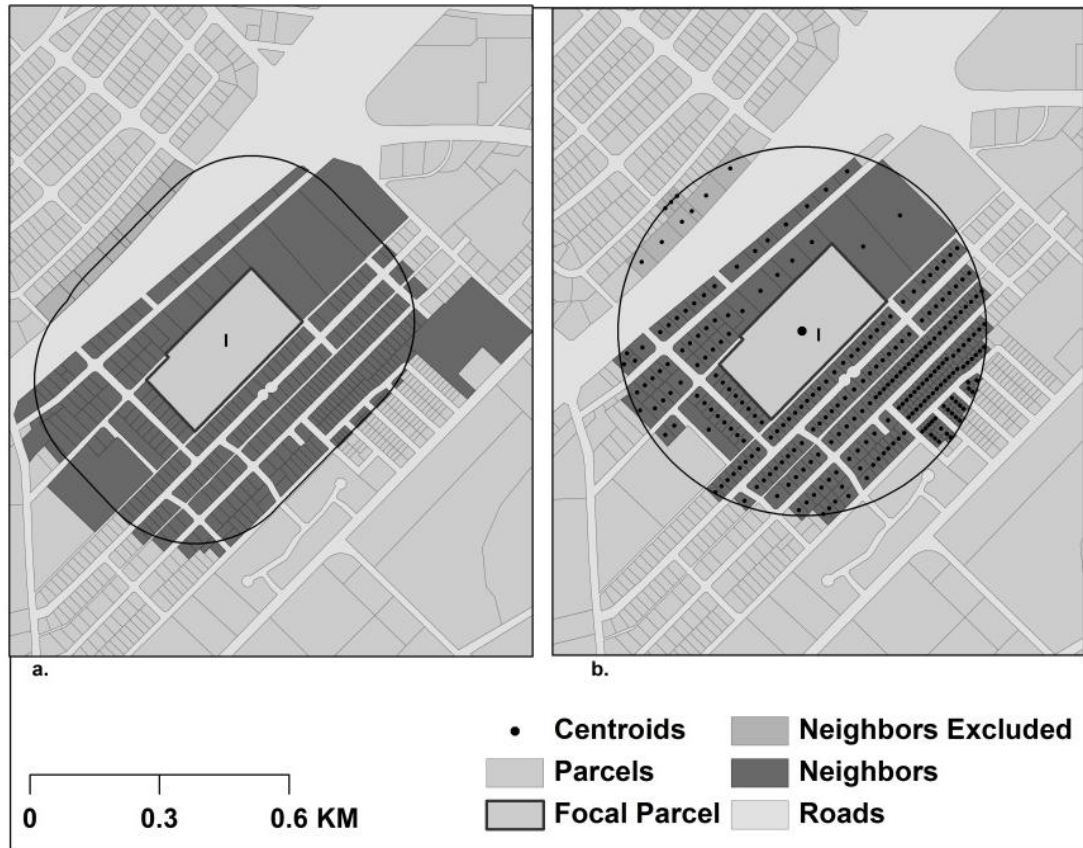


Figure 4.3. Intercepted buffer neighborhood. a) Defined by a buffer of 240m from the boundary of polygon I but excluding the area beyond the main highway; b) defined by a buffer of 360m from the centroid of polygon I but excluding the polygons across the main highway.

4.3.4 Extent-wide neighborhood

In order to include the influence of actions and entities at distance, this type of neighborhood includes the whole geographic extent of the simulation (Figure 4.4). Thus, the number of neighbors is same for each polygon in the study area. However, all the neighbors do not have equal influence on the focal polygon. Nearer neighbors exert higher amount of impact than the distant ones. This is taken into account by introducing a distance decay function. For simplicity, I assume that the influence of surrounding units on the focal polygon is optimum within a close proximity, which fades off linearly in the

outward direction before reaching zero at a specified distance threshold. As illustrated in Figure 4.5, a value of 1 is assigned to all the polygons that are within a user-assigned distance (35m in the figure) from the focal polygon to indicate that these neighbors have had a full influence and are referred as the core neighbors. Likewise, the polygons beyond the buffer of a user-assigned distance (115m in the figure) are assigned a value of 0 to indicate that they have no influence on the central parcel. The impact value for all the intermediary polygons is calculated as:

$$NI_i = (Max_i - Value_i) / Max_i \quad (4.1)$$

where, NI_i is neighborhood impact of polygon i on the focal polygon, Max_i is the distance of the farthest polygon from the focal parcel, and $Value_i$ is the distance of polygon i from focal polygon. Sum of impact values for all the polygons in the study area are added up to get the neighborhood impact for the focal polygon. The buffer distance assigned to capture the core neighbors influences the total impact score for a polygon. In this study, four buffer distances were assigned to demarcate the core neighboring areas resulting in four different scenarios. In boundary extent-wide neighborhood, the buffer is drawn from the boundary of the focal polygon and all other polygons intersecting the buffer are counted as neighbors, whereas the centroid extent-wide neighborhood is characterized by a buffer from the centroid. Only the polygons having their centroids within the buffer are selected as the core neighbors.

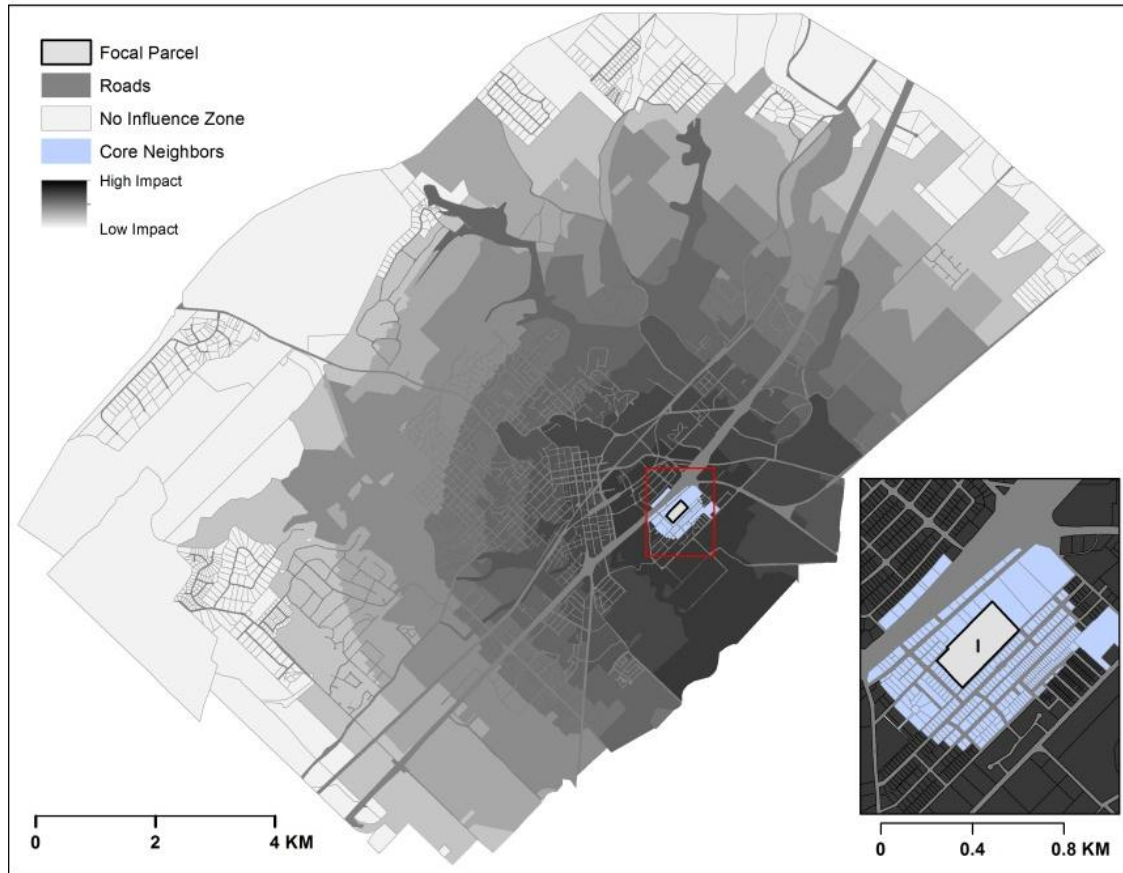


Figure 4.4. Extent-wide neighborhood showing the core polygons exerting high impact on central polygon I and decreasing amount of impact along the distance from the focal polygon. The inset map highlights the core part of the central polygon I.

