LARGE-SCALE SPATIOTEMPORAL MODELING OF URBAN GROWTH WITH CYBERINFRASTRUCTURE: A SURROGATE-BASED APPROACH

by

Wenpeng Feng

A dissertation submitted to the faculty of
The University of North Carolina at Charlotte
in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in
Geography and Urban Regional Analysis

Charlotte

2017

Approved by:

____
Dr. Wenwu Tang

____
Dr. Eric Delmelle

____
Dr. Harrison Campbell

____
Dr. David K. Pugalee

ABSTRACT

WENPENG FENG. Large-scale spatiotemporal modeling of urban growth with cyberinfrastructure: a surrogate-based approach. (Under the direction of Dr. Wenwu Tang)

Spatiotemporal simulations can provide critical insights to understand the underlying mechanisms of complex geographic phenomena. Therefore, spatiotemporal simulations play a vitally important role in solving the global geographic problems such as habitats loss, climate change, and deforestation. Due to complex mechanisms and big spatial data, computational intensity greatly hinders the application of spatiotemporal simulations at large scale. Cyberinfrastructure has been recognized as a promising tool to tackle computational intensity in spatiotemporal simulations. However, a challenge lies in the accurate estimation of its computing performance, which may prevent an efficient utilization of cyberinfrastructure. This dissertation demonstrates a surrogate-based approach to appropriately estimate the computing performance of parallel spatiotemporal simulations within cyberinfrastructure environments. A generalized computational framework is developed to integrate surrogate-based models, spatiotemporal simulations, and cyberinfrastructure. I applied the computational framework to simulate urban growth in North Carolina as a case study. Results show that surrogate-based approaches accurately estimate the computing performance. Kriging has a better prediction performance than linear regression surrogate-based model in this study. With the support

of surrogate-based approaches, the computational framework substantially supports spatiotemporal simulations by efficiently handling computational intensity.

DEDICATION

To My Parents Mr. Guoliang Feng and Mrs. Xiurui Li, and My wife, Mrs. Yuqi Chen for their endless love and support.

ACKNOWLEDGEMENTS

I would not complete this dissertation without the support and guidance from my advisors, friends, and family. I deeply appreciate their contributions to this dissertation.

First, I must sincerely thank my advisor, Dr. Wenwu Tang, for his endless support during my PhD study in Geography. He introduced me to the world of geocomputation and spatiotemporal modeling which serve the foundation of this dissertation and my career. More important, his enthusiasm encourages me to face challenges even in stressful academic and life situations. I also am grateful to Dr. Eric Delmelle, who is my first advisor and opens the door of GIS and spatial optimization for me. He always encourages me when I am not confident. I gratefully acknowledge Dr. Harrison Campbell for his generous help and guidance in terms of geography theories. I would like to thank Dr. David K. Pugalee for his valuable time to improve my dissertation.

I own many thanks to my friends who have helped me through the past few years. In particular, I would like to express my gratitude to all professors and my study companies in our departments.

Last, but not least I would like to thank my parents, Mr. Guoliang Feng and Mrs. Xiurui Li, and my wife, Mrs Yuqi Chen. They give me unconditional, endless love and support, which encourage me to complete this dissertation.

TABLE OF CONTENTS

LIST OF TABLES x

LIST OF FIGURES xi

LIST OF ACRONYMS xiii

CHAPTER 1: INTRODUCTION 1

1.1 Background 1

1.2 Research objectives 4

1.3 Road map 6

CHAPTER 2: LITERATURE REVIEW 8

2.1 Coupled human and natural systems 8

2.2 Spatial simulation models 13

2.2.1 Markov models 13

2.2.2 System models 17

2.2.3 Cellular automata models 21

2.2.4 Agent-based models 27

2.3 CyberGIS 34

2.4 Parallel spatial simulation model 36

2.4.1 Evaluation of parallel computing performance 36

2.4.2 Parallel computing architectures and methods 38

2.4.3 Parallel strategies related to parallel spatial modeling 46

2.5 Summary 54

CHAPTER 3: SURROGATE-BASED MODEL..58

3.1 Background of surrogate-based models... 58

3.2 Components of surrogate-based models 63

3.3 Significance of surrogate-based models 68

3.4 Designs for spatiotemporal simulations... 69

CHAPTER 4: COMPUTATIONAL FRAMEWORK...74

4.1 Research objective 1 methodology 74

4.1.1 Computational framework of spatiotemporal simulation 74

4.2 Research objective 2 methodology 76

4.2.1 Parallel computing 76

4.2.2 Scientific workflow middleware... 78

4.2.3 Automation 79

4.3 Research objective 3 methodology 81

4.3.1 ABM urban growth simulation 81

CHAPTER 5: CASE STUDY..87

5.1 Study area and data 87

5.2 Parallel calibration of urban growth simulation.. 89

5.3 Implementation 92

CHAPTER 6: EXPERIMENTS..93

6.1. Hypotheses 93

6.2. Experiments 95

6.2.1 Relationship between computing performance and spatial characteristics/content 97

6.2.2 Impacts of sample size and types of surrogate-based approach 104

6.2.3 Impact of surrogate-based approach on computing performance....................... 111

6.3 Concluding Discussion 116

CHAPTER 7: CONCLUSIONS 119

REFERENCES 124

LIST OF TABLES

Table 1. Table of data used in this dissertation 89

Table 2. Correlation of average computing time and model variables 102

Table 3. Average root-mean-square errors of surrogate-based models for different sample size 109

Table 4. *P*-value of *t*-test for different sample size 110

Table 5. Estimated and real computing performance of parallel urban growth model calibration 114

Table 6. Comparison of computing performance of two treatment groups (Group 1: without surrogate-based load balancing; Group 2: with surrogate-based load balancing) 114

LIST OF FIGURES

Figure 1. Number of surrogate-based modeling articles published by year 60

Figure 2. Research area of surrogate-based modeling 61

Figure 3. Components of surrogate-based modeling 64

Figure 4. Sequential procedure of surrogate-based modeling 71

Figure 5. Surrogate-based computational framework of parallel spatiotemporal simulations 75

Figure 6. Workflow of spatial sampling and validation 80

Figure 7. An example of sampling table 80

Figure 8. Agent-based modeling integrated urban growth model 82

Figure 9. Map of North Carolina 88

Figure 10. Workflow of urban growth model calibration 90

Figure 11. Urban percentage of NC for year 2001 99

Figure 12. Calibrated discount factor of each county in NC 99

Figure 13. Calibrated patch mean of each county in NC 100

Figure 14. Calibrated patch range of each county in NC 100

Figure 15. Simulated urban growth in NC for the year 2001 101

Figure 16. Scatter plot of computing time and model variables of urban growth (A: number of cells; B: urban percentage; C: demand) 103

Figure 17. An example of response surface of computing time in Kriging surrogate-based model.. 107

Figure 18. Average root-mean-square error of surrogate-based models using different sample size 110

Figure 19. Average root-mean-square errors for linear regression and Kriging surrogate-based models 111

Figure 20. Plot of real and estimated speedup of parallel urban growth model calibration 115

Figure 21. Plot of speedup of with surrogate-based model group (blue line) and without surrogate-based model group (red line) 115

Figure 22. Plot of efficiency of with surrogate model group (blue line) and without surrogate model group (red line) 116

LIST OF ACRONYMS

GIS	Geographic information system
ABM	Agent-based modeling
CHANS	Coupled human and natural systems
CPU	Central processing units
SOA	Service-oriented architecture
GPU	Graphics processing units
OpenMP	Open Multi-Processing
NLCD	National land cover database
MPI	Message passing interface

CHAPTER 1: INTRODUCTION

1.1 Background

Geographic information system (GIS) has been developed to acquire, manage, analyze, and visualize geographic data, exemplified by remote sensing and census data, to solve complex geographic problems such as deforestation and climate change (Goodchild, 1992; Clarke and Gaydos, 1998; Goodchild, 2001; Weber and Puissant, 2003; Goodchild, 2004). Taking advantage of GIS, more and more scholars in geographic information science (GIScience) have developed different types of spatiotemporal models, including statistics, optimization, and simulation to investigate complex geographic phenomena (Armstrong, 2000; Goodchild, 2003; Openshaw, 2004; Rindfuss et al., 2004; Turner et al., 2008).

Spatiotemporal simulation models have been developed to help researchers understand the underlying mechanisms of complex geographic phenomena such as urban growth, habitats loss, and sea-level rise (Gilbert and Troitzsch, 2005; Parker et al., 2008; O'Sullivan and Perry, 2013). Urban growth has received increasing attention from researchers in a series of scientific domains such as geography, public policy, urban planning, and ecology (Clarke and Gaydos, 1998; Waddell, 2002; Herold et al., 2003; Weber and Puissant, 2003; Pijanowski, 2014). Attributed to interactions among spatial and non-spatial factors and processes operating at different temporal and spatial scales, research on urban growth plays a critical role in solving global geographic problems such as habitat loss, climate change, sea-level rise, and greenhouse gas emission (Foley et al., 2005; Verburg et al., 2009). Recognized as a complex problem, urbanization substantially impacts social-ecological systems by decreasing and fragmentizing non-urban areas with a variety of land cover types (Batty, 1991; Batty and Xie, 1994; Batty et al., 1999).

Spatiotemporal simulation models represent processes driving land use transitions (e.g., from open space to urban, or from agriculture to urban) of urban growth over time. Taking advantage of spatiotemporal simulation models, researchers can tap into spatial variables of interest and their interrelationships in their study regions. Furthermore, decision alternatives and their consequences can be represented, analyzed and visualized to support the decision making process of policy makers using what-if scenario analysis (O'Sullivan and Perry, 2013).

As a representative of spatiotemporal simulation models, agent-based modeling (ABM) has been increasingly recognized as an appropriate approach to simulate urban growth (Batty and Longley, 1994; Epstein and Axtell, 1996; Epstein, 1999; Batty et al., 2012). Compared with other spatiotemporal simulation models, ABM is more suitable to investigate decision making processes at the individual level in a complex system. ABM allows to investigate impacts from individual behavior to the spatial patterns of land cover change at an aggregated level (Tang et al., 2011a; Tang et al., 2011b; Tang and Bennett, 2012; O'Sullivan, 2008). However, ABMs always require a large amount of data and computation, especially at a large spatial scale (e.g., state level urban growth simulations). This results in unaffordable computational costs (e.g. computer memory, data storage, and computing speed) out of the reach of traditional desktop computers (Armstrong, 2000).

In studies using spatiotemporal models to simulate urban growth at a large scale with fine resolution, the degree of complexity of such models can dramatically increase in multiple ways (Marurngsith, 2014). First, the number of individuals involved in urban growth (e.g. farmers and residents) can be very large. Second, spatial impacts from urban growth affect a large area. Third, underlying urban growth mechanisms and driving forces do not follow linear and constant trends; rather, they are dynamic and non-linear (Golledge and Stimson, 1997; Gahegan, 2003). Fourth, empirical data used to simulate urban growth are evolving to big data in terms of volume, variety, and velocity (Manyika et al., 2011; Zikopoulos and Eaton, 2011). For these reasons, tackling model complexity of urban growth simulations at large scale requires advanced computational capabilities. Consisting of high-performance and parallel computing, massive data handling, and virtual organization, cyberinfrastructure is a promising tool to address this bottleneck (NSF, 2007). Besides reduction in the computing time of large-scale urban growth simulations, cyberinfrastructure is capable of improving and validating the simulation results with computationally intensive methods (e.g. Monte Carlo tests). Also, better statistical analysis can be conducted based on handling large size or fine resolution empirical data. Furthermore, new theories and hypotheses related to urban growth can be discovered and tested, taking advantage of computational progress (Openshaw and Turton, 2005).