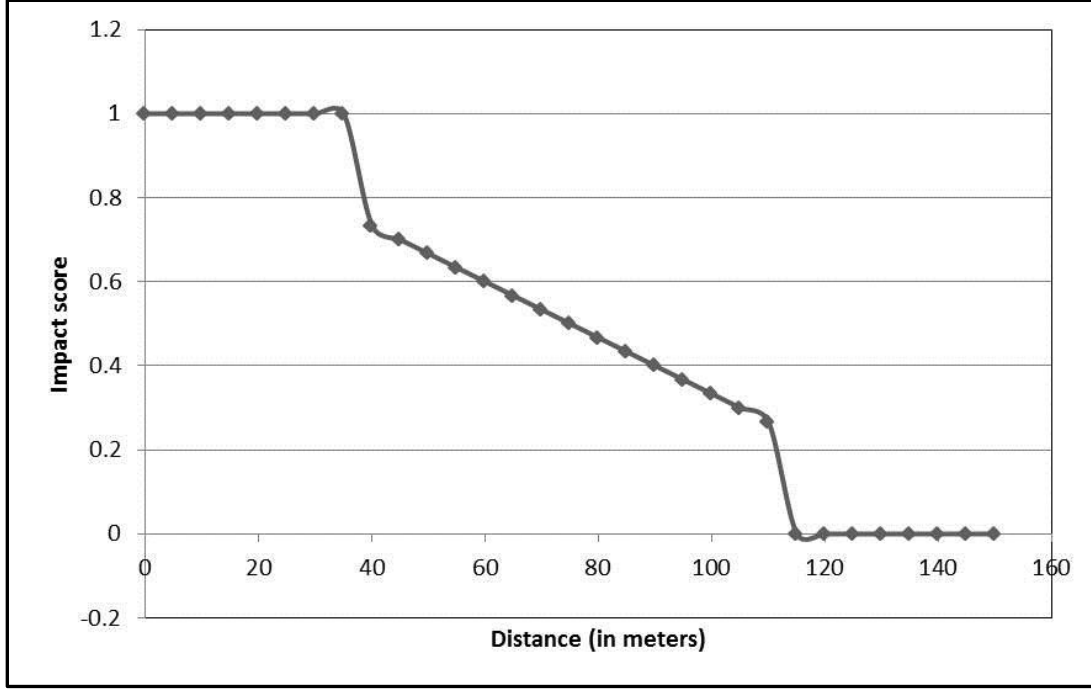


Figure 4.5. Distribution function of impact scores for polygons across the entire study area

4.4 Urban growth simulation

Following the conceptual framework of the AIIA (Subsection 2.4), an irregular cellular automata model was developed and implemented to simulate urban land use dynamics in San Marcos, Texas. The primary function of the model, which uses cadastral parcels as the unit of operation, is to determine the conversion of undeveloped lands of different land use categories into developed state. Each vector polygon (i.e. the parcel) has a development state of either ‘developed’ or ‘undeveloped’ in addition to one of the six land use attributes as shown in Figure 4.6. State conversion of an undeveloped parcel is the function of its development suitability and neighborhood impact expressed as:

$$DP_i^t = f(S_i^t, NI_i^t) \quad (4.2)$$

where, DP_i^t is development potential of parcel i at time t , S_i^t is its suitability at time t for future development, and NI_i^t is the influence of developed neighbors on parcel i at time t .

State of an undeveloped parcel is updated to ‘developed’ at time $t+1$ based on its development potential at time t .

Neighborhood impact is calculated as mentioned in the previous section as a total number of developed neighbors, while development suitability of a parcel is calculated based on the influence of various driving forces. The urban growth drivers included in this study, and their respective weights were derived from the coefficients of a logistic regression analysis (Table 4-2). The development status of parcels (i.e. whether developed or undeveloped) was used as dependent variable and the mentioned factors as independent variables for the baseline year of 2000 as mentioned in Subsection 2.4. Each factor receives different development weights for residential, commercial and industrial land use categories. The model computes the suitability score for each polygon each time step based on the following equation after Barredo et al. (2003):

$$S_i^t = \sum_i^n (W_i * F_i) + \varepsilon \quad (4.3)$$

where, S_i^t is the suitability score of parcel i at time t , F_i is the value of a factor for parcel i , W_i is the weight of that factor at time t_0 , n is the number of factors included (11 in this case), and ε is the stochastic disturbance factor computed as a random number ranging between 0 and 1. To examine the impact of neighborhood variation, other parameters were held constant except for the neighborhood type and size. Similarly, the stochastic term was given a value of 0 to minimize random noise and control the subsequent variability. Each parcel has three different suitability scores corresponding to three different weights or preferences for residential, commercial and industrial development based on the assumption that a parcel may be more preferable for one land use development over the others. For example, a parcel closer to already developed

residential neighborhood receives higher preference score for residential development than for commercial or industrial. The same applies to commercial and industrial categories as well.

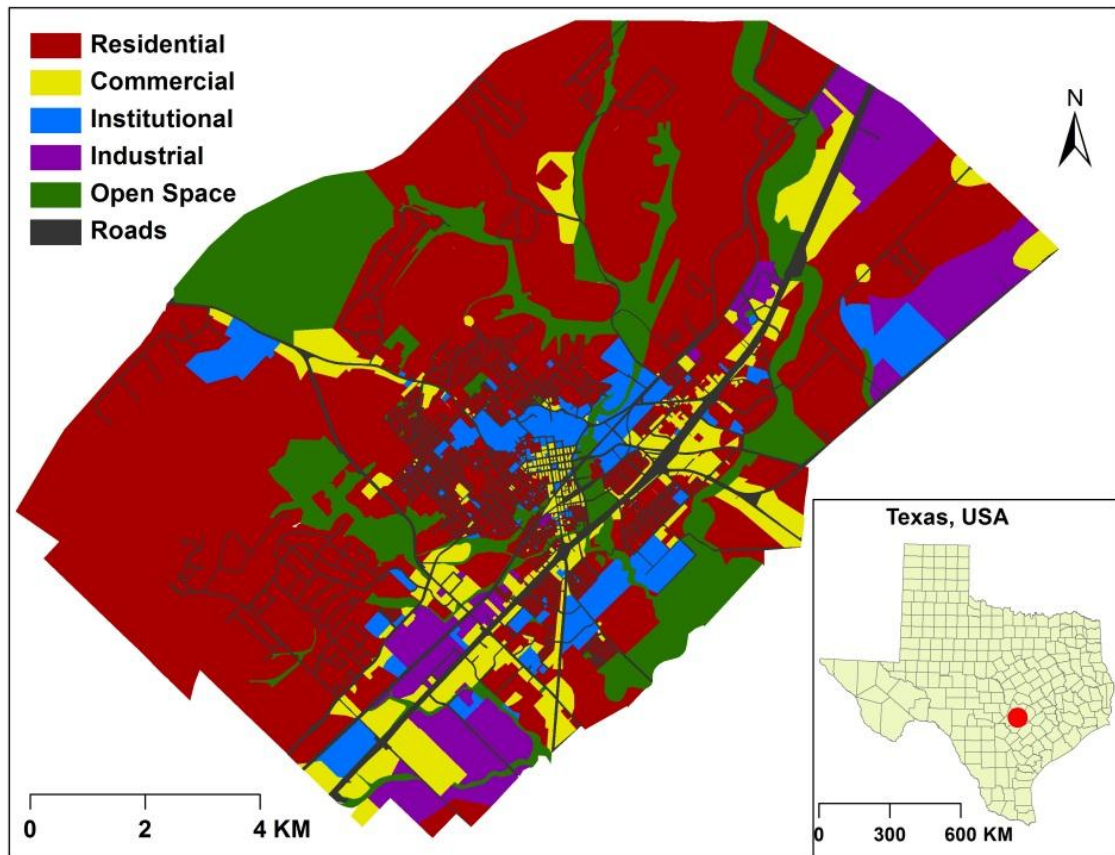


Figure 4.6. Geographic location of San Marcos, Texas, the study site, and six land use categories used in the model.

Table 4-2. Urban growth factors included in the present study and corresponding weights for three different categories of land use development.

Factors		Development weights		
		<i>Residential</i>	<i>Commercial</i>	<i>Industrial</i>
Environmental	Elevation	1.8	-3.0	0.0
	Slope	-1.2	-7.8	-7.7
	Distance to rivers	0.4	1.3	0.0
Transportation accessibility	Distance to IH-35	-5.6	-3.9	0.0
	Distance to railroads	0.0	3.9	-1.7
	Distance to major roads	-4.9	-4.1	-2.4
	Distance to airport	-3.8	2.0	-2.9
Centrality influence	Distance to Texas State University campus	1.2	-31.4	0.0
	Distance to city hospital	-1.2	-2.4	1.7
	Distance to San Marcos outlet mall	-0.5	-2.0	-2.8
	Distance to CBD	-0.9	-29.4	0.0

The model estimates the total area required to be developed at time $t+1$ for residential, commercial, industrial and institutional land use zones separately as described in Subsection 2.4.1. Undeveloped parcels with the higher development potential score are selected each time step for development in each of these categories (except for institutional) to the estimated amount. The model assumes a strict implementation of zoning regulations, limiting development only within the assigned land use categories. For instance, if the model estimates a total of 100 acres of residential land for development in time $t + 1$, then parcels with the highest development potential score on the ascending order are selected until the total area exceeds 100 acres within residential land use zone. The selected parcels are subdivided in accordance with their shape, size and orientation. Roads are extended to the parcel selected for development if it is disjoint to the existing roads. The ‘undeveloped’ state of the subdivided parcels in residential,

commercial and industrial land use categories is updated to ‘developed’. But, for the development of lands in public and institutional land use category, the estimated amount of undeveloped area within this zone is selected randomly and the state updated to ‘developed’. The programs and algorithms used for the subdivision of parcels were adapted from Chapter 3. A single iteration in the simulation represents a year, and the model runs for 12 years giving out scenario developments for the year 2012.

4.5 Results and model evaluation

In this study, 26 scenarios were defined based on different neighborhood configurations. A simulation map of urban growth was created for each of the scenarios by using datasets for the base year 2000. For the illustration purpose, simulated results for three scenarios and reference map are shown in Figure 4.7. Each map in the figure shows developed area in the base year, newly developed area or the development occurred during the simulation period, undeveloped or available lands, and exclusion zones or the areas prohibited for future development. Each of the output maps was evaluated against the reference land use data, which delineates developed and undeveloped states of parcels in different land use categories. As the goal was to investigate the variation in model outputs due to changes in neighborhood type and size, different accuracy assessment and pattern-based metrics were computed for the scenario results (Table 4-3).

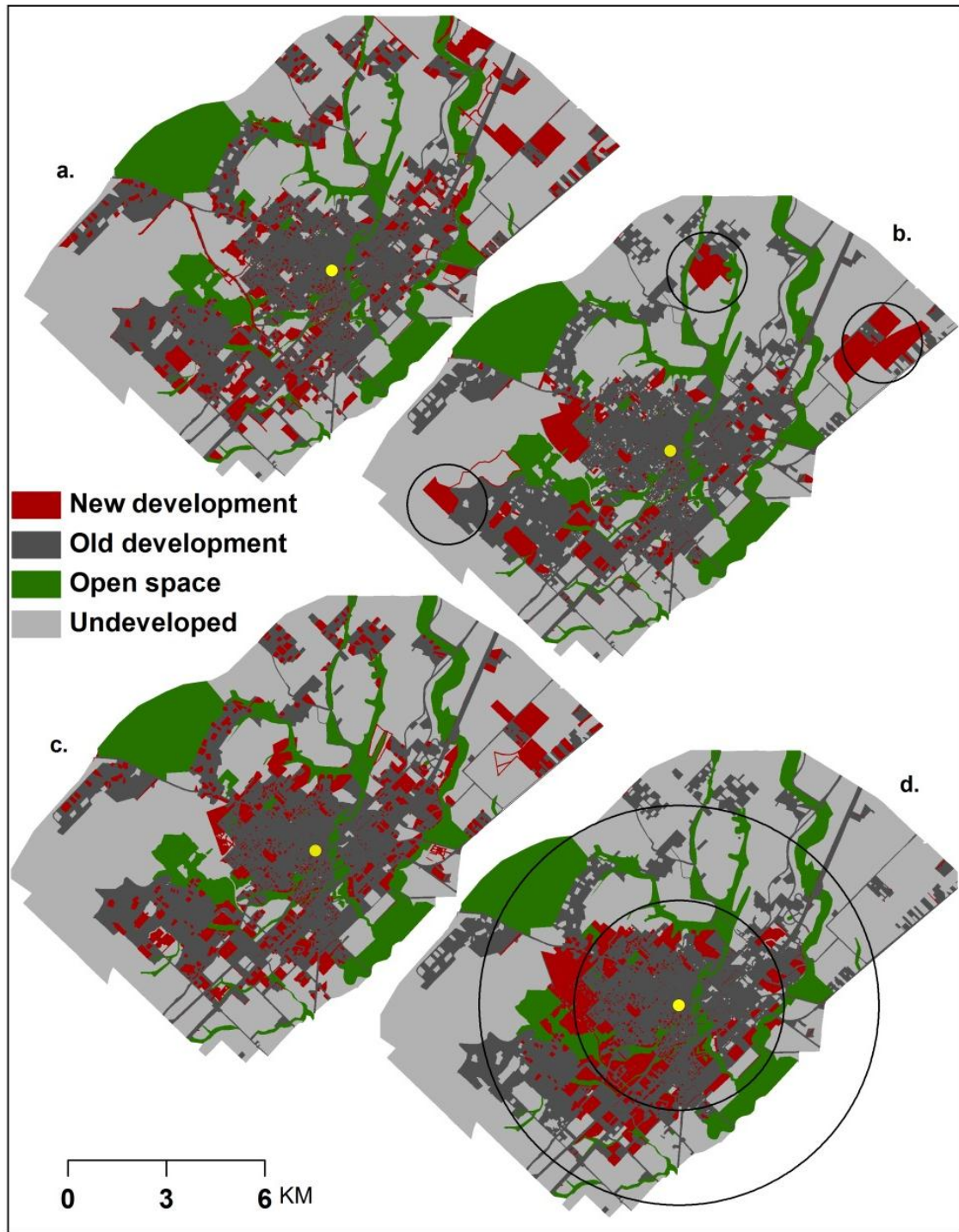


Figure 4.7. Results for scenarios defined by three selected neighborhood configurations. a) Reference map for year 2012; b) output for extended adjacency neighborhood where three black circles emphasize peripheral urban clusters; c) output defined by the neighborhood of centroid intercepted buffer with 400 feet distance; d) output for centroid extent side neighborhood with 1600 feet buffer and two concentric circles. The yellow dot represents the city center.

In order to assess how well the model captures the rate of urban expansion vis-à-vis the observed data, Error Percentage (Error %) was computed for each of the 26 scenario maps as:

$$Error\% = (Dev_{Ref} - Dev_{Sim}) / Dev_{Ref} \quad (4.4)$$

where, Dev_{Ref} is the amount of total developed area in reference map and Dev_{Sim} is the amount of total developed area in simulated map. This statistic shows non-site-specific accuracy (or error), in which only the total areas of modeled categories are compared without considering whether the compared sites spatially agree or disagree. Table 4-3 shows the error of less than 5% (or an accuracy of more than 95%) for all the scenarios, suggesting that the model performs well in terms of the overall rate of development. The fact that this statistic is almost the same for all scenarios means the variation of neighborhood type and size has a negligible influence in the total amount of modeled development. Thus, the rate of development is not sensitive to neighborhood configurations but other parameters such as the annual growth rate and the size of geographic units resulting from subdivision that occurs during simulation.

Apparently, the non-site-specific statistic does not account for the spatial coincidence between the modeled output and reference data. Therefore, Kappa Index (Congalton and Green 2009) – a measure of spatial similarity – was computed. Based on Table 4-3, it is clear that the Kappa values substantially vary with the changes in both the type and size of neighborhood. The variation due to neighborhood type is more remarkable than the variation due to size. This is evident because the Kappa values are substantially different between the neighborhoods defined by boundary-proximity (0.60-

0.65) and centroid-proximity (0.75-0.81) (Table 4-3). But, the values are only slightly different for the neighborhood types along the four different values of buffer distance. It is also clear from Table 4-3 that the centroid intercepted buffer neighborhood yields the best spatial match of the newly developed sites. The proximity neighborhood using centroids rather than the external buffer results in a better match. Within the centroid-based proximity neighborhood, spatial accuracy slightly decreases with the increased neighborhood size.

As the Kappa value is often inflated because of the presence of comparatively large amount of undeveloped lands in the study area, thereby tending to overestimate the correspondence between modeled and observed maps (Wang et al. 2011), another measure of spatial match called Lee and Sallee Index (Jantz and Goetz 2005) was computed. This index is calculated as a ratio of the intersection to the union of the simulated and actual developed areas for the final simulation year. As expected, the values of Lee and Sallee index are smaller than the corresponding Kappa values. This statistic also shows variability in model results due to the changes in size and type of neighborhood. Table 4-3 reveals that values of the Lee-Sallee index are proportionally related to the values of Kappa index. In other words, Lee-Sallee index is higher for the neighborhood configurations that have resulted in higher Kappa values and lower for those that have lower Kappa values.