Despite these benefits, it is recognized that the expertise and steep learning curve associated with the use of high performance and parallel computing hinder its use and development for non-programmers (Parry and Bithell, 2012; Scheutz and Harris, 2012). Moreover, appropriate estimation of the computing performance of spatiotemporal simulations plays a critical role in the efficient utilization of cyberinfrastructure. However, it is extremely challenging to predict computing performance, since underlying mechanisms of spatiotemporal models are complex (Wang and Armstrong, 2009). Consequently, there exists a gap between computation methods and spatiotemporal simulations, which limits the use of cyberinfrastructure in studies of urban growth. The challenge thus lies in how to efficiently leverage the cyberinfrastructure to benefit the studies of urban growth.

A surrogate-based model is a black-box approach that simulates the relationship between model input and output in an approximate way (Forester et al., 2008). Surrogate-based models have been widely applied in many scientific domains, such as engineering and computer science, to overcome computational challenge for high-fidelity simulations (Queipo et al., 2005; Forester et al., 2008; Kleijnen, 2009; Forrester and Keane, 2009). In order to fill the gap between computation methods and spatiotemporal simulations, I develop a surrogate-based computational framework to support the efficient utilization of cyberinfrastructure in spatiotemporal simulations.

1.2 Research objectives

The goal of this research is to investigate the (more) efficient utilization of cyberinfrastructure in a large-scale spatiotemporal simulation of urban growth using surrogate-based approaches. I integrate surrogate-based modeling, spatiotemporal simulations, and cyberinfrastructure together to solve complex geographic problems. The surrogate-based modeling can help researchers without computing background to predict the computing performance. Given these estimated computing performance, researchers can better design their research and budget (time and funding) within cyberinfrastructure environments. In spatial analysis, arbitrary assumptions (such as homogeneous spatial characteristics) are usually used to reduce computational and data burden. With the support of cyberinfrastructure, these assumptions can be greatly removed by leveraging high performance computing and big data handling capabilities. Furthermore, novel insights related to urban growth can be discovered by comparing the results between simulations with fine spatiotemporal resolution and coarse-grained approaches in the context of urban growth.

Based on the research goal, there are three specific objectives of this work:

a) Develop a surrogate-based computational framework to identify and address key computational challenges to efficiently handle computational intensity in large-scale spatiotemporal simulations.

b) Investigate how to realize the computational tractability, and make the computational framework as simple as possible for researchers without computing background in the context of urban growth.

c) Apply the proposed computational framework to support a large-scale parallel urban growth simulation, and evaluate its computing performance.

In alignment with existing methodologies, this research provides an alternative to facilitate the spatiotemporal simulation of urban growth. With support of the proposed computational framework, scholars can further investigate how human decision making process at individual level can impact the urban growth at aggregate level. Computational issues due to increased complexity at large scale can be efficiently tackled with our computational framework by employing high performance and parallel computing methods. Furthermore, taking into account impacts from human decisions, our framework can provide solid support to explore massive parameter space based on alternative scenarios of future urban plans and policies within an affordable computing time. As a case study, I apply the proposed computational framework to simulate urban growth from 1992-2001 for North Carolina.

1.3 Road map

The rest of my dissertation will be organized as follows: Chapter 2 provides a thorough literature review on four main parts: (1) the theory of coupled human and natural systems, (2) mainstream spatial simulation modeling in geography, (3) concept and application of cyberGIS, and (4) existing research on parallel spatial simulation modeling. Chapter 3 focuses on introducing surrogate-based models. It first reviews existing work related to the application of surrogate-based models in other scientist domains. Then, the main components and processes of building a surrogate-based model are presented. With respect to the previous literature review, a surrogated-based computational framework is proposed in Chapter 4 to support spatiotemporal simulations within cyberinfrastructure environments. Main framework and corresponding methodologies for three specific research objectives are illustrated in this chapter. In Chapter 5, I introduce a case study of a spatiotemporal simulation of urban growth in North Carolina. In addition, a parallel method is designed for model calibration.

Based on the research objective, Chapter 6 selects three hypotheses to examine the utility of surrogate-based approaches in the estimation of computing performance within parallel spatiotemporal simulations. Correspondingly, three experiments are designed to test three hypotheses: (1) Computing intensity will be correlated with spatial characteristics/content in spatiotemporal simulations. (2) Sample size and the type of surrogate-based approaches will impact the prediction ability for computing intensity of spatiotemporal simulations. (3) The application of surrogate-based approaches will improve the computing performance of parallel spatiotemporal simulations. The computing performance of our proposed surrogate-based computational framework is evaluated in experiments as well. Chapter 7 concludes this dissertation by emphasizing contributions and significances of my work, and presents the future research directions.

CHAPTER 2: LITERATURE REVIEW

This chapter conducts the literature review of existing study in terms of spatiotemporal simulations and cyberinfrastructure. I first introduce the concept of coupled human and natural systems. Then, I demonstrate mainstream approaches to simulate urban growth. At last, development and application of high performance and parallel computing are discussed.

2.1 Coupled human and natural systems

Urban growth can be studied as a complex adaptive system in which agents interact with each other and adapt environment (Batty, 2007). A complex system has characteristics of self-organization, nonlinearity, emergency, path-dependence, and adaption (Manson, 2001; Lansing, 2003). Complexity theory is developed based on general systems theory which has been applied in scientific domains such as geography, physics, biology and social sciences (Bertalanffy, 1968; Warren et al., 1998; An et al., 2005). The main objective of complexity theory is to understand the underlying mechanisms of complex adaptive systems (Holland, 1992, 1995, 2000). A set of heterogeneous subsystems or entities comprise a complex adaptive system by interacting with each other. These subsystems or entities can improve their situations within the complex adaptive system by leaning or adapting from interactions (Holland 2006). With respect to these characteristics, it is impossible to fully predict or control a complex adaptive system, since the complexity could be introduced in many forms (Solé and Goodwin, 2008). However, complexity theory can help researchers better understand and improve the complex adaptive system by means of investigating the nonlinear relationship and interactions (for example feedback loops) among entities (or subsystems) in a complex adaptive system (Manson, 2001;Crawford et al., 2005).

The traditional land cover modeling is based on determinist and aggregated top-down approaches, for example Markov models (Drewett, 1969; Bourne, 1971; Bell, 1974; Bell and Hinojosa, 1977; Robinson, 1978; Jahan, 1986; Muller and Middleton, 1994) and the system dynamic models (Forrester, 1969; Randers, 1980; Richardson and Pugh, 1981; Han et al., 2009; Liu et al., 2013). Undoubtedly, these modeling approaches can shed light on the relationships between urban development and other ecological systems, for example the forest system, in a holistic view. However, from the aspect of complexity theory, these traditional land cover modeling approaches have hindered the discovery of new complexity patterns and underlying mechanisms, as well as interactions among sub-systems within a complex adaptive systems (An et al., 2005; Liu et al., 2007; An and Liu, 2010). Based on these traditional methods, simulation usually concentrates on land use and land cover changes at an aggregated level, without explicitly considering the impacts from human decision making at an individual level. Thus, it is hard to gain important insights about the linkage between individual behaviors (for example the land development from real estate agents) at the micro-level and their consequences in the complex adaptive system. Moreover, the limitation of synthetic interdisciplinary analysis in traditional land cover change modeling approaches makes them weak to simulate and understand the characteristics of complexity, such as the nonlinearity and thresholds, within urban growth process.

CHANS (Coupled human and natural systems) is a modeling approach which integrates human decision making and natural systems together. It has been increasingly applied to examine the relationship between human society and ecological systems (An et al., 2005; Liu et al., 2007; An and Liu, 2010). In comparison with other traditional modeling approaches, the greatest advantage of CHANS is the synthetic analysis for both socioeconomic behaviors and ecological impacts, instead of excluding the human decision making processes or ecological services in the simulation process. Summarized by Liu et al., (2007), there are six characteristics in CHANS as follows: reciprocal effects and feedback loops, nonlinearity and thresholds, surprises, legacy effects and time lags, and heterogeneity. Based on these characteristics, both human social components and ecological components are taken into account, as well as the relationship between these two components within complex adaptive systems. In addition, these characteristics of CHANS can substantially facilitate the resolution of geographic problems incorporating the impacts from human decision making processes by means of mimicking individual behaviors at the micro-level (An et al., 2005; Liu et al., 2007; An and Liu, 2010).

Due to advances in complexity theory and modeling techniques, a series of CHANS have been implemented to solve real-world problems based on empirical data. Two bottom-up modeling approaches, cellular automata and ABM (agent-based modeling), have been widely coupled within CHANS (Parker et al., 2003; An et al., 2005; Matthews et al., 2007; Liu et al., 2007; Verburg and Overmars, 2009; An and Liu, 2010).

Since the cellular automata models have strength on representing the spatial dimension and spatial interactions of a system, a lot of CHANS have been developed by integrating the cellular automata approach for landscape modeling. Recently, a myriad of scholars have devoted to use ABM to mimic the human decision making process in CHANS (Deadman and Gimblett, 1994; Epstein and Axtell, 1996; Railsback, 2001; Gimblett, 2002; Parker et al., 2003; Bousquet and Le Page, 2004; Evans and Kelley, 2004; Deadman et al., 2004; An et al., 2005; Brown et al., 2005; Evans et al., 2006; Brown and Robinson, 2006; Manson, 2006; Bennett and Tang, 2006; Matthews et al., 2007; An, 2012) .

The underlying process can be better explored with ABM by comparing the simulation patterns with the observed patterns (Goodchild, 2004). Taking advantage of integration with ABM, CHANS is capable of simulating the decision making process for each individual agent or agent group in a complex adaptive system (Gimblett, 2002; An, 2012). With support of ABM, complex spatial patterns and processes can be explored directly at an individual level instead of operating as a deterministic or stationary simulations at the aggregate level based on rigid assumptions, for example their linear relationship. Also, ABM enables CHANS the capability of investigating the relationships of agent-agent, agent-system, as well as system-system interactions. Furthermore, interactions among heterogeneous agents can be modeled at the micro-level to investigate the causes and consequences in complex adaptive systems, according to a set of decision rules of agents. In the work of Li et al. (2015), the authors developed an analytical framework to the complex human-nature interaction which drives the land use change and urban growth in eastern China. The complex adaptive behaviors of individual agents are simulated in three stages: agent learning, agent decision making, and agent action.

With respect to the hydrologic domain, Hu et al. (2015) have applied CHANS to simulate water conflict between the requirement of ecologic system and the demand of human irrigation within the Republican River basin (RRB) in the U.S. Midwest. The CHANS model consists of a multi-agent irrigation decision making system model and the Republican River Compact Administration groundwater model to investigate how the human decision making process can impact the water use in study area. The decision rule of agents is to maximize their individual utilities, which can be updated through the Bayesian learning process that incorporates the prior knowledge and observations using Bayesian statistics. A large number of interactions exist between agents and their environment, which give rise to significant computational intensity. In order to handle the computational intensity, Hu et al. (2015) have applied multithreaded programming to increase the computational efficiency, in which a single CPU (central processing unit) executes multiple computing threads/tasks at the same time. Moreover, a thread safe program has been designed to make sure the update of data in shared memory is correct. With the support of multithread programming, computation time was greatly decreased from one hour with sequential run to twelve minutes in parallel version running on eight CPU. In addition, to facilitate the accessibility for public use, Hu et al. (2015) have coupled the model with the SOA (service-oriented architecture) architecture implemented with a Hadoop-based cloud computing environment. Load-balancing and task scheduling have been realized by taking advantage of MapReduce in Hadoop.

2.2 Spatial simulation models

According to the inherent structure and mechanism, spatial simulation models could be implemented with support of two main approaches: top-down and bottom-up approaches. Markov and system dynamic models are representatives of top-down spatial simulation models, while bottom-up spatial simulation models primarily include cellular automata models and agent-based models.