Table 4-3. Evaluation of 26 simulation scenarios characterized by neighborhood types and sizes using different spatial pattern and accuracy metrics. Average patch size is in hectare.

Neighborhood types and proximity		Similarity (accuracy)			Pattern metrics			Radial development (% developed area)		
		Aggregate (% error)	Spatial (Kappa)	Lee-Shalee Index	No. of patches	Average patch size	Moran's I	Core	Sphere	Outland
Adjacency		2.32	0.62	0.34	658.00	22.08	0.16	25	39	36
Extended adjacency		3.95	0.65	0.37	604.00	24.05	0.19	29	42	29
Boundary proximity										
Buffer distance (meter)	120	3.83	0.62	0.35	449.00	32.35	0.18	40	48	12
	240	0.55	0.60	0.33	354.00	41.03	0.23	44	44	13
	360	1.56	0.61	0.25	355.00	40.92	0.20	45	43	12
	480	4.66	0.65	0.26	343.00	42.35	0.20	49	41	10
Centroid proximity										
Buffer distance (meter)	120	4.89	0.81	0.63	139.00	104.47	0.22	35	42	23
	240	4.49	0.79	0.59	159.00	91.35	0.29	45	39	16
	360	3.39	0.77	0.53	191.00	76.04	0.25	49	40	11
	480	2.76	0.75	0.44	215.00	67.55	0.24	46	44	9
Boundary intercepted buffer										
Buffer distance (meter)	120	0.19	0.69	0.34	419.00	34.65	0.18	34	49	17
	240	3.28	0.62	0.35	361.00	40.32	0.19	39	47	14
	360	4.91	0.60	0.33	364.00	39.92	0.10	38	47	14
	480	2.49	0.61	0.34	370.00	39.31	0.12	41	46	13
Centroid intercepted buffer										
Buffer distance (meter)	120	4.20	0.92	0.74	146.00	99.48	0.29	36	43	21
	240	3.19	0.91	0.73	166.00	87.51	0.23	40	43	17
	360	2.78	0.88	0.69	185.00	78.54	0.27	45	42	13
	480	0.89	0.85	0.66	230.00	63.15	0.19	49	39	12
Boundary extent-wide										
Buffer distance (meter)	120	4.00	0.66	0.41	330.12	43.11	0.41	55	43	2
	240	4.33	0.68	0.42	333.00	43.62	0.43	54	44	2
	360	4.00	0.69	0.43	334.27	43.76	0.45	54	43	2
	480	3.15	0.68	0.42	333.00	43.62	0.49	54	43	2
Centroid extent-wide										
Core radius (meter)	120	3.80	0.72	0.49	287.00	50.61	0.39	57	42	2
	240	3.01	0.73	0.47	284.00	51.14	0.41	57	43	1
	360	4.00	0.71	0.76	283.11	50.72	0.41	55	42	2
	480	4.54	0.73	0.47	284.00	51.14	0.44	57	41	2

Although comparison-based statistics of accuracy assessment such as the Kappa index are useful, they are limited because of their overemphasis on locational match-finding at the expense of revealing spatial patterns. Moreover, pattern-based indices, rather than the comparison statistics, are more appropriate to assess the simulation results of complex systems like cities (White and Engelen 2000). The number of patches (NP) and average patch size were calculated to measure landscape configuration, which highlight the frequency and size of developed sites. A higher NP indicates spatial heterogeneity or a spotted distribution of urbanized areas in the landscape. Table 4-3 shows there are fewer urban patches for the neighborhood types (e.g. centroid proximity and centroid intercepted-buffer) that yield a higher spatial match as measured by Kappa index. Obviously, the number of patches and the average patch size are inversely proportional. That is, the larger the number of patches is for a neighborhood type the smaller the average patch size.

Moran's Global I, which identifies whether development is clustered or dispersed, was also computed. The results show that the boundary extent-wide neighborhood produces the most clustered development. Moran's I is the largest for this type of neighborhood when the size is set to a buffer of 480m. The development appears to be more dispersed as the buffer distance increases. Because of higher centrality influence, this neighborhood type promotes development in and around the city center. For example, even a parcel without any immediate developed neighbor can be developed if it is in the core because the sum of distance-weighted impact value for development is higher there. In other words, although an undeveloped parcel in the periphery is surrounded by many developed parcels, it may not have impact score high enough to be

selected for development. As a result, the fringe development is discouraged by this neighborhood type.

A city is often seen as a diffusive process, composed of centrifugal layers of urban areas (Burgess 1925). In order to observe morphological patterns of the outgrowth, a radial analysis was done. For each of the 26 scenario results, the study site was divided into three different concentric zones: core, sphere and outland (Figure 7d). Then, the amount of newly developed area within each of the zones was calculated. In this study, the core comprises the central part of the city, encompassing areas that are within a radius of 3,048m (10,000ft) from the CBD. Sphere covers the annulus region defined by two radii of 3,048m and 6096m (20,000ft), while the areas beyond 6,096m constitute the outland zone. Table 4-3 shows that extent-wide neighborhoods promote development in and around the city center. The centroid extent-wide neighborhood causes the highest concentration of development in the core region. Some 57% of the total developed area during the simulation period has fallen within the core, whereas only about 2% has been developed in the outland region (Table 4-3). The varied buffer distances used to define the maximum influence area in the two extent-wide neighborhoods has only a slight impact in terms of radial development. Also, there are only modest differences when using the centroid or the external boundary. This type of neighborhood promotes more organic development concentrated in and around the city center. Adjacency neighborhoods, on the other hand, promote periphery development. New development is likely to emerge around any developed parcel in the study site regardless of its radial zone with this neighborhood.

4.6 Discussion and Conclusion

Computation of neighborhood influence on a focal cell within a defined localized region plays a critical role in CA modeling. It has been well documented in the literature that the output of regular CA is highly sensitive to neighborhood configurations, which ultimately determines the overall quality of the model. The present study has examined how model outcomes correspond with the variation in neighborhood shape and size of irregular CA. In addition, it explored eight different types of neighborhood applicable to irregular CA models. This addresses the first research question of this project (i.e. RQ 7) that eight neighborhood types (shown in Table 4-1) were defined and used to conduct sensitivity analysis. The urban growth model, which uses parcel data at the cadastral scale to represent the geographic space, was run to produce simulation scenarios characterized by 26 different neighborhood configurations. When appraised using different accuracy assessment and landscape metrics, the results verify the existence of model sensitivity to neighborhood configuration. This addresses the second research question (i.e. RQ 8) that configuration of neighborhood affect the outcome of the model by bringing changes in accuracy of the model and patterns of the simulated development as shown in Table 4-3 and Figure 4.7.

Although the rate of urbanization as measured by total developed area during simulation is not sensitive to neighborhood type and size, the overall accuracy as measured by Kappa and Lee-Sallee indices vary substantially with the variation in neighborhood type and size. The neighborhood type defined by a centroid intercepted buffer yielded the most accurate results for the city of San Marcos, which is characterized by sprawl development. This is obvious when Figure 4.7a and Figure 4.7c are visually

inspected. This answers the third research question (i.e. RQ 9). The centroid extent-wide neighborhood produced the most clustered development with higher concentration of the new urbanized lands near the city center, and thus discouraging the sprawl and leapfrog development. With the centrality influence it exerts, this neighborhood promotes organic development, whereby undeveloped parcels farther from the city center tend to be developed only when the developable parcels in the central area are unavailable. The urbanization spread outward in a radial fashion with more than 50% of the new development occurring within the inner circle (i.e. the core) as shown in Figure 4.7d. On the other hand, the adjacency neighborhoods also produced clustered development but not necessarily only near the city center. With this neighborhood, undeveloped parcels in peripheral area also are likely to be developed if they have immediate neighbors that are developed. The three circled clusters of newly developed area in Figure 4.7b illustrate that this neighborhood promotes leapfrog and spotted cluster development. Thus, it is shown that different development patterns emerge as a result of changes in neighborhood configurations, which address the fourth research question of this project (i.e. RQ 10).

The results reveal that spatial accuracy and morphological patterns of simulated urban development is sensitive to the neighborhood type and size. The centroid intercepted neighborhood with a buffer of 120m is the recommended configuration to the case of San Marcos, as it produced the highest spatial match with the reference data. However, this neighborhood type may not produce similar results in other geographic contexts. Other neighborhood configurations can be more appropriate for other urban areas that have unique features and locational variations. In order to select the appropriate neighborhood for a given study site and thereby ensure the quality of simulation, this

study, similar to Kocabas and Dragicevic (2006), confirmed the importance of sensitivity analyses as an integral part of the irregular CA modeling. This will further contribute to the understanding of model behaviors as well as a more accurate interpretation of results.

While a comprehensive neighborhood typology was presented in the study, additional neighborhood definitions are possible with irregular CA. Additional neighborhood types not investigated in the study can be defined by making different combinations of topological operations and buffer distances. For instance, a new type of neighborhood can be defined by using an external buffer that would capture the neighbors based on point-in-polygon operation using centroids. Future research should focus on this aspect, and explore additional, more efficient neighborhood types of irregular CA. It would be also interesting to examine the sensitivity of polygon size in the future work because the number of neighbors in a specific neighborhood is variable for each polygon in an irregular CA lattice. Case in point, a neighborhood defined by a buffer of the same radius for a particular focal polygon will have fewer parcels but larger area if the neighboring parcels are of larger size, and vice-versa. This variation is likely to impact the model outcome. Likewise, sensitivity analysis of model outputs to geographic extent should be included in the future work. Since the impact of the spatial extent of the study area on regular CA modeling has been well documented (Pan et al. 2010), the same is likely to apply to the irregular CA also.

5. SUMMARY AND CONCLUSION OF THE DISSERTATION

This dissertation research focuses on urban growth simulation modeling with irregular geometries as the unit of operation (i.e. the use of vector data). It has adopted a hybridization approach to build an integrated model that would bring together the CA and ABM systems. As the main research goal was to build a software model that is more robust than the currently existing urban growth simulation models, a prototype model called AIIA was developed. In addition, the subdivision component of the model was explained in detail, and the sensitivity of the model to neighborhood configuration tested. The findings of the overall research expand the corpus of literature on spatial urban growth simulation at the same time shedding important insights on urban theories.

The first part of the dissertation (Chapter 2) outlines the development and implementation of the AIIA mode. The model, which was calibrated and tested for the city of San Marcos, Texas, employs vector GIS (both data model and operations) at cadastral level. AIIA integrates agents' decision processes into irregular automata that are represented by cadastral land parcels. By encompassing the site selection behaviors also of commercial, industrial and public/institutional agents in addition to planner and household agents, the AIIA provides a holistic framework capable of capturing the processes of urban growth. This increases the overall, user's and producer's accuracies as well as the total percentage of matched area between simulated and reference maps, thus directly answering in affirmation RQ 1 (*"Does incorporating commercial, industrial and institutional agents improve accuracy of modeling urban development?"*) Change in urban land use occurs as a result of locational decisions made by the agents based on their

preferences or utilities, which are derived from empirical data in contrast to the existing practice of employing heuristic rules to govern the decision of the agents. This data-driven procedure avoids subjective biases inherent in the behavioral rules. As an improvement to the existing irregular automata-based modeling, AIIA includes classification of housing developer agent into two categories: single-family residential and multi-family residential. This has made the simulation more realistic, and contributed to the predictive power of the model. Similarly, the categorization of household agents into four types based on socio-demographic backgrounds has yielded important findings about the micro-dynamics of urban processes such as the distribution of low-income versus high-income households across the city. These two statements answer RQ 2, *“How does the categorization of household agents into four different types and residential developer into two types improve the modeling?”* AIIA provides an effective facility for producing different future scenarios, thereby being useful for testing various management strategies. The question *“What are the impacts of urban growth policies”* (i.e. RQ 3) is addressed by the results of the scenario simulation. It is revealed that the scenario that introduces ‘growth boundary’ promotes more compact built-in development compared to the scenario of continued development with current trends. Similarly, the results document that urban growth has increased linearly with a scattered sprawling pattern, and the city has witnessed a proportional increase in commercial development and a decrease in industrial development during the simulation period. This statement answers RQ 4 (*“What are the spatial patterns of modeled land development in San Marcos in years 2010 & 2020?”*).

The second research project (Chapter 3) is about *Parcel-Divider*, a vector-based

urban land partitioning toolset integrated as the subdivision component of the AIIA. It offers multiple subdivision styles to match the geometric and locational attributes of a land parcel selected for development. *Parcel-Divider* generates subdivisions of comparable realism. The solution has passed the tests of maintaining egress to each generated lot, ensuring the optimum utilization of available space, and creating lots of uniform size and regular shape. Moreover, the tools maintain continuation of the road networks by connecting to the candidate parcels that were disjoint to the existing road network, are user-friendly with ample support documentation and tutorials, and are faster and easier to operate while remaining computationally simple in algorithm and code logic. Statistical evaluation verifies that errors range between 0-6 percent, which means the tools produce lots with an accuracy of more than 94%. This answers RQ 5 (*“What is the percentage error of simulated lots in terms of different shape and size indices?”*). Visual inspection reveals that resemblance of observed and modeled subdivisions is higher. This answers RQ 6 (*“How closely do the modeled subdivisions resemble the observed subdivisions?”*).

The third part (Chapter 4) explores eight different types of neighborhood applicable to irregular CA models. This statement addresses RQ 7 (*“What are the possible neighborhood definitions (types) in irregular CA/ABM models?”*). Driven by RQ 8 (*“How does neighborhood configuration affect the outcome of irregular CA/ABM modeling?”*), the chapter examines how model outcomes correspond with the variation in neighborhood shape and size of irregular CA. When appraised using different accuracy assessment and landscape metrics, the results are very sensitive to neighborhood configuration. The results show that centroid intercepted buffer neighborhood produced

the most accurate model outcome, which is the answer to RQ 9 (*“Which neighborhood type yields the most accurate results?”*) Different neighborhood types promote different geographic patterns of the simulated cities. For example, a centroid intercepted buffer yielded the development characterized by sprawl. Centroid extent-wide neighborhood produced the most clustered development with higher concentration of the new urbanized lands near the city center, and thus discouraging the sprawl and leapfrog development. Adjacency neighborhoods also produced clustered development but not necessarily only near the city center, promoting leapfrog and spotted development. These statements answer RQ 10 (*“What kind of development patterns emerge as a result of variation in neighborhood types?”*) The study recommends the exploration of neighborhood configuration through sensitivity analysis as an integral part of the irregular CA modeling.

5.1 Significance of the study

The main contribution of the research is to improve the existing automata-based urban growth modeling by providing a holistic framework for fusing agents into irregular CA. It has examined the importance of commercial and industrial agents and utility-based functions in modeling urban dynamics. Specifically, the research findings have documented that incorporating commercial and industrial agents improve the modeling. Similarly, incorporating social theory (e.g. utility functions) improves modeling the urban development. This has advanced our understanding towards constructing a holistic framework that is capable of capturing important processes of the urban growth. Furthermore, the research has shed useful insights into the existing theories of urban development. For example it examined whether a development plan enforcing the

strategy of growth boundary would significantly promote mixed-use, high-density development.

The three specific articles have their own research implications and importance, which is mentioned in the following paragraphs.

The AIIA model

The AIIA model is directly applicable to the formulation of urban growth management policies. The framework serves as a useful instrument in understanding how various biophysical and socioeconomic variables affect urbanization. The planners and stakeholders can analyze and visualize how urban expansion occurs with different policy choices and scenarios. By offering scenario-based future land use maps to evaluate development alternatives, it provides guidance to the authorities to identify challenges of a sustainable urban future. The officials at the planning body of a metropolitan area find the model a hands-on tool to produce scenario-based land use zoning maps. Specifically, the local results and findings of the research will be useful to the San Marcos community (i.e. the study area).

The parcel subdivision toolset

The *Parcel-Divider* toolset can be useful for researchers as they can integrate it as a subdivision module into their urban simulation modeling. Likewise, the toolset can be used by land developers and planning officials to assess and monitor different designs of urban layouts in the developable lands. It is useful while conducting build-out analysis to estimate residential density and allocating land areas for roads and open spaces. Surveying technicians can utilize the tools for initial warm-up work which can give them guidance to conducting their field works. As part of an exploratory exercise, they can

apply the tool to the parcel they are going to partition, producing different scenarios for varying parameter values. This may provide insights on different possibilities.

Neighborhood sensitivity of the irregular CA models

The research has confirmed the importance of sensitivity analyses as an integral part of the irregular CA modeling. Sensitivity analysis contributes to the selection of appropriate neighborhood for a given study site and thereby ensuring the quality of simulation. Furthermore, this is useful to understand model behaviors as well interpret the model results more accurately.