2.2.1 Markov models

Geographers have extensively applied Markov models to study and analyze the dynamical land use and land cove change (Drewett, 1969; Bourne, 1971; Bell, 1974; Bell and Hinojosa, 1977; Robinson, 1978; Jahan, 1986; Muller and Middleton, 1994). The Markov model is a stochastic model which has the capability to mimic the land use and land cover change using discrete state spaces (all possible values that a Markov process can take) and fixed sojourn time (time spent during each step). In the Markov model, there are two important components: transition matrix and transition probability matrix. The transition matrix and transition probability matrix are yield based on empirical data in the study area, which can capture the historical trend of land use and land cover change. The transition matrix and transition probability matrix are used to simulate and predict the future land use and land cover change (Baker, 1989). In most of the applications of Markov model, the state of land cover and land use in next time step only depended on the current state instead of historical steps.

Based on the empirical land cover and land change data, a transition matrix can be built by summarizing the number of converted cells from each land use category to all other categories in study area between two observed time points. With the same order of the transition matrix, the transition probability matrix indicts the percentage of land use change in multiple directions among all the land use categories which are mutually exclusive (Burnham, 1973; Bell, 1974; Turner, 1987; Muller and Middleton, 1994). The transition probability of land use category *i* to shift to category *j* is calculated based on the following equation:

$$P_{ij} = \frac{N_{ij}}{N_i} \quad (1)$$

where P_{ij} implies the percentage of land cells shifting from land use category *i* to category *j* during the observed time period. The term N_{ij} is the number of cells changed, while N_i is total number of cells of land use category *j* at the beginning of the time period under consideration.

The Markov model has been widely applied in land use and land cover change as a simulation and prediction tool. In the work of Burnham (1973), the author presented a Markov land use simulation model to evaluate the alternative land use scenarios and policies based on historical observation data. The study area was located in the Southern Mississippi Alluvial Valley with total land acreage of 24,079,000. The transition matrix was developed based on the converted area of each land use group to all other groups from 1950 to 1969, including six groups: cropland, grassland, transition, forest, urban and other. According to the empirical transition matrix, the transition probabilities were represented with the ratio of transition area from one group shifting to all other groups over the total area of this group at the beginning of the observation time period in the transition probability matrix. Then the transition probability matrix was applied to project the land use change in study area for two time periods: 1969 to 1988 and 1988 to 2007, with the same amount of years with the observation time period. The impacts of four alternative land use scenarios were estimated by means of adjusting the responding probabilities in the transition probability matrix.

Compared with the constant transition probability matrix of Burnham (1973), Turner (1987) developed a spatial simulation model with a dynamic transition probability matrix to investigate land use pattern change in Oglethorpe County, Georgia, including five land use categories. The net land use transition rate was obtained from historical aerial photography. The spatial resolution was 1 ha, and the simulation was conducted for two time periods: 1942 to 1955 and 1955 to 1980. The impacts of spatial influences from alternative spatial simulations, including random simulation, von Neumann neighborhood (four adjacent cells) and Moore neighborhood (eight adjacent and diagonal surrounding cells), were modeled and evaluated using landscape metrics at the patch level (mean number and size, fractal dimension and amount of edges).

With the advances of remote sensing technology and GIS, the Markov model has been increasingly used into large scale land use land change modeling (Muller and Middleton, 1994; Brown et al., 2000; Weng, 2002). Instead of using the actual transition area, data sampling strategies were employed to develop the transition function used to predict the land use change. In the work of Muller and Middleton (1994), sample points were classified into three land use categories (urban, agricultural and natural uses) to extract the transition information for each category based on aerial photography from

1935 to 1981. Brown et al. (2000) selected 136 simple sites following stratified sampling scheme and principal-components analysis. Parcel maps and aerial photography were used to develop an approach in which the relationship between land cover change and land use change was investigated in the Upper Midwest, USA. Aiming at improving the analysis by reducing data uncertainty, Weng (2002) integrated satellite remote sensing, GIS and a Markov model together to simulate and monitor the urbanization in Zhujiang Delta, China, with a total area of 15,112 square kilometers. With the satellite remote sensing and GIS, land use and land cover change were detected and summarized taking advantage of Landsat TM data for the study area. Then, the Markov model was built to simulate and project the land cover change based on the historical trend.

However, there exist native challenges and limitations in the land use and land cover change simulation with a Markov model, due to assumptions applied in the Markov model (Burnham, 1973; Bell, 1974; Turner, 1987). First, according to the assumption of stationary, the transition probability is constant along both temporal and spatial scales. Because the land use and land cover change is closely related to human and natural driving factors (for example, the urban sprawl), it is implausible to use a constant changing rate in a long simulation time period or large study area. Second, the Markov land change models always ignore spatial autocorrelation. In other word, it assumes that the land use change of each cell is only affected by the current land use state of this cell, independent with other cells. However, the land use change of each cell is likely influenced by its neighborhood cells. For example, the cell surrounded by an urban area is more likely to be developed to an urban area because of the higher development pressure from neighboring regions. Third, different equations are applied in Markov land use and land cover modeling to simulate the land use categories converted from one category to another one. As we all know, the phenomenon of landscape change is caused by the interaction of a variety of exogenous or endogenous factors, especially human activities. With a simple difference equation, it is hard to capture or simulate the contributions of these driving factors in this complex process.

2.2.2 System models

The system dynamic model was first introduced as a modeling methodology in the work of Forrester (1970). Arising from system thinking, approaches to building system dynamic models are well described and discussed in the work of Randers (1980), as well as Richardson and Pugh (1981). As a classic representative of a top-down simulation method, the system dynamic model is well tailored to simulate the process of complex systems over time with macro-level spatial (for example topography) or non-spatial (such as population) variables which are the driving forces of the occurrence of spatial patterns or phenomena. Components and interrelationships within a complex social economic system are mimicked and analyzed in the system dynamic models. By means of representing feedback loops with mathematical equations, system dynamic models are capable of predicting the output of complex and large scale systems under different what-if scenarios with various changes in macro-level social economic driving variables (Liu et al., 2013; Han et al., 2009).

Generally speaking, a complex system simulated in system dynamics is treated as a group of sub-systems. Within the fixed boundary of a complex system, these subsystems interconnect and interact with each other. Through describing the interactions and interconnections among these sub-systems, system dynamics models can reveal how the variables of interest change through time. Therefore, researchers can achieve solutions to better improve the status of social economic and ecological systems taking advantage of what-if scenario analysis with a system dynamic model (Vlachos et al., 2007).

A system dynamics model consists of three components: variables, flows and feedback loops. A variety of variables are connected by two types of flows: physical flow and information flow. Through these two flows, variables can interact with each other to construct a stock-flow diagram of the system dynamics model. The feedback loops represent the cause-effect relationship in the stock-flow diagram. Positive and negative feedbacks can operate at the same time to describe the dynamic change of variables of interest. Moreover, the strength of feedbacks can be varied at different time steps (Chang et al., 2008).

System dynamics models have been widely used to study complex systems, and solve the recourse planning and management problems caused by the social economic driving forces in geography, including human social systems (Qu and Barney, 1998; Sterman, 2000), environmental resource management and planning (Mashayekhi, 1990; Vizayakumar and Mohapatra, 1992a; Vizayakumar and Mohapatra, 1992b; Vezjak et al., 1998; Ford, 1996; Ford, 2000; Wood and Shelley, 1999; Abbott and Stanley, 1999; Guo et al., 2001; Dyson and Chang, 2005), decision and policy making support (Nail et al., 1992; Saysel et al., 2002), ecological modeling (Wu et al., 1993; Grant et al., 1997), as well as urban planning and land cover change simulation (Wolstenholme, 1983; Mohapatra et al., 1994; Guo et al., 2001; Liu et al., 2007; Shen et al., 2007; Chang et al., 2008; Han et al., 2009; Liu et al., 2013).

Wood and Shelley (1999) built a system dynamics model to represent the major influences of pore water metal activity in a constructed wetland. Specially, model testing and validation processes were discussed with several approaches. Saysel et al. (2002) developed a system dynamics model to analyze and estimate policies to support policy making process, which was calibrated with empirical data. Dyson and Chang (2005) applied Stella (software offering visualization tools) to develop a system dynamics model, in order to predict the solid waste generation according to limited empirical data samples in the city of San Antonio, Texas (USA).

Since land cover and land change is usually driven by a set of interactions among components in a complex system, a system dynamics model is well tailored to resolve problems related to land use change and urban development. Several researchers have implemented system dynamic methods to simulate land cover and land use changes at different scales (Wolstenholme, 1983; Mohapatra et al., 1994; Guo et al., 2001; Liu et al., 2007; Shen et al., 2009; Chang et al., 2008; Han et al., 2009; Liu et al., 2013). Guo et al. (2001) proposed a system dynamics model with a name of ErhaiSD to simulate a large scale environmental system. In this model, interactions among four subsystems, including population, industry, pollution control, and water quality subsystem, were investigated taking into account dynamic information feedback among these subsystems. The ErhaiSD was operated with empirical data to support decision-making related to the water-quality deterioration in the Lake Erhai Basin, China. To estimate impacts on the water-quality from various policies, the ErhaiSD ran with four different decision alternatives within a time range of 15 years, starting from 1996. The base run predicted the future of the environmental system under the current environmental and social-economic policy, while other three alternatives were designed with different focuses based on previous planning studies, for example emphasizing economic development.

In the work of Chang et al. (2008), a system dynamics based decision support system (DSS) was proposed to support the environmental planning and management for the Coral reef ecosystem in Kenting, Taiwan. Blending with the integrated coastal zone management (ICZM) concept, the system dynamics based DSS substantially facilitated the resolution of environmental problems by means of conducting scenario analysis. Shen et al. (2009) applied a system dynamics method to study the sustainability of land use and urban growth in Hong Kong, including five subsystems: population, economy, housing, transport and urban/developed land.

Despite the advantages, a system dynamics has its own drawbacks: (1) it does not easily capture the trend of spatial pattern of change; (2) because of the fixed boundary in the system dynamics model, it excludes the external driving factors, which also makes important contributions to changes in a complex system. To overcome the disadvantage of ignoring the spatial pattern, a system dynamic can be integrated with a cellular automata (CA) model (introduced in section 2.2.3). Shen et al. (2007) combined the system dynamic model (Forrester’s urban dynamics) with a CA model to explore the urban growth in Beijing, China. In this study, the system dynamics model was applied to simulate the dynamic change in population and economic growth. The CA model focused on the analysis of spatial interaction and structure in the study area. Han et al. (2009) integrated system dynamic and CA models together to assess the urban growth in Shanghai, China. With the system dynamics model developed in the paper of Han et al. (2009), urban land increase was forecasted based on the feedback loops among three components: residential, public and commercial land.

The system dynamics model also can be combined with a spatial optimization models to solve the problems related to land use allocation. In the paper of Liu et al. (2013), the authors developed a System Dynamic and Hybrid Particle Swarm Optimization Land use Allocation Model (SDHPSO-LA) to optimize the land use pattern for a large study area: Panyu, Guangdong, China, with a total area of 786 km^2. The system dynamics model in this paper focused on the prediction of urban demand based on the analysis of cause effects with feedback loops. The impacts from macro level social economic driving factors, for example population and economic growth, were incorporated and analyzed in the optimization of land use allocation.

2.2.3 Cellular automata models

The phenomena of land use and urban growth are usually modeled in a top-down and deterministic manner based on driving forces in macro scales, for example the system dynamics model. According to complex system theory, however, global spatial patterns and forms can rise from local actions among components composing the complex system, which might be ignored and missed in traditional top-down approaches. Models in land use and urban growth appeal to bottom up approaches in which local actions following simple transition rules lead to the emergency of global patterns (see Batty, 1995). Instead of macro-level driving forces, the bottom-up approaches focus on the changes of spatial and temporal dimensions. The cellular automata model is one of the classical representatives of bottom-up approaches to simulate discrete dynamic systems.