5.2 Limitations and the ways forward

The present study has demonstrated that the integration of CA and ABM systems using vector data format is an efficient approach toward developing predictive simulation models of urban land use dynamics. Specifically, the AIIA model developed as part of the study has been proved more accurate and robust compared to the existing urban growth simulation models. However there is still further potential for improving the modeling and extending its applicability. Some of the major steps for future research are to:

- Integrate fuzzy logic into the hybridized modeling approach in order to compute more accurate suitability score or development potential for each of the vector units (such as parcels or census units), which ultimately enhances the predictive power of the modeling. This will add to the reliability and validity of the model.
- Develop a module to carry out sensitivity analyses of the modeling also to transitional rules to be applied, weights to be assigned to the driving factors of

urbanization, and stochasticity. Comparative studies should be carried out using weights derived from different methods including regression analysis, neural network, expert system, and analytical hierarchy process as part of the calibration of the model.

- Use a separate spatial data management system to handle geometry operations of the geographic objects, and improve the optimization of source codes so that processing time could be shortened and therefore being suitable for application to bigger cities as well.
- Make the model more generic so as to make it applicable to other cities.
- Distribute online (the improved version of the) AIIA to interested stakeholders, researchers, planning officials and students (free of cost to the extent it would be possible). The software modeling package should be accompanied by support documentation, multimedia tutorials (e.g. YouTube videos), webinars etc.

Future directions specific to the three research articles of the dissertation are to:

The AIIA model

- Create a module capable of carrying out logistic regression and then passing the coefficients (weights) to the suitability calculation module. This would enhance automation of the software because the present software requires that users carry out regression analysis separately in SPSS or other statistical packages and then input the coefficients into the model manually.
- Include additional urban growth driving factors that are important to the local context.
- Address errors inherent to the GIS data and model uncertainty.

The parcel subdivision toolset

- Include additional tools in the toolset as *Parcel-Divider* still falls short on generating the variety of subdivision layouts exhibited in the real-world urban landscape. For instance, there is no tool for appropriately subdividing an upside-down T-shaped parcel.
- Introduce appropriate sinuosity to the tool-generated streets because new roads created by the current tools are always straight.
- Take into account the terrain attributes like slope which in totality result in the variability of lot configurations as well as subdivision designs.
- Introduce appropriate shape recognition algorithm for detecting the parcel's shape so as to automatically decide which of the subdivision tools to be applied for a particular parcel. Currently, the tools recognize the size and orientation but not the shape of the parent parcel. Users should decide which of the tools to apply for a particular parcel.

Neighborhood sensitivity of the irregular CA models

- Explore additional neighborhood types not investigated in the article (i.e. except the eight types mentioned).
- Investigate the sensitivity of polygon size as well because the number of neighbors in a specific neighborhood is variable for each polygon in an irregular CA lattice.
- Conduct sensitivity analysis of model outputs to geographic extent.

In conclusion, the AIIA model built as part of this dissertation using irregular

spatial units produces accurate simulation scenarios of urban growth. It has been a promisingly useful decision support tool for developers and urban management authorities. The further upgrade of the model in the days ahead will only add to the model's robustness and universal applicability.

REFERENCES

- Al-Ahmadia, K., Seea, L., Heppenstall, A. and Hogg, J., 2009. Calibration of a fuzzy cellular automata model of urban dynamics in Saudi Arabia. *Ecological Complexity*, 6(2). 80-101.
- Almeida, C. M., Monteiro, A., CaMara, G., Silveira, B., Soares-Filho, B. S., Cerqueira, G. C., Pennachin, C. and Batty, M., 2005. GIS and remote sensing as tools for the simulation of urban land-use change. *International Journal of Remote Sensing*, 26(4). 759-774.
- Arendt, R., 2004. Linked landscapes: Creating greenway corridors through conservation subdivision design strategies in the northeastern and central United States. *Landscape and Urban Planning* 68(3), 241-269.
- Augustijn-Beckers, E., Flacke, J., and Retsios, B., 2011. Simulating informal settlement growth in Dar es Salaam, Tanzania: An agent-based housing model. *Computers, Environment and Urban Systems*, 35(2), 93-103.
- Baetens, J., Van der Weeen, P., and Baets, B., 2012. Effect of asynchronous updating on the stability of cellular automata. *Chaos, Solitons and Fractals*, 45(4), 383-394.
- Ballestores, F. and Qui, Z., 2012. An integrated parcel-based land use change model using cellular automata and decision tree. *Proceedings of the International Academy of Ecology and Environmental Sciences*, 2(2), 53-69.
- Barredo, J.I., Kasanko, M., McCormick, N., and Lavallo, C., 2003. Modeling dynamic spatial processes: Simulation of urban future scenarios through cellular automata. *Landscape and Urban Planning*, 64(3), 145-160.
- Barredo, J. I., Demicheli, L., Lavallo, C., Kasanko, M. and McCormick, N., 2004. Modeling future urban scenarios in developing countries: An application case study in Lagos, Nigeria. *Environment and Planning B: Planning and Design*, 31(1), 65-84.
- Batty, M., 1997. Cellular automata and urban form: A primer. *Journal of the American Planning Association*, 63(2), 266-274.
- Batty, M., 2005. *Cities and Complexity: Understanding Cities with Cellular Automata, Agent-Based Models, and Fractals*. London, England: The MIT Press.
- Batty, M. and Longley, P. A., 1994. *Fractal Cities: A Geometry of Form and Function*. San Diego, CA: Academic Press.

- Batty, M. and Xie, Y., 1997. Possible urban automata. *Environment and Planning B*, 24(2), 175-192.
- Bhatta, B., 2010. *Analysis of Urban Growth and Sprawl from Remote Sensing Data*. New York: Springer.
- Benenson, I. and Torrens, P.M., 2004. *Geosimulation: Automata-Based Modeling of Urban Phenomena*. London: John Wiley & Sons.
- Benenson, I., Birfur, S. and Kharbash, V., 2006. Geographic automata systems and the OBEUS software for their implementation. In J. Portugali, ed. *Complex Artificial Environments: Simulation, Cognition, and VR in the Study and Planning of Cities*. Berlin: Springer, 137-153.
- Benenson, I., Omer, I., and Hanta, E., 2002. Entity-based modeling of urban residential dynamics: The case of Yaffo, Tel Aviv. *Environment and Planning B: Planning and Design*, 29(4), 491-512.
- Blair, J. and Premus, R., 1987. Major factors in industrial location: A review. *Economic Development Quarterly*, 1(1), 72-85.
- Bothwell, S., Gindroz, R., Lang, R., 1998. Restoring community through traditional neighborhood design: A case study of Diggs town public housing. *Housing Policy Debate*, 9(1), 89-114.
- Boucher, T., 1993. *The process of residential real estate development*. Master's thesis. University of Florida, Gainesville.
- Bowman, T., Thompson, J., Tyndall, J., 2012. Resident, developer, and city staff perceptions of LID and CSD subdivision design approaches. *Landscape and Urban Planning*, 107(1), 43-54.
- Brown, D.G., 2006. Agent-based models, the Earth's Changing Land. In: H. Geist, ed. *An Encyclopedia of Land-use and Land-Cover Change*. Westport CT: Greenwood Publishing Group, 7-13.
- Brown, D., Robinson, D., An, L., Nassauer, J., Zellner, M., Rand, W., Riolo, R., Page, S., Low, B., and Wang, Z., 2008. Exurbia from the bottom-up: Confronting empirical challenges to characterizing a complex system. *Geoforum*, 39(2), 805-818.
- Burgess, E. W., 1925. The growth of city: An introduction to a research project. In: R. Park, E. Burgess and R. McKenzie, eds. *The City*. Chicago: The University of Chicago Press, 47-62.

- Castle, C. and Crooks, A.T., 2006. Principles and concepts of agent-based modeling for developing geospatial simulations. *Working Papers Series, 110*, UCL Center for Advanced Spatial Analysis, University College London, London.
- Cheng, J. and Masser, I., 2003. Modeling urban growth patterns: A multi-scale perspective. *Environment and Planning A*, 35(4). 679-704.
- Christaller, W., 1933. *Central Places in Southern Germany*. New Jersey: Englewood Cliffs.
- Clarke, K. C. and Gaydos, L. J., 1998. Loose coupling a cellular automaton model and GIS: Long-term growth prediction for San Francisco and Washington/Baltimore. *International Journal of Geographical Information Science*, 12(7), 699-714.
- Clarke, K. C., Hoppen, S. and Gaydos, L. J., 1997. A self-modifying cellular automaton model of historical urbanization in the San Francisco bay area. *Environment and Planning B: Planning and Design*, 24(2), 247-261.
- Coffman, L., 2000. *Low-impact development design strategies: An integrated design approach*. Maryland: Prince George's County.
- Congalton, R. and Green, K., 2009. *Assessing the Accuracy of Remotely Sensed Data*. Boca Raton: CRC Press.
- Couclelis, H., 1985. Cellular worlds: A framework for modeling micro-macro dynamics. *Environment and Planning A*, 17(5), 585-596.
- Couclelis, H., 1997. From cellular automata to urban models: New principles for model development and implementation. *Environment and Planning B: Planning and Design*, 24(2), 165-174.
- Crooks, A., 2010. Constructing and implementing an agent-based model of residential segregation through vector GIS. *International Journal of Geographical Information Science*, 24(5), 661-675.
- Crooks, A. T., Castle, C. J. E. and Batty, M., 2008. Key challenges in agent-based modeling for geo-spatial simulation, computers. *Environment and Urban System*, 32(6), 417-430.
- Demetriou, D., Stillwell, J., See, L., 2012. *LandParcels: A module for automated land partitioning*. Working Paper 12/02, University of Leeds, School of Geography, Leeds, United Kingdom.

- Dietzel, C. and Clarke, K. C., 2006. The effect of disaggregating land use categories in cellular automata during model calibration and forecasting. *Computers, Environment and Urban Systems*, 30(1), 78-101.
- Downs, A., 2005. Smart growth: why we discuss it more than we do it. *Journal of the American Planning Association*, 71(4), 367-378.
- Duany, A., Plater-Zyberk, E., Speck, J., 2001. *Suburban nation: The rise of sprawl and the decline of the American dream*. New York: North Point Press.
- Engelen, G., White, R., Uljee, I. and Drazan, P., 34. Using cellular automata for integrated modeling of socio-environmental systems. *Environmental Monitoring and Assessment*, 34(2), 203-214.
- ESRI (Environmental Systems Research Institute). 2012. *Calculation of polygon main angle* [online]. ArcGIS Resource Center. Available from: <http://help.arcgis.com/en/arcgisdesktop/10.0/help/index.html#/007000000028000000> [Accessed December 10, 2012].
- ESRI (Environmental Systems Research Institute), 2013. *ESRI CityEngine* [online]. ArcGIS Resource Center. Available from: www.esri.com/cityengine [Accessed July 10, 2013].
- Flache, A. and Hegselmann, R., 2001. Do irregular grids make a difference? Relaxing the spatial regularity assumption in cellular models of social dynamics. *Journal of Artificial Societies and Social Simulation*, 4(4), <http://jasss.soc.surrey.ac.uk/4/4/6.html>
- Fulton, W., 1996. *The new urbanism: Hope or hype for American communities?* Cambridge, MA: The Lincoln Institute for Land Policy.
- Garcia, A., Sante, I., Crecente, R., and Miranda, D., 2011. An analysis of the effect of the stochastic component of urban cellular automata models. *Computers, Environment and Urban Systems*, 35(4), 289-296.
- Gardner, M., 1970. Mathematical games: The fantastic combinations of John Conway's new solitaire game "life". *Scientific American*, 223, 120-123.
- Harris, C. and Ullman, E., 1945. The nature of cities. *Annals of the American Academy of Political and Social Sciences*, 242(1), 7-17.
- Hoyt, H., 1939. *The Structure and Growth of Residential Neighborhoods in American Cities*. Washington D.C.: Federal Housing Administration, the US Government Printing Office.

- Itami, R., 1994. Simulating spatial dynamics: Cellular automata theory. *Landscape and Urban Planning*, 30(1-2), 27-47.
- Jantz, C. and Goetz, S., 2005. Analysis of scale dependencies in an urban land-use-change model. *International Journal of Geographical Information Science*, 19(2), 217-241.
- Jantz, C. A., Goetz, S. J., Donato, D. and Claggett, P., 2010. Designing and implementing a regional urban modeling system using the SLEUTH cellular urban model. *Computers, Environment and Urban Systems*, 34(1), 1-16.
- Jjumba, A. and Dragičević, S., 2012. High resolution urban land-use change modeling: Agent iCity approach. *Applied Spatial Analysis and Policy*, 5(4), 291-315.
- Johnston, K., 2013. *Agent-Analyst: Agent-based modeling in ArcGIS*. Redlands, California: ESRI Press.
- Kaplan, D., Wheeler, J., and Holloway, S., 2008. *Urban Geography*. York, PA: John Wiley & sons.
- Kim, J.H., 2013. Spatiotemporal scale dependency and other sensitivities in dynamic land-use change simulations. *International Journal of Geographical Information Science*, 18(2), 1-22.
- Kocabas, V. and Dragicevic, S., 2006. Assessing cellular automata model behavior using a sensitivity analysis approach. *Computers, Environment and Urban Systems*, 30(6), 921-953.
- Kocabas, V. and Dragičević, S., 2009. Agent-based model validation using Bayesian networks and vector spatial data. *Environment and Planning B: Planning and Design*, 36(5), 787-801.
- Kone, D., 2006. *Land Development*. Washington DC: BuilderBooks, National Association of Home Builders.
- Levinson, D. and Krizek, K., 2008. *Planning for Place and Plexus: Metropolitan Land Use and Transport*. New York: Routledge.
- Li, X. and Liu, X., 2007. Defining agents' behaviors to simulate complex residential development using Multi-criteria Evaluation. *Journal of Environmental Management*, 85(4), 1063-1075.
- Li, X. and Liu, X., 2008. Embedding sustainable development strategies in agent-based models for use as a planning tool. *International Journal of Geographic Information Science*, 22(1), 21-45.

- Li, X., Chen, Y., Liu, X., Li, D. and He, J., 2011. Concepts, methodologies, and tools of an integrated geographical simulation and optimization system. *International Journal of Geographical Information Science*, 25(4), 633-655.
- Ligtenberg, A., Bregt, A. and van Lammeren, R., 2001. Multi-actor-based land use mode spatial planning using agents. *Landscape and Urban Planning*, 56(1), 21-33.
- Liu, Y. and Phinn, S. R., 2003. Modeling urban development with cellular automata incorporating fuzzy-set approaches. *Computers Environment and Urban Systems*, 27(6), 637-658.
- Liu, X., Li, X., Liu, L., He, J. and Ai, B., 2008. A bottom-up approach to discover transition rules of cellular automata using ant intelligence. *International Journal of Geographical Information Science*, 22(11-12), 1247-1269.
- Lösch, 1954. *The Economics of Location*, 2nd edition. Translated into English by W. H. Woglom and W. F. Stolper. New Haven: Yale University Press.
- Loibl, W. and Toetzer, T., 2003. Modeling growth and densification processes in suburban regions—simulation of landscape transition with spatial agents. *Environmental Modeling Software*, 18(6), 553-563.
- Macal, C. M. and North, M., 2008. Agent-based modeling and simulation: ABMS examples. In: S. J. Mason, R. R. Hill, L. Monch, O. Rose, T. Jefferson, J. Fowler, eds. *Proceedings of the 2008 Winter Simulation Conference*. Argonne: Center for Complex Adaptive Systems Simulation, Decision and Information Sciences Division, Argonne National Laboratory.
- Menard, A. and Marceau, D., 2005. Exploration of spatial scale sensitivity in geographic cellular automata. *Environment and Planning B – Planning and Design*, 32(5), 693-714.
- Moreno, N., 2008. *A vector-based geographical cellular automata model to mitigate scale sensitivity and to allow objects' geometric transformation*. PhD Dissertation. Department of Geomatics Engineering, University of Calgary, Alberta.
- Moreno, N., Menard, A., and Marceau, D., 2008. VecGCA: a vector-based geographic cellular automata model allowing geometric transformations of objects. *Environment and Planning B: Planning and Design*, 35(4), 647-665.
- Moreno, N., Wang, F., and Marceau, D., 2009. Implementation of a dynamic neighborhood in a land-use vector-based cellular automata model. *Computers, Environment and Urban Systems*, 33(1), 44-54.