Generally speaking, the cellular automata model consists of three main components: transition rule, state and neighborhood (Toffoli and Margolus, 1987; Batty, 1995; Itami, 1994; Wolfram, 1984). With respect to the principle of self-organization, a set of transition rules are designed and incorporated into the cellular automata model to guide land conversion occurring in each cell from one specific land use type to others in study region. The future state of each cell is determined by the current state value of itself, as well as the current state values of its proximate neighborhoods. Simple self-organization interactions among local neighborhoods can lead to a fractal pattern (the repeating pattern observed at different scales) which reflects the geometric characteristics of urban forms (Batty, 1991; Batty and Longley, 1994; White and Engelen, 1993; White and Engelen, 1997).

The neighborhood in cellular automata can be classified into two types: the first type is the exact immediate neighborhood, for example first-order neighborhood; and the second one is the neighborhood cells which influence the current cell based on information or material flow within a certain distance (distance based neighborhoods). States of different cells are updated simultaneously in the study region. By this way, the future patterns are predicted, taking into account the influences from both the interactions among local neighborhoods and the past trend of urban growth. The general cellular automata model can be described by the following equation:

$$S(global) = f(N, S, T) \tag{2}$$

where the global state of complex system is represented by *S(global)*, which is a function of the transition rule *T* according to the neighborhood *N* and the states of neighborhood *S* at each time step.

The cellular automata model, as a bottom-up model, imitates a complex system within a lattice of discrete cells over discrete time steps, which has been widely used in spatial diffusion modeling such as the urbanization process (Wolfram, 1984; Li et al., 2010). The first cellular automata model was originally developed with two dimensions by mathematician John von Neumann in the late 1940s (Itami, 1994; Santé et al., 2010). Based on the work of Tobler (1979), this bottom-up modelling approach was introduced into the geography domain. Then, Wolfram (1984) defined the characteristics of cellular automata with the following characteristics (Itami, 1994):

(1) The cellular automata model consists of a grid of cells with a finite set of states;

(2) In discrete time steps, cells can change their states with the constant and deterministic transition rules;

(3) The neighborhood is composed of the proximity cells.

There are two stages in the development of cellular automata model in geography: the theoretical application and the real-world application stages. In the late 1980s and 1990s, following the work of Tobler (1979), a large number of geographers employed the abstract cellular automata simulation to investigate the theoretical geography phenomena with respect to urban structure, complexity and sprawl (Couclelis, 1985, 1989; Phipps, 1992; Batty and Xie, 1994a; Itami, 1988; Wu, 1998; Portugali and Benenson, 1995; Cecchini, 1996).These early theoretical applications of the cellular automata model paved the way for applying cellular automata in real-world urban simulation by allowing for the exploration for the theories and hypotheses related to urban growth (Batty, 2007). Specially, in the work of Couclelis (1985), the author discussed the limitations of abstract cellular automata models. To overcome these inherent limitations, the author proposed to relax the traditional definitions in terms of the spatial space, neighborhood and transition functions.

In the 1990s, with the emergence of GIS in spatial simulation and modeling, researchers have explored the integration of GIS and CA to better investigate complex systems (Itami, 1994; Batty and Xie, 1994b; Couclelis, 1997; White and Engelen, 1997; Takeyama and Couclelis, 1997). Due to the inherent grid based structure, CA can be easily incorporated with GIS, and the states and transition rules are greatly improved by the spatial and temporal information provided by GIS.

With the support of GIS, the limitation of uniform space can be relaxed taking advantage of the data processing functions, which makes the cellular automata model much closer to the real-world situation (Batty and Xie, 1994b). A variety of GIS data sources can be used as inputs in cellular automata models such like historical records and maps (Batty and Xie, 1994a; White and Engelen, 1993; Clarke et al., 1997; Clarke and Gaydos, 1998) and remote sensing datasets (Li and Yeh, 2000; Wu and Webster, 1998).

Takeyama and Couclelis (1997) presented the language of geo-algebra, an extension of map algebra, to express the proximal space in cellular automate model. The geo-algebra enabled the capability of manipulating GIS data with three classes of operands: maps, relational maps and metarelational maps (Couclelis, 1997). In the work of Wu (1998), the author combined cellular automata, GIS and multicriteria evaluation (MCE) in a tight coupling way to simulate land conversion between urban and rural. The multicticriteria evaluation implemented with the approach of analytical hierarchy process (AHP) was used to determine the transition rules in the cellular automata model. Land suitability was estimated to determine the transition probability. The model was applied to simulate the urban growth in Guangzhou, China, covering a total area of 225 km^2. Empirical land use data was used based on Landsat Tm-5 digital images. However, Wu (1998) did not take account the impact from human decision making to urban growth.

After the exploration of theoretical cellular automata models and their integration with GIS, cellular automata models have increasingly been used to mimic land use change and urban growth in real-world situations. In comparison with early theoretical models, these models are more sophisticated, because they relax the limitations of the standard cellular automata model (homogeneous space, exactly immediate neighborhoods, and constant transition rules). Instead of being deterministic, probabilistic transition rules are widely spread, taking into account development probability (Batty and Xie, 1994; Batty, 1997) or land use suitability (White and Engelen, 1997; Wu, 1998; Wu, 2002; Wu and Webster, 1998; Wu and Martin, 2002).

White et al. (1997) built a cellular automata model to simulate land use change in Cincinnati, Ohio. A regular grid consisted of 80*80 cells with two different states representing three types of land use: commerce, industry and housing. A suitability value for each land use was calculated and used to determine the transition probability. The effect of distance decay in neighborhoods was considered in transition rules. Empirical land use data of year 1960 was applied to calibrate the model with a trial and error approach. Sensitivity analysis was conducted to test the reliability of the prediction with respect to the stochasticity of the model.

Developed by Clarke et al. (1997), the SLEUTH (Slope, Land cover, Excluded area, Urban, Transportation, Hillshade) model has been applied by a number of researchers to conduct land use and urban simulation in different regions of the world, reported by Clarke et al. (1996), Clarke et al. (1997), Silva and Clarke (2002), Yang and Lo (2003), Herold et al. (2003) and Mahiny and Gholamalifard (2007). The name of the model indicates the six required input data layers: slope, land cover, excluded, urban, transportation and hill shade. Based on these layers, four transition rules were defined to guide the urban growth in study area, including spontaneous growth, new spreading centers, edge growth and road influence growth. Traditionally, the calibration of the SLEUTH model is in a brute force way, according to five parameters: Diffusion, Breed, Spread, Slope, and Road Gravity. Spontaneous growth rule is controlled by diffusion factor. Breed factor is in charge of new spreading centers rule. Spread and slope factors work together to control edge growth rule. Road influence rule takes account all factors except spread. Each parameter ranges from 0 to 100. Monte Carlo simulations are applied based on all combinations of five parameters. According to the evaluation metrics, the optimal combination of parameter sets can be found and used in the prediction of future land use and urban change in the study area.

Besides the SLEUTH model, the constrained cellular model has been developed by Li and Yeh (2000) and White et al.(1997). In the constrained cellular automata model, urban growth and land conversion are regulated by constrained space. For example, cells in a river cannot be transferred into urban space. In the work of Li and Yeh (2000, 2002), an urban cellular automata model was presented to simulate the alternative scenarios of sustainable development. Specially, grey cell was designed to represent the land development percentage. Three levels of constraints (local, regional and global) were defined to ensure sustainable urban development. Among these constraints, land suitability represented by accessibility was considered, as well as environmental factors.

2.2.4 Agent-based models

Usually, agent-based models consist of a set of decision makers, i.e., agents, and the environment where they fit in. Agents are described and defined with a set of attributes and behaviors representing different preferences and uniqueness in agent-based models. Guided by their decision rules, heterogeneous agents interact with each other and environment. With the objective to achieve better situation in the complex system, agents are capable of learning from each other and modifying their own decision rules to adapt to environment in a more appropriate way (Grimm, 1999; Bousquet and Le Page, 2004; Farmer and Foley, 2009). Agent-based models allow us to tap into the impact from individual dynamics on the global pattern. Furthermore, the heterogeneity of spatial pattern can be displayed and captured in the spatial simulation process using a bottom-up way (O'Sullivan, 2008; Tang et al, 2011a; Tang et al., 2011b; Tang and Bennett, 2012; Goodchild, 2004).

In alignment with the development of complexity theory, many modeling advancements have been made to support the study of complex adaptive systems. Agent-based modeling (ABM) is a major one of these advancements to test complexity theory and simulate the complex adaptive system. There were two major development tracks in social science and ecology pushing the advancement of ABM respectively. Within ecology, researchers have developed individual-based modeling (IBM) which focuses on the heterogeneity among individual agents in terms of their attributes and behaviors. On the other hand, ABM, in social science, concentrates on the decision making process of individual agents (Bousquet and Le Page, 2004).

A myriad of agent-based models have been applied in theoretical and empirical spatial simulations, such as land use and land cover change (Axelrod et al., 2000). Similar to Cellular Automata models, agent-based models are computerized simulation approaches in a bottom-up way (Batty and Longley, 1994; Epstein and Axtell, 1996; Batty et al., 2012). Agent-based models are better suited for simulating the individual behaviors and human decision making process, while the Cellular Automata model focuses on spatial interactions in a complex system. In agent-based modeling, individual agents can move in spatial and temporal dimensions. ABM is capable of predicting the output of complexity systems based on assumptions derived from real world phenomena. Thus, what-if questions can be answered in scenario analysis with ABM.

In agent-based models, decision rules always play a critical role in modeling the human decision making process and subsequent actions. Usually, the decision rule is composed of two parts, "if" and "then" parts. Once the predefined conditions (the "if" part) are triggered, agents will act accordingly (the "then" part). Through interacting with and learning from each other, agents are able to improve their decision rules to achieve better situations in the complexity system. In terms of the underlying mechanism, decision rules of agent-based models are mainly derived from the following approaches: utility function, spatial suitability, empirical data and survey, and machine learning algorithms.

With the utility function, the benefit or profit of an agent can be measured and quantified based on a certain combination of non-spatial economic and social factors.

$$utility_i = f(E_i, S_i, N_i) \tag{3}$$

where $utility_i$ represents the value of utility in location i for an agent. E_i, S_i, N_i are the environmental, social, and neighborhood factors in location i, respectively. Each factor has an associated weight indicating the contribution of this factor to the utility value. The objective of agents is to maximize their utility values, that is, the benefits they can achieve in the real world (Brown et al., 2004; Bennett and Tang, 2006; Parker and Meretsky, 2004; Brown and Robinson, 2006; Li and Liu, 2008; Reeves and Zellner, 2010).

In the work of Parker and Meretsky (2004), an agent-based model was designed to investigate how edge-effect externalities can impact the land use pattern. The model simulates the decision making process of individual parcel managers in terms of the conflicts between two possible land use types: urban and agricultural. Individual parcel managers were intended to maximize their profits based on their knowledge. Economic factors, such like the transportation cost, are taken into account in the decision rules of individual parcel managers. Brown and Robinson (2006) presented an agent-based model of residential location to simulate the process of residential development within southeastern Michigan. The heterogeneity in the characteristics and behaviors of actors is represented with the different preferences of the selection of residential location. The utility value of a location was determined by a set of social economic factors, including jobs, aesthetic quality, and the similarity of neighborhood, and the corresponding preference weights.

Instead of economic and social factors, the spatial characteristics of locations are considered by means of calculating the potential suitability for land use transition. With respect to this approach, each agent generates a suitability map for candidate locations based on their own preference. The suitability is related to the accessibility (represented by Euclidian distance or travel time) to social facilities and the main transportation network, for example school, hospital and highway, from the candidate locations, as well as preference weights. According to the suitability value, the conversion probability of each candidate location can be derived with probability function. Thus, agents can select the candidate locations with higher probability to convert (Loibl and Toetzer, 2003; Manson, 2006; Yin and Muller, 2007). In Loibl and Toetzer (2003), a multi agent simulation was designed to investigate urban sprawl driven by population migration and commercial startups in the suburban Vienna Region, Austria. Land use change was impacted by the migration and allocation decision making of a spatial agent model based on the attractiveness and constraints at different scales, which fully considered accessibility including traveling time to the core city and access to motorways. Similarly, Manson (2006) designed a scenario-based model to project the land use and land cover change with different assumptions in Mexico. The spatially explicit model integrates the agent-based model with a Cellular Automata model and GIS to study the underlying driving factors of land use change. Specially, within the ABM component of this integrated model, the suitability values of each cell of the Cellular Automata model are estimated according to the preference weights of the corresponding agent.

Decision rules also can be yield from empirical data and survey, which follows an induction scientific research method. With this approach, researchers have to collect large amounts of empirical data first. Then, data mining and statistical analysis can be used to classify agents into groups and derive decision rules. Specially, neural networks, statistical regression and decision trees are commonly used data analysis methods for the generation of decision rules (Evans et al., 2006; Acosta-Michlik and Espaldon, 2008; Millington et al., 2008; Valbuena et al., 2010). In Millington et al. (2008), the authors applied an agent-based model to simulate agricultural land use decision making and investigate the impact of land tenure and land use on the landscape change in the future in Mediterranean agriculture landscape. Based on empirical data, the authors classified all individuals into two groups with different actions: commercial agents and traditional farmers. While the commercial agents try to achieve profits by means of acting economically rational, traditional farmers conserve their land from a cultural aspect.

The last approach of deriving decision rules is based on the application of machine learning algorithms. In this domain, genetic algorithms and evolutionary programming have been extensively applied to yield decision rules of agents (Manson, 2005; Manson and Evans, 2007). In evolutionary algorithms, a mass of individuals compose a population, where each individual is encoded with a solution of the problem. With respect to varied approaches, the strategies of encoding are different. In genetic algorithms, an individual is encoded with the binary string (zero or one) to represent a specific solution. However, in evolution programming, integer numbers or real numbers are applied in the encoding process. For example, real numbers could be an index of potential facility site.

Generally speaking, at the beginning of an evolutionary algorithm, a population of solutions including a set of individual land use plans is randomly generated. According to the objective of land use change process, a fitness function is used to evaluate each individual. Individuals with high fitness values have a high possibility to be selected to create the population of next generation. In this way, the fitness is preserved for next generation. There are two operators involved in the process of creating the next generation: recombination and mutation. In the process of recombination, selected individuals exchange their components with intention of increasing the fitness. To avoid the local maximum or minimum, mutation operations randomly alter part of the current population to generate the next generation. The new generation repeats the same process of selection, recombination and mutation until the qualified solution is achieved (Xiao et al., 2007). In the work of Manson (2005), the authors developed a land change model taking advantage of the combination of genetic programming, Cellular Automata modeling and agent-based modeling to simulate the land cover change in the Southern Yucatan Peninsula Region of Mexico. Specially, genetic programming is applied to model the human decision making process by serving as symbolic regression solutions, according to the multi criteria evaluation. A set of parent land use plans and strategies compete with each other and evolve to better offspring strategies with higher fitness value, through the genetic programming.

In the book chapter of Parry and Bithell (2012), the authors conducted a review of main stream approaches of building large scale agent based models. As discussed in the article, challenges of large scale ABM can be attributed to massive number of agents, as well as managing complexity of the simulation. In this work, three widely used approaches have been compared and examined: super-individuals, agent-parallel and environment parallel. The first one is an aggregation approach in which individual objects are grouped into a super-agent. With less modification of the model formulation, substantial improvement in terms of computing performance can be achieved with the super-individual approach. However, this aggregation can result in a significant challenge about how to spatial temporally link super-individuals to individuals in an appropriate way. The last two approaches are both based on high performance and parallel computing. The authors summarized the following challenges in parallel ABM simulation: “load balancing among computing processors, synchronizing events to ensure causality monitoring of the distributed simulation state, managing communication between modes and dynamic resource allocation” (Parry and Bithell, 2012, p278). Agent-parallel approach aims to distribute individuals among available computing processors, which is similar to functional decomposition. With this approach, each computing core has to be updated with the information of the complete environment and neighborhood agents. The load balancing of this approach is straightforward, measured by the number of agents. Instead of agents, the environment-parallel approach divides the entire environment into a set of sub domains and allocates them into computing processors, which is similar to domain decomposition. However, when agents have high mobility and the density of agents is spatially heterogeneous, load balancing will be very challenging.

2.3 CyberGIS

With the extreme increase in the complexity and size of data, the computational capability and capacity becomes a bottleneck of spatiotemporal simulations. Regarding this computational issue, researchers have devoted their efforts to synthesize spatial thinking and computational thinking by means of CyberGIS. CyberGIS, empowered by state-of-the-art advances in computer science, consists of three main components: GIS, cyberinfrastructure and spatial analysis in a loosely coupled approach (Wang, 2010). With regard to the tremendous computation power, cyberinfrastructure paves the way for applying CyberGIS to tackle computational issues in spatial analysis and modeling at large spatial and temporal scales.

Exemplified by XSEDE (Extreme Science and Engineering Discovery Environment; http://www.xsede.org) and NSF TeraGrid (http://www.teragrid.org/), cyberinfrastructure substantially facilitates and supports the process of knowledge discovery in a wide array of scientific domains. Cyberinfrastructure consists of high-performance and parallel computing, massive data handling and virtual organization, which bridges the advanced technologies in computer science and solutions of domain specific problems that are infeasible with traditional desktop computing platforms (NSF, 2007).

As the main computing power of cyberinfrastructure, high-performance and parallel computing has been increasingly applied to cope with domain specific problems in terms of computational intensity. To fully harness distributed computing resources, high-performance and parallel computing is always organized as three forms: cluster computing, grid computing and cloud computing (Armbrust et al., 2010). The general idea of high-performance and parallel computing is so called divide-and-conquer, which divides a large scale problem that is computationally intensive into sub-problems that can be simultaneously handled in multiple computing nodes. The decomposition of a problem can be based on the data or functions, guided by domain decomposition and function decomposition respectively. Moreover, sub-problems, which are individual computing tasks, are mapped to each available computing node taking advantage of task scheduling.

Tang et al. 2011b aimed to build a service-oriented ABM simulation framework by means of integrating service-oriented computing (SOC) architecture, ABM (GAIA: geographically aware intelligent agents) and GIS. With support of this framework, computation and data intensity in ABM simulations can be overcomed by leveraging the high performance computing power of cyberinfrastructure (CI) enabled computing resource. The framework encapsulates functionalities into services and is capable of assembling and integrating domain specific functionalities for ABM. In the case study, an elk migration ABM has been implemented with the proposed framework based on supercomputing computing resource on the TeraGrid.

2.4 Parallel spatial simulation model

With the rapid advances of computer science and GIS, the application of high-performance and parallel computing in spatial simulating modeling becomes a crucial research thread in geography. A large number of geographers are dedicated to the exploitation of the parallelism in large scale geographical spatial analysis which was not infeasible for traditional sequential simulation models. In this section, I am going to review efforts invested in this research direction, focusing on three aspects: evaluation of computing performance, parallel computing architectures, and parallel strategies.

2.4.1 Evaluation of parallel computing performance

With regard to evaluating parallel computing performance, there exist in general two common used quantitative methods: speedup factor and efficiency (Wilkinson and Allen, 2004). These two methods are both based on the execution time of best sequential algorithm executed on the single-processor system and the execution time of parallel algorithm with multi-processor system. The speedup factor can tell us how fast the parallel algorithm is, compared with the sequential algorithm, with the following mathematic equation (see Ding and Densham, 1996; Abbott et al., 1997; Hazen and Berry, 1997; Nagel and Rickert, 2001; Owczarz and Zlatev, 2002; Wang et al., 2006; Nichols et al., 2008; Parry and Evans, 2008; Tang and Bennett, 2009; Tang et al., 2011a; Gong et al., 2012):

$$sf = \frac{T_{sequential}}{T_{parallel}} \tag{4}$$

where *sf* is the speedup factor. $T_{sequential}$ and $T_{parallel}$ are execution time of sequential and parallel algorithms, respectively. Based on Amdahl's law, the maximum speedup cannot beyond $1/f$, where f is the percentage of computation which has to be executed in sequential computing.

Efficiency can indicate the percentage of time in which processors execute a parallel algorithm, which always is calculated as following (see Ding and Densham, 1996; Nagel and Rickert, 2001; Tang and Bennett, 2009; Tang et al., 2011a):

$$efficiency = \frac{T_{sequential}}{T_{parallel} \times NP} \tag{5}$$

where *NP* is the number of processors which are used in computation.

Besides the two approaches discussed above, the computation /communication ratio can be used to evaluate the computing performance in message-passing computing (Wilkinson and Allen, 2004). When we parallelize a spatial simulation model with message-passing approach, the execution time actually consists of two parts: the computation part and the communication part which is spent on the inter-processors communication. To achieve better performance on computation, we have to reduce the communication overhead among processors. The computation/communication ratio is defined as following:

$$C = \frac{T_{computation}}{T_{communication}} \tag{6}$$

where $T_{computation}$ represents the time of computation and $T_{communication}$ indicates the time of communication among processors.

2.4.2 Parallel computing architectures and methods

We can divide parallel computing architectures into two categories: SIMD (Single Instruction stream, Multiple Data stream) and MIMD (Multiple Instruction stream, Multiple Data stream) (Ding and Densham, 1996). SIMD is well tailored to the data parallel in which multiple processors executed the same sequential operation on different datasets at the same time. Instead of the same operation, with respect to MIMD, multiple processors concurrently conduct multiple operations with different datasets.

Based on parallel computing architectures, there are two common used paradigms for high-performance and parallel computing: multi-core and many-core computing. Multi-core computing is the natural paradigm to extend single core computing. Moreover, multiple memory modules are connected to each other and can be accessed by multiple processors with support of share memory approach, exemplified by quad processor shared memory multiprocessors (Wilkinson and Allen, 2004). Furthermore, a large amount of computer clusters have been built based on the multi-core machines.

In terms of many-core computing, the computing power of graphics processing units (GPUs), which usually are attached to CPU, has been extensively exploited by researchers in many scientific domains such like geography, biology and physics. GPUs are originally developed with intent to support the display of graphics. Around the middle of 2000s, the general-purpose computing on graphics processing units (GPGPU) was implemented to accelerate scientific computations. GPUs are well tailed to data-parallel paradigm, in which a computationally intensive problem can be divided and executed in a mass of streaming processors (SPs) (Tang and Bennett, 2012). Compute unified device architecture (CUDA) is the programming model and platform which was developed by NVIDIA Corporation in 2007. Taking advantage of CUDA, we can trigger the kennel function, which are conducted on GPUs. Computation results are gathered and copied back to CPU in the host machine (Tang and Bennett, 2009).

To fully harness these two computing paradigms, there are three kinds of parallel approaches: embarrassingly parallel, shared memory, and message passing (Wilkinson and Allen, 2004). Three parallel approaches should be applied in different parallel simulations in terms of a communication mechanism among processors. Embarrassingly parallel suits the so-called complete decomposition, in which there is no communication among multiple processors. On the other hand, if processors need to exchange data with each other, shared memory and message passing approaches should be taken into account in parallel simulation. In terms of shared memory, each memory module can be accessed by multiple processors. In this way, the exchange of data is addressed by common memory space. The message-passing computing has been extensively employed in distributed-memory systems. The data among processors is exchanged in a form of message sending or receiving. When a processor requires data in other processor, required data has to be encoded into a message and sent from one processor to another processor. In summary, both for message-passing and shared memory approaches, the communication among multiple processors lower the computing performance by bringing in communication overhead.

During the end of 1980s and 1990s, a host of researchers exploited the application of parallel spatial simulation modeling in terms of parallel computing architectures and methods. With respect to the application of SIMD, Franklin et al., (1989) presented a parallel algorithm to detect line intersection. The algorithm was well implemented on a Sequent Balance 21000 computer including 16 processors. Bestul (1989) generally discussed the methodology to develop SIMD algorithms according to pointer-based quadtrees structures, in which one processor was assigned to each quadtree node. Li (1992) detailed the application of SIMD computing architecture in the spatial data analysis. On the other hand, parallel strategies, represented by domain decomposition, implemented on MIMD computing architectures were generally discussed by the work of Armstrong and Densham (1992) and Ding and Densham (1996). Hopkins et al. (1992) investigated the scalability of GIS algorithm related to parallel polygon overlay, which was conducted on a MIMD system. Uziel and Berry (1995) parallelized an individual-based model to simulate animal migrations in Northern Yellowstone national park. In this work, the authors employed a 32-processor Thinking Machines CM-5, in which different programs were executed on each processing node with various data.

Besides computing architecture, many scholars have shown increasing interests in various applications of parallel computing approaches. With respect to shared memory approach, Nugala et al. (1998) implemented a parallel individual-based model to simulate the movement of ants on a network of UNIX workstations. Owczarz and Zlatev (2002) presented a parallel spatial simulation related to air pollution with support of shared memory approach. Gong et al. (2012) have applied a hybrid parallelization approach to deal with both computation and data intensity related to land cover and land change modeling. With the proposed parallelization approach, a commonly employed neural network model Fuzzy ARTMAP (Adaptive Resonance Theory MAP) has been modified into a parallel version, based on the sequential version. The results of experiments indicate that the larger the training dataset is, the better the computing performance is.

In the work of Gong et al. (2013), the authors proposed a parallel agent based modeling framework to investigate individual-level spatial interaction. The proposed parallel ABM framework can simulate information change, spatial diffusion of opinion development, and opinion consensus processing. The authors focused on examining how the computing performance can be impacted by two key spatial properties of spatial interaction systems, including the extent and the range of spatial interactions. A parallel algorithm has been developed in this work based on multicore computing environment which is a coarse-grained shared-memory system. The parallel algorithm has been implemented with C++ and OpenMP with domain decomposition, inter-thread data access and synchronization strategies. An equal size decomposition strategy has been applied to reduce the interdependence among sub domains. Also, mutual exclusion algorithm is used to solve the race conditions occurring when multiple threads access and modify the same address in the shared memory space at the same time. Furthermore, the authors used a barrier method to implement the synchronization for iterations to make sure coherence and data integrity.

Besides CPU, shared memory approach has been extensively employed in the application of GPUs within spatial simulation. Richmond et al. (2009) detailed a parallel agent-based model to simulate the movement of pedestrian taking advantage of GPUs. Erra et al. (2009) discussed the implementation of the application of GPUs in large scale individual-based simulation. Tang and Bennett (2009) presented a parallel land-use opinion model with support from GPUs. Spatial opinion exchange, representing the spatial diffusion related to human decision making process, was simulated with ABM in which the individual agents with heterogeneous preference can communicate with each other and improve their own opinions. To fully leverage the high performance computing power of GPU, domain decomposition (mapping agents to threads) and mutual exclusion (shared memory only can be modified by one thread at one time) have been taken into account in the key algorithm.

With regard to message-passing approach, the work of Berry and Minser (1997) discussed the application of message-passing approach in land cover simulation taking advantage of PVM (Parallel Virtual Machine) and MPI (Message Passing Interface). To investigate how alternative land use management scenarios can affect the environmental services, Hazen and Berry (1997) presented a parallel version of Land-Use Change Analysis System (pLUCAS) based on PVM. The pLUCAS has been applied to two spatially distinct study areas: the Little Tennessee river basin (LTRB) in North Carolina and the Olympic Peninsula in Washington State. To tackle the large commuting time and demand at large scale, the authors have utilized the PVM consisting of a network of arbitrary workstations. An embarrassingly parallel computing approach has been implemented, in which different simulation scenarios can be allocated to available computing nodes. The authors discussed the computing issues related to management of interprocessor communication and task scheduling. And to handle these issues, centralized task management was used. In the experiment section, parallel version achieved a peak speedup of 10.77 using 20 computing nodes, compared with a sequential version.

In the work of Guan and Clarke (2010), the authors have developed a general purpose parallel raster processing programming library (pRPL) to parallelize raster processing algorithms. The pRPL has been implemented based on MPI and C++. In terms of parallel strategies, the pRPL can improve the computing performance of parallel cellular automata models by means of incorporating domain decomposition, load balancing and task scheduling. Besides of regular domain decomposition including row-wise, column-wise and block-wise, a quad-tree decomposition has been implemented in pRPL. The quad-tree decomposition can divide workload more even when the study area is highly heterogeneous. However, because of native recursive mechanism, quad-tree decomposition could result in more computational overhead. Taking advantage of static and dynamic load balancing, data parallelism with computing nodes in same group and task parallelism among computing node groups have been provided by pRPL. In order to demonstrate how to utilize the pPRL, a classical CA land use and land change mode SLEUTH has been parallelized in the case study.

And with the trend of applying agent-based model to simulate complex system, researchers have implemented parallel ABM within message-passing computing environment (Parry and Evans, 2008; Tang et al., 2011a). Abbott et al. (1997) conducted a study about white-tailed deer with a parallel individual-based model with support of message-passing approach. Wang et al. (2006) paralleled an agent-based model related to fish population with respect to message-passing mechanism. In the work of Timm and Pawlaszczyk (2005), a conceptual parallel framework of multi-agent simulation has been presented to simulate large scale networks of autonomous decision-makers in logistic domain. To tackle challenges in the management of infrastructures, grid computing approaches have been applied to build a decentralized scalable architecture for grid-based multi-agent simulation, and peer-to-peer technology was used to support the direct communication among nodes within the network.

Reported by Parry and Even (2008), the authors discussed how to utilize limited computing resource to simulate large complex spatial system in ecology with individual based model. Two common used approaches, parallel computing and “super-individuals” have been implemented and examined in Parry and Even (2008). The parallel computing approach can maintain the structure of original sequential model and generate comparable simulation result with support of parallel strategies including messaging passing, load balancing and synchronization. In the experiment part, significant improvement in terms of computing performance (speed up and memory availability) can be observed after five computing processors are used. Besides parallel computing, a super-individual can also be used to handle the challenge of large number of individuals. The basic idea of a super-individual is to reduce the number of individuals by aggregating individuals in a population into super-individuals. Compared to parallel computing, the super-individual approach can substantially improve the computing performance. However, since it greatly modified the model dynamic, the super-individual approach could be inappropriately used in a density-dependent model without considering individual variability. Thus, the authors pointed out that parallel computing was a better choice to solve the computation issue caused by large number of individuals.

Tang and Wang (2009) have developed a hierarchical parallel framework (HPABM) to handle the computational intensity in agent-based model for solving geospatial problems at large scale. With support of HPABM, domain decomposition is conducted to divide the whole big agent based model into sub models which can be grouped into super models. Taking advantage of super model and sub model design, HPABM can leverage the high performance computing power of cyber infrastructure by means of loosely integrating agent model with parallel computing architectures. In the experiment section, the influences of spatial granularity and scalability have been tested with a theoretical agent-based model (StupidModel) parallelized using HPABM framework.

In Kim et al. (2015), the original Schelling model has been parallelized. In comparison with this work, existing parallel ABM work usually have three modeling limitations: 1. the communications and interactions among agents are limited within adjacent neighborhood; 2. Instead of a real world case, they all were tested with an artificial world 3. Most of them use study regions with regular geometric shape (Tang et al., 2011; Shook et al., 2013; Gong et al., 2013; Tang and Jia, 2014). Three domain decomposition strategies (equal area, unequal area, and irregular shape) have been applied and estimated in experiment based on the number of housing units. The unequal area decomposition yielded a better computation performance than equal area, due to the unevenly distributed housing units. In terms of communication, two-stage all-to-all communication between subdomains, as well as exchange boundary information through ghost zone, have been implemented with MPI. Furthermore, the proposed parallel Schelling model was validated with the 2010 Decennial Census.

2.4.3 Parallel strategies related to parallel spatial modeling

A serial of parallel strategies, such as synchronization and decomposition, have been applied in parallel spatial simulation modeling to cope with challenges related to parallel spatial modeling. In this section, we will detail these parallel strategies.

The objective of decomposition strategies is to divide the geographical problem into a myriad of sub-problems which can be simultaneously executed on multiple processors which are available. Generally speaking, decomposition strategies can be categorized into three approaches: complete decomposition, domain decomposition and control decomposition (Ding and Densham, 1996). Complete decomposition is often implemented with a master-slave computing pattern. In this pattern, a computing processor acts as master who distributes tasks and collects results. Other computing processors act as slave in which tasks are executed. Taking advantage of complete decomposition, a problem can be divided and distributed to a mass of processes which do not communicate with each other in terms of computation. Because there is no communication among processes, synchronization is not necessary with complete decomposition. Domain decomposition divides the total dataset into individual data elements based on the characteristics of spatial data. In contrast, control decomposition breaks the entire computational process into a mass of individual functions which are allocated to multiple processors (Armstrong and Densham, 1992; Wang and Armstrong, 2003; Wang and Armstrong, 2009).

After partitioning a problem with decomposition strategy, a problem is divided into isolated subtasks. When we allocate numerous subtasks to available computer resources, load balancing has to be taken into account (Wilkinson and Allen, 2004). If some processors are allocated with much smaller amount of work, they will finish their work much earlier than others, and computer resource will be wasted. In order to improve computation performance by means of minimizing the execution time, we have to allocate workload to each available processor as even as possible.

With an objective of mapping each subtask to available processors, task scheduling strategy provides static and dynamic approaches to balance the workload on each processor. Taking advantage of static task scheduling strategy, we can assign subtasks to available computer resource before execution on each processor. Static task scheduling strategy well suits homogeneous problem which can be partitioned into subtasks with equal workload. Instead of homogeneous problem, dynamic task scheduling strategy is well tailored to tackle inhomogeneous problem in which the workload of subtasks are various. Dynamic task scheduling strategy assigns subtasks to available processors during execution on multiple processors. Furthermore, according to hierarchy and structure among processors, there are two types of dynamic task scheduling: centralized and decentralized (Wilkinson and Allen, 2004).

With regard to communication among multiple processors, we have to take into account synchronization in MIMD systems. According to communication mechanism in shared memory, methods, exemplified by semaphore method, are implemented to coordinate the execution process on each processor. In contrast, based on controlling the message passing process among processors, synchronization is realized on distributed-memory systems, such as the synchronous message passing method (Ding and Densham, 1996).

Geographers have tapped into parallel strategies which can be applied to solve computational challenges related to parallel spatial simulation (Ding and Densham, 1996; Armstrong, 1992; Wang and Armstrong, 2003; Wang and Armstrong, 2009; Nagel and Rickert, 2001). Taking a broad view, Ding and Densham (1996) summarized a set of parallel strategies such as decomposition, load balancing, synchronization and data dependencies. Furthermore, the authors classified the spatial modeling for parallelism and respectively discussed the parallel strategies according to varied spatial characteristics.

Armstrong (1992) is a classical article focusing on discussing the domain decomposition in parallel spatial modeling and how to handle the massive computational requirements with parallel processing. The author introduced the geographical modeling with sequential architectures, vector pipeline architectures, as well as parallel processing architectures. There are two types of parallel computing architectures discussed in this article: SIMD (Single Instruction on Multiple Dataset) and MIMD (Multiple Instructions on Multiple Dataset). In comparison with SIMD, there is more overhead introduced by more communication and movement of instructions and data in MIMD, reducing the computing performance of parallel spatial modeling. Pointed out by the author, two objectives have to be achieved within domain decomposition process. The first objective is to make the working load as even as possible among computing processors, while the second one is to minimize overhead in the task management, for example task scheduling. According to two characteristics of spatial domain, regularity and homogeneity, domain decomposition can be classified into four combinations. Furthermore, the author discussed three types of data dependencies and how to address them in domain decomposition. To better demonstrate the domain decomposition, a *p*-median location-allocation model (selecting facility sites from candidates) has been parallelized in the case study of this article.

To handle the computation intensity in spatial interpolation, Wang and Armstrong (2003) has developed a distributed spatial interpolation algorithm (inverse-distance-weighted (IDW)) algorithm taking advantage of Globus Toolkit (GT) based on Grid computing technology. In this article, the authors have implemented a quad tree approach to conduct domain decomposition. In the work of Wang and Armstrong (2009), a theoretical approach has been presented to capture the computational intensity in geographical analyses. The authors have developed a generic means to formalize the computation intensity of geographical analyses with computational transformations, based on the spatial computational domain representation. The novelty of this paper is that the authors establish computational domains to support the parallel strategies such as domain decomposition, load balancing and task scheduling. In the methodology framework, region quadtrees and space filling curves have been applied to conduct domain decomposition based on the spatial computational domain. In the case study, two types of spatial algorithms, IDW and G* (d) spatial statistic (measuring spatial association among observation points), have been used to illustrated the proposed theoretical approach and associated methodology framework.

Nagel and Rickert (2001) discussed how to parallel the Transportation Analysis and SIMulation System (TRANSIMS). The TRANSIMS takes advantage of microscopic simulation of large transportation system to support the transportation planning, which uses individual objects to represent entities in transport system such as vehicles and travelers. In domain decomposition, the street network was divided into sub domains with similar size, and each sub domain was allocated to a single computing node. Two approaches can be applied in the graph partitioning to efficiently divide the whole transportation network graph: orthogonal recursive bisection and METIS (a graph partitioning library). Then, an adaptive load balancing strategy has been designed and implemented based on execution time of links and nodes. Furthermore, the authors discussed how to systematically predict the computing performance of parallel TRANSIMS model, which can guide the planning of computing budget.

Besides general work mentioned above, parallel strategies has increasingly been applied by a host of scholars in parallel specific spatial simulation to tackle challenges with regard to communication and synchronization among processors, workload balancing, and memory management. In the work of Uziel and Berry (1995), domain decomposition and task scheduling were implemented to cope with irregular shape of study area in parallel individual-based models. To investigate the impacts of alternative hydrologic planning scenarios on wildlife, in Abbot et al. (1997), the paper authors have parallelized an individual based model to simulate the population change of deer and panthers in the Everglades region of south Florida. The authors have selected the simpdel model (spatially explicit individual-based simulation model of Florida panther and white-tailed deer in the Everglades and Big Cypress landscape) as the individual model. A row-wise block-striped partitioning domain decomposition has been applied to divide the landscape into a set of sub groups. When distributing the sub groups among available computing processors, instead of total number of rows, total number of cells on each processor is used to guide the load balancing with respect to the irregular shape of study area. The communication among computing processors in the deer mating process has been implemented by MPI. In the experiment part, the parallel simpdel model achieved a peak speedup of 27 with a population size of 2,000 using 32 computing processors.

Hazen and Berry (1997) designed a centralized task management to realize the inter-processor scheduling of tasks in a distributed computing environment. In Nugala et al. (1998)'s work, the authors evaluated and compared the computing performance of dynamic and static task scheduling with respect to the speedup and communication cost. Static and dynamic load balancing were extensively employed in the partitioning process of parallel spatial simulation models (Lan et al., 2002; Owczarz and Zlatev, 2002; Wang et al., 2006; Parry and Evans, 2008; Nichols et al., 2008).

Nichols et al. 2008 have proposed a parallel ABM framework of nonlinear structured population model taking advantage of message-passing parallel approach. Nichols et al. (2008) discussed the significance and impact of load balancing for the structured population model. Thus, a load balancing algorithm has been designed and implemented to increase the scalability and reduce the computing time of structured population model in this article. In this work, a predator-prey model was parallelized to examine the proposed parallel scheme. Two types of agents, daphnia and fish, have been connected with the feeding mechanism. And due to the large computational requirement, the authors focused on the lifecycle of daphnia. Two strategies local combining and global combining have been examined in terms of the computing performance. And the global combining has been proven to be a better solution than the local one because it can minimize the total workload among computer nodes. Furthermore, a rebalancing algorithm has been designed to balance the workload on each node after a fixed rebalancing period. A superliner speedup was observed in experiments when 2 and 4 processors were used. The authors discussed this result and pointed out the reason was the fully utilization of cache memory on each computing node.

Shook et al. (2013) identified the major computational bottleneck as inter-processor communication in scaling parallel spatially explicitly agent-based models (SE-ABM). In order to overcome this challenge, inter-processor communications have been classified into four categories in the conceptual design in this work, including entity interaction, entity transfer, simulation management, and model parallelization. The entity interaction mainly handles three types of interactions: agent-agent, agent-environment and environment-environment interactions. Entity transfer contains the movement of agents among multiple computing cores. Simulation information such as the status, statistics and I/O (Input/Output) can be retrieved in the simulation management. A set of parallel strategies including domain decomposition, load balancing and synchronization are encompassed in model parallelization. The proposed communication framework can reduce the directly management of inter-processor communication in SE-ABM. There are four interrelated methods in this communication framework: group organization and operation, rectilinear domain decomposition (RDD), a load balancing strategy, and entity proxies. In this work, a distributed load balancing strategy is designed and implemented with three steps: preparation, partition calculation, and redistribution. The computation intensity is calculated by the number of agents time a weight factor, for example the estimated transfer time of one agent. A theoretical agent-based model (Sugarscape model) has been paralleled in the experiments to illustrate the proposed communication framework.

With regard to communication and synchronization, ghost zones are widely used to exchange data and state information among multiple processors in message-passing approach (Wang et al., 2006; Li et al., 2010; Tang et al., 2011a). In order to efficiently simulate land use and land cover change, Li et al. (2010) have developed a cellular automata urban simulation with parallel computing techniques (MPI). To test the computing performance, the proposed framework has been applied to simulate the urban dynamics in the Pearl River Delta, a fast developing area in China. The authors have proposed a line-scanning method calculating the computing time of each sub domain, to conduct domain decomposition based on equal workloads instead of equal area. Furthermore, a ghost area has been designed in each sub domain to exchange necessary information among processors.

In the work of Tang et al. (2011a), the authors have designed and implemented a parallel agent-based model to examine how the opinion of individual agents (ranchers) about the land use policy can achieve consensus through interactions with each other. The ABM of opinion formation is from a land use simulation which is developed to investigate the land use dynamic in southwest Montana. Within the agent-based model, individual agents can randomly select other agent to communicate with in a perceptual window defined by users. Moore neighborhood (eight adjacent and diagonal surrounding cells) has been applied in the process of neighborhood research. If the selected neighboring agent has a close opinion value, the current agent will update its own opinion according to the decision rule. To handle the computation intensity, MPI and C++ have been employed to parallelize this opinion agent based model for large scale study area. In this work, three parallel strategies have been used in the parallel model, including spatial domain decomposition, communication and synchronization among computing nodes. In terms of spatial domain decomposition, the entire landscape was divided with a row-wise regular strategy. Thus, the number of rows in each sub domain can be used to measure the workload on each computing node. Also, ghost zone strategy was used to facilitate the communication among computing nodes. Ghost zones are the overlapping parts among sub domains, through which the neighborhood information on neighboring computing nodes can be accessed and updated. The size of ghost zone was determined by the size of perceptual window. Furthermore, the authors implemented the synchronization at each iteration step with barrier approach.

2.5 Summary

In this chapter, we conducted a thorough literature review to investigate the utility of spatiotemporal simulations and cyberinfrastructure in geographic problem solving.

Spatiotemporal simulations are able to assist urban planners by providing future scenarios analysis among alternative decisions. While simulating urban growth in spatial and temporal dimensions, traditional simulation approaches always ignore the impact from human decision making process. ABM is well tailored to explicitly represent the decision making process in urban growth. By integrating with other modeling approaches such as CA, ABM allows to investigate how agent behaviors at individual level can generate spatial land use patterns at aggregate level.

Cyberinfrastructure has been applied to provide support for solving computational intensity associated with ABM. With the support of high performance and parallel computing, we can substantially improve the modeling capability of ABM. Empirical data with large data size and fine spatial resolution can be used to derive decision rules and calibrate model parameters used in ABM. Furthermore, parallel strategies (such as decomposition, task scheduling, and communication) and evaluation metrics of computing performance have been developed based on the characteristics of spatiotemporal simulations.

Load balancing has been recognized as an important objective in the design of parallel strategies such as decomposition and task scheduling (Armstrong, 1992; Ding and Densham, 1996; Wilkinson and Allen, 2004; Timm and Pawlaszczyk, 2005).The application of load balancing can substantially improve the efficiency of parallel simulations. Load balancing requires the accurate estimation of computing intensity. However, existing work did not provide a method for the estimation of computing intensity. Most parallel simulations only roughly estimated computing intensity based on their own simulation mechanisms (Abbot et al., 1997; Nagel and Rickert, 2001; Wang and Armstrong, 2003; Nichols et al., 2008; Guan and Clarke, 2010; Parry and Bithell, 2012; Shook et al., 2013; Pijanowski, 2014).These previous work did not present an general approach to capture the relationship between spatial characteristics/content and computing intensity. Consequently, it is hard to apply these methods in spatial simulations with different mechanisms. Furthermore, they did not report a method which can validate their estimation of computing intensity.

In particular, Wang and Armstrong (2009) has developed a theoretical approach to assess computational intensity in geographical analysis. However, the limitation of this approach is that it requires sufficient and detailed knowledge about the underlying mechanisms which is hard to obtain due to the dynamic and complex characteristics of a spatiotemporal simulation. In conclusion, a major limitation of previous work is that there is no general approach to appropriately estimate the computing performance and validate the estimation. My dissertation will fill this gap by presenting an empirical approach to estimate the computing performance of spatiotemporal simulations. This approach ignores the underlying mechanisms, which can be generally applied in parallel simulations. The estimation accuracy of computing performance is also provided by this approach for validation purpose.

In Sum, applying cyberinfrastructure in solving computational challenges in spatiotemporal simulations has greatly improved our capability of modeling complex adaptive systems, which represents a research frontier in the study of complex adaptive systems. Although researchers have developed parallel frameworks to support spatiotemporal simulations based on different computing architectures (such as SIMD and MIMD) and parallel strategies, it is still not clear how to appropriate estimate its computing intensity. Therefore, the efficient use of cyberinfrastructure in spatiotemporal simulations has not been adequately studied.

CHAPTER 3: SURROGATE-BASED MODEL

This chapter focuses on the exploration of the utility of surrogate-based models. First, I broadly illustrate the background of surrogate-based modeling in section 3.1. In section 3.2, I generally introduce components of surrogate-based modeling based on literature review. In section 3.3, I discuss the important role of surrogate-based models in supporting parallel spatiotemporal simulations. At last, a sequential procedure of the construction of surrogate-based models is presented to demonstrate how to build and apply a surrogate-based model in a spatiotemporal simulation.

3.1 Background of surrogate-based models

Surrogate-based models, also known as response surface models or metamodels, have been extensively employed to solve computational intensity for high-fidelity simulations in different domains, including mathematic, computer science, and engineering domains (Forester et al., 2008; Kleijnen, 2009). Since these high-fidelity simulations usually have multiple objectives which are mutually competing with each other, it might consume long computation time for specific combinations of parameters. Therefore, routine tasks in simulation such as sensitivity analysis and model optimization are impossible to complete within affordable computing time based on a large parameter space.

Surrogate-based approaches can greatly alleviate this computing intensity using an approximation way. While spatiotemporal simulations capture the causal model of phenomena or systems, surrogate-based modeling focuses on the approximation of transformation from input to output, ignoring the exact processes within a system. Surrogated-based models are selected to fit I/O data generated in spatiotemporal simulation runs, and then provide fast predictions for simulation output (Queipo et al., 2005; Forrester et al., 2008; Forrester and Keane, 2009).

To investigate the development of surrogate-based models, I conduct citation and subject searching for research journals. The database used is Web of Science (ISI), in which 5,741 articles are found under the topic “surrogate model” or “response surface model” or “metamodel”. The research of the utility of surrogate-based models can be traced back to 1970’s. Harvey J. Greenberg published the very first article of surrogate model in the journal of Operations Research in 1973 with 21 citations. The author has developed a generalized surrogate model based on the monotonic-penalty-function (Greenberg, 1973).

The work of Stockwell and Peterson (2002) has the highest citation number of 482 in this domain. In their work, the authors have applied surrogate-based models to simulate ecological niches and predict geographic distributions. The occurrence data of bird species (around 300,000 records in total) was used in this study, which was extracted from the Atlas of the Distribution of Mexican Birds. Through re-sampling for the species data, the authors explored how the sample size can affect the accuracy crossing surrogate models.

Figure 2 shows the number of publication for each year from 1973 to 2017. We can see the exploration of surrogate-based model keeps increasing with the advances in computational power (Sacks et al., 1989). Figure 3 shows the top 10 research areas where these publications appear. Based on Figure 3, we can see the research of surrogate-based models is dominated by two research areas, engineering and computer science, with a percentage of 87%. Surrogate-based models have been most used in engineering with a percentage of 44%. Close to engineering, computer science has the second highest number of publications (43%) related to surrogate-based models. Following engineering and computer science, the number of publications of surrogate-based models only takes a percentage of 7.7% in mathematics.

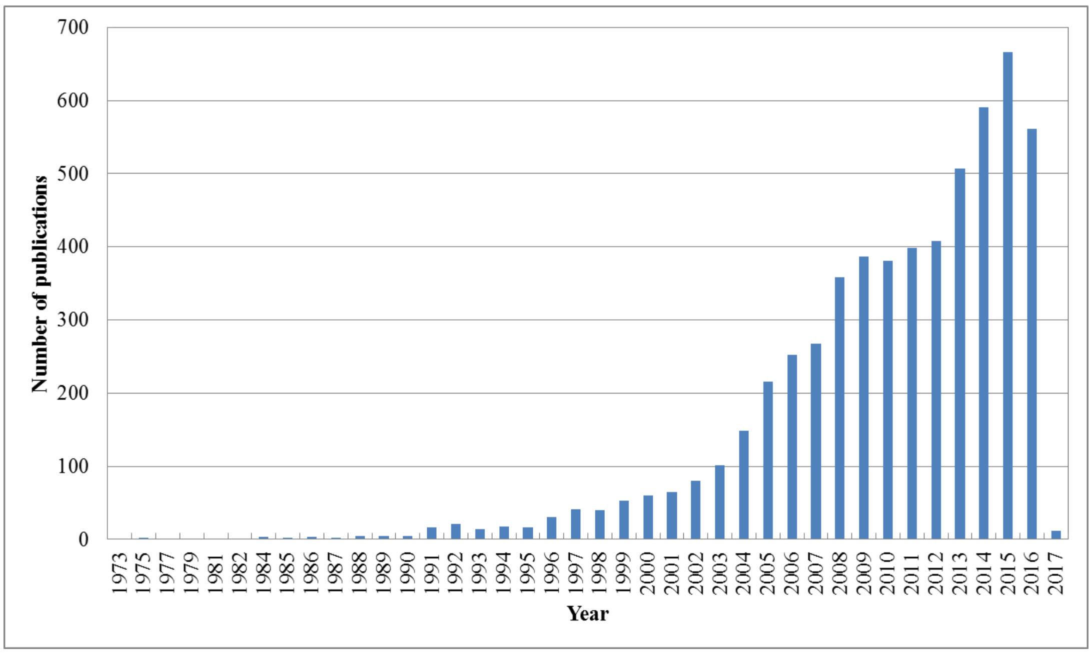

Figure 1. Number of surrogate-based modeling articles published by year

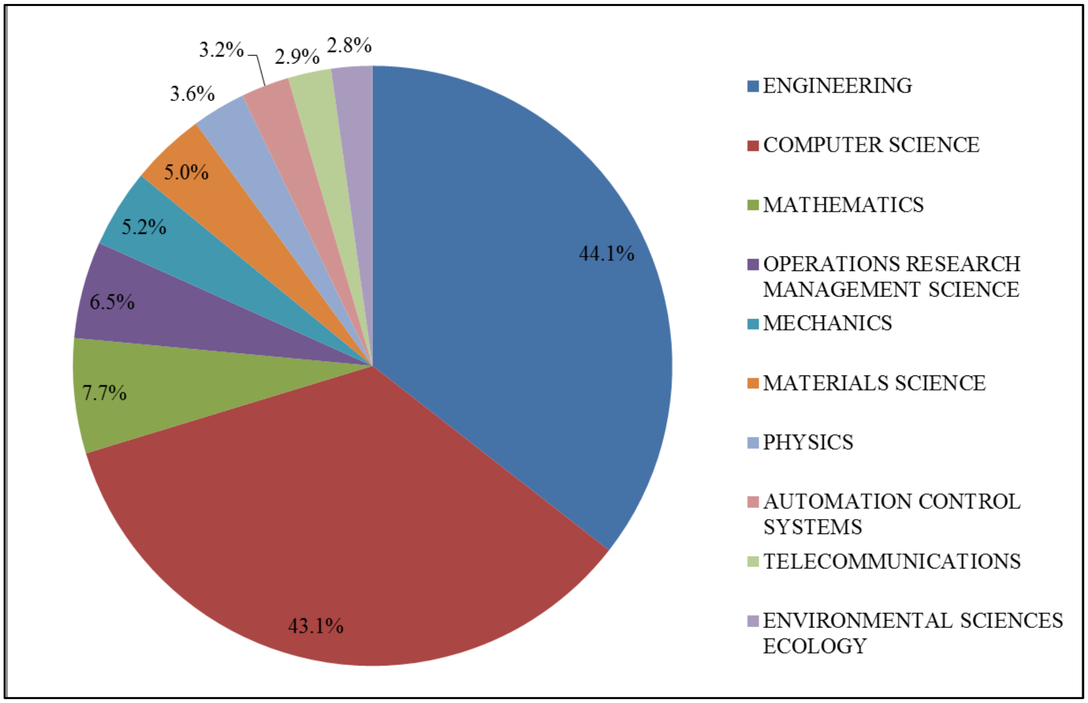

Figure 2. Research area of surrogate-based modeling

Instead of using metamodel in an ad hoc way in previous studies, Kleijnen and Sargent (2000) firstly demonstrated a methodology for the development of metamodel. The methodology of Kleijnen and Sargent (2000) mainly focuses on distinguishing the construction and validation of a metamodel. By emphasizing the relationships among problem entity, simulation model, and metamodel, Kleijnen and Sargent (2000) summarized and discussed that there are four objectives in metamodeling: understanding, prediction, optimization, and verification and validation. In particularly, Kleijnen and Sargent (2000) explicitly illustrated the procedure of metamodeling for random simulation using linear-regression.

Take a broad view, many researchers have dedicated to review the development of surrogate-based modeling with a varied of simulation goals (Queipo et al., 2005;

Forrester and Keane, 2009; Keleijnen, 2009). Generally speaking, a surrogate-based model is built through the following three steps: 1) design of sampling strategy; 2) construct the surrogate model; 3) validate the performance of surrogate model (Queipo et al., 2005; Forrester and Keane 2009). Queipo et al. (2005) identified the significance of SBAM (surrogate-based analysis and optimization) in handling computational challenges within the development of aerospace systems. In the work of Queipo et al. (2005), the authors discussed several fundamental issues in SBAM, including sampling strategy, the selection and validation of model, surrogate model construction, sensitivity analysis, and surrogate-based optimization.

Following the work of Queipo et al. (2005), Forrester and Keane (2009) conducted a deeper review to illustrate the pros and cons of mainstream approaches in the construction of surrogate-based optimization such as polynomials, moving least-squares, radial basis functions, kriging, and support vector regression. Keleijnen (2009) focused on the review of using Kriging metamodeling in deterministic and random simulations. The work of Keleijnen (2009) demonstrated that how the bootstrapping method can be used in the estimation of the variance in Kriging metamodeling. Keleijnen (2009) also compared Kriging metamodeling to classic polynomial regression. Furthermore, this paper respectively presented the designs of Kriging metamodeling for two basic goals of simulation (sensitivity analysis and optimization).

With respect to the application of surrogate-based modeling, Goel et al. (2007) proposed a response surface approximation approach for multi-objective optimization in rocket injector design. The authors of Goel et al. (2007) applied a quintic polynomial to approximate the Parceto optimal front (POF) which helps researchers choose compromise designs by representing trade-offs among multiple objectives. Gorrissen et al. (2010) demonstrated a software framework and developed a toolkit, Matlab SUrrogate MOdeling (SUMO), to construct surrogate models for computer simulations. In particularly, Gorrissen et al. (2010) implemented an active learning (or sequential design) process in the sampling process.

3.2 Components of surrogate-based models

A surrogate-based modeling usually consists of three components: spatiotemporal simulation, design of experiments, and model selection (see Figure 1). In the component of spatiotemporal simulation, we should determine which simulation variables should be taken into account in surrogate-based modeling. Simulation variables can be selected based on their relative importance to the output of simulation. The value of selected variables will be varied in experiments of simulations. Based on the underlying mechanism of simulation, we can choose a subset of all variables as the initial sample design at first. According to simulation results and analysis of the initial sample design, we then can further add or remove variables to improve the initial sample design.

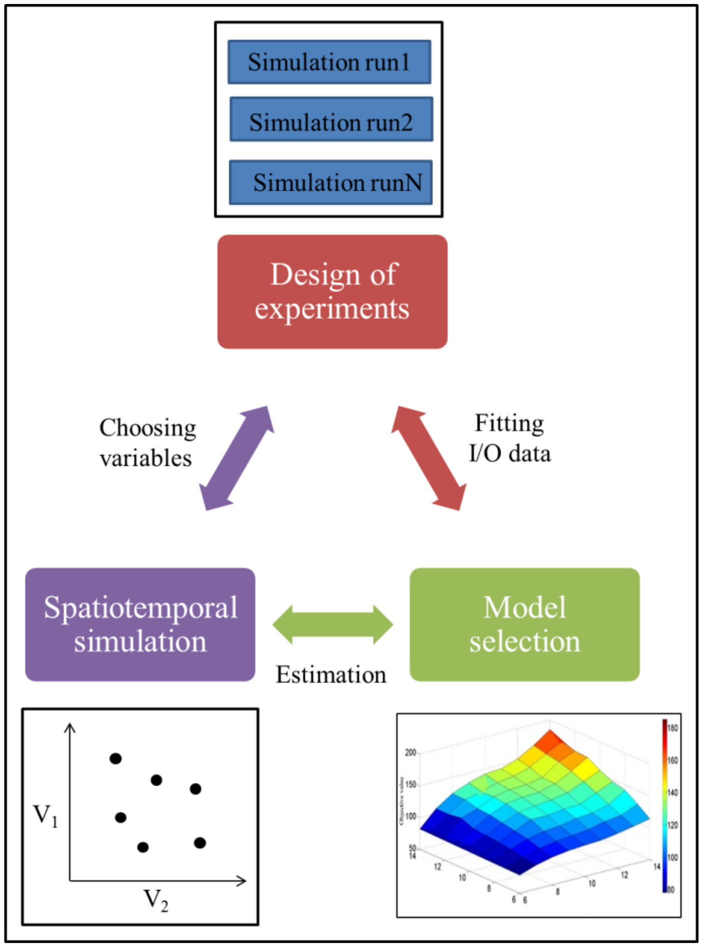

Figure 3. Components of surrogate-based modeling

The second component is design of experiments