- Nechyba, T. and Walsh, R., 2004. Urban sprawl. *The Journal of Economic Perspectives*, 18(4), 177-200.
- Oguz, H., Klein, A. and Srinivasan, R., 2007. Using the SLEUTH urban growth model to simulate the impacts of future policy scenarios on urban land use in the Houston-Galveston-Brazoria CMSA. *Research Journal of Social Sciences*, 2(1), 72-82.
- O'Sullivan, D., 2001. Exploring spatial process dynamics using irregular cellular automaton models. *Geographical Analysis*, 33(1), 1-18.
- O'Sullivan, D. and Torrens, P., 2000. Cellular models of urban systems. *Working Papers Series*, 22, UCL Center for Advanced Spatial Analysis, University College London, London.
- Pacione, M., 2005. *Urban Geography: A Global Perspective*. London: Rutledge.
- Pan, Y., Roth, A., Yu, Z., and Doluschitz, R., 2010. The impact of variation in scale on the behavior of a cellular automata used for land use change modeling. *Computers, Environment and Urban Systems*, 34(5), 400-408.
- Parker, D.C., Manson, S.M., Janssen, M.A., Hoffmann, M.J., and Deadman, P.J., 2003. Multi-agent systems for the simulation of land-use and land-cover change: A review. *Annals of the Association of American Geographers*, 93(2), 314-337.
- Portugali, J., 2000. *Self-Organization and the City*. Berlin: Springer-Verlag.
- Samat, N., 2006. Characterizing the scale sensitivity of the cellular automata simulated urban growth: A case study of the Seberang Perai Region, Penang State, Malaysia. *Computers, Environment and Urban Systems*, 30(6), 905-920.
- Sante, I., Garcia, A., Miranda, D., Crecente, R., 2010. Cellular automata models for the simulation of real-world urban processes: A review and analysis. *Landscape and Urban Planning*, 96, 108-122.
- Schmitz, A., 2004. *Residential Development Handbook*. Washington DC: Urban Land Institute.
- Shi, W. and Pang M., 2000. Development of Voronoi-based cellular automata – an integrated dynamic model for geographical information systems. *International Journal of Geographical Information Science*, 14(5), 455-474.
- Simon, H. A., 1962. The architecture of complexity. *Proceedings of the American Philosophical Society*, 106(6), 467-482.

- Stevens, D. and Dragičević, S., 2007. A GIS-based irregular cellular automata model of land-use change. *Environment and Planning B: Planning and Design*, 34(4), 708-724.
- Stevens, D., Dragičević, S. and Rothley, K., 2007. iCity: A GIS-CA modeling tool for urban planning and decision making. *Environmental Modeling & Software*, 22(6), 761-773.
- Sudhira, H., Ramachandra, T., Wytzisk, A. and Jagger, K., 2005. *Framework for integration of agent-based and cellular automata models for dynamic geospatial simulations*. Bangalore: Centre for Ecological Sciences, Indian Institute of Science.
- Thapa, R. B. and Murayama, Y., 2011. Urban growth modeling of Kathmandu metropolitan region, Nepal. *Computers, Environment and Urban Systems*, 35(1), 25-34.
- Tian, G., Ouyang, Y., Quan, Q., Wu, J., 2011. Simulating spatiotemporal dynamics of urbanization with multi-agent systems - a case study of Phoenix metropolitan region, USA. *Ecology and Society*, 22(5), 1129-1138.
- Tobler, W. R., 1979. Cellular geography. In: S. Bale and G. Ollson, eds. *Philosophy in Geography*. Dordrecht: Kluwer, 379-386.
- Torrens, P. M., 2001. New tools for simulating housing choices. *Institute of Business and Economic Research and Fisher Center for Real Estate and Urban Economics, Program on Housing and Urban Policy Conference Paper Series (C01-006)*. Berkeley: University of California.
- Torrens, P. M. and Benenson, I., 2005. Geographic automata systems. *International Journal of Geographic Information Science*, 19(4), 385-412.
- Torrens, P. M. and Nara, A., 2007. Modeling gentrification dynamics: A hybrid approach. *Computers, Environment and Urban Systems*, 31(3), 337-361.
- United Nations (UN), 2005. *World Urbanization Prospects: The 2005 Revision*. New York: Population Division, Department of Economic and Social Affairs, UN.
- United Nations (UN), 2010. *World Urbanization Prospects: The 2009 Revision*. New York: Population Division, United Nations.
- US Census (United States Census Bureau), 2000. *Census 2000 Summary File1 (SF-1)* [online]. Census 2000 Gateway. Available from: <http://www.census.gov/census2000/sumfile1.html> [Accessed July 12, 2012].
- US Census (United States Census Bureau), 2012. *State and County QuickFacts: Data derived from Population Estimates* [online]. American Community Survey, Census of

- Population and Housing, County Business Patterns, Economic Census, Survey of Business Owners, Building Permits, Census of Governments. Available from: <http://quickfacts.census.gov/qfd/states/48/4865600.html> [Accessed July 12, 2012].
- Vanegas, C., Aliaga, D.G., Benes, B., and Waddell, P., 2009. Visualization of simulated urban spaces: Inferring parameterized generation of streets, parcels, and aerial imagery. *IEEE Transactions on Visualization and Computer Graphics*, 15(3), 424-435.
- Vanegas, C., Aliaga, D., Benes, B., Waddell, P., 2009 (b). Interactive design of urban spaces using geometrical and behavioral modeling. *ACM Transactions on Graphics*, 28, 111(1)-111(11).
- Vanegas, C., Kelly, T., Weber, B., Halatsch, J., Aliaga, D., Muller, P., 2012. Procedural generation of parcels in urban modeling. *Eurographics*, 31(2), 681-690.
- Vaughan, J., 2008. *Growth and change in a paradigmatic region: Is it sustainable? Does planning make a difference?* Ph. D. Dissertation. Texas State University-San Marcos.
- Vliet, J., White, R., and Dragicevic, S., 2009. Modeling urban growth using a variable grid cellular automaton. *Computers, Environment and Urban Systems*, 33(1), 35-43.
- Von Neumann, J., 1951. The general and logical theory of automata. In: L. A. Jeffress, ed. *Cerebral Mechanisms in Behavior: The Hixon Symposium*. London: John Wiley & Sons, 1-31.
- Von Thunen, J. H., 1826. *Der Isolierte Staaat in Beziehung auf Landschaft und Nationalökonomie, Hamburg* (English translation by C.M. Wartenburg). London: Oxford University Press.
- Waddell, P., Wang, L., Charlton, B., and Olsen, A., 2010. Micro-simulating parcel-level land use and activity-based travel: Development of a prototype application in San Francisco. *The Journal of Transport and Land Use* 3(2), 65-84.
- Wakchaure, A.S., 2001. *An ArcView tool for simulating land subdivision for build out analysis*. Master's thesis. Polytechnic Institute and State University, Virginia.
- Wang, F., Hasbani, J.G., Wang, X., and Marceau, D., 2011. Identifying dominant factors for the calibration of a land-use cellular automata model using Rough Set Theory. *Computers, Environment and Urban Systems*, 35(2), 116-125.
- White, R. and Engelen, G., 1997. Cellular automata as the basis of integrated dynamic regional modeling. *Environment and Planning B: Planning and Design*, 24(2), 235-246.

- White, R. and Engelen, G., 2000. High-resolution integrated modeling of the spatial dynamics of urban and regional systems. *Computers, Environment and Urban Systems*, 24(5), 383-400.
- White, R., Engelen, G., Uljee, I., 1997. The use of constrained cellular automata for high-resolution modeling of urban land-use dynamics. *Environment and Planning B: Planning and Design*, 24(3), 323-343.
- Wickramasuriya, R., Chisholm, L., Puotinen, M., Gill, N., Klepeis, P., 2011. An automated land subdivision tool for urban and regional planning: Concepts, implementation and testing. *Environmental Modeling & Software*, 26(12), 1675-1684.
- Wolfram, S., 1984. Cellular automata as models of complexity. *Nature*, 311(4), 419-424.
- Wu, F., 1998. Simulating urban encroachment on rural land with fuzzy-logic-controlled cellular automata in a geographical information system. *Journal of Environmental Management*, 53(4), 293-308.
- Wu, F., 1998. Simulation of land development through the integration of cellular automata and multi-criteria evaluation. *Environment and Planning B: Planning & Design*, 25(1), 103-126.
- Wu, N. and Silva, E., 2010. Artificial intelligence solutions for urban land dynamics: A review. *Journal of Planning Literature*, 24(3), 246-265.
- Yang, X. and Lo, C. P., 2003. Modelling urban growth and landscape changes in the Atlanta metropolitan area. *International Journal of Geographical Information Science*, 17(5), 463-488.
- Yang, Q., Li, X. and Shi, X., 2008. Cellular automata for simulating land use changes based on support vector machines. *Computers & Geosciences*, 34(6), 592-602.
- Yeh, A. G. and Li, X., 2003. Simulation of development alternatives using neural networks, cellular automata, and GIS for urban planning. *Photogrammetric Engineering and Remote Sensing*, 69(9), 1043-1052.
- Zellner, M., Page, S., Rand, W., Brown, D., Robinson, D. N., J., and Low, B., 2009. The emergence of zoning policy games in exurban jurisdictions: Informing collective action theory. *Land use Policy*, 26(2), 356-367.
- Zhang, H., Zeng, Y., Bian, L. and Yu, X., 2010. Modeling urban expansion using a multi agent-based model in the city of Changsha. *Journal of Geographical Sciences*, 20(4), 540-556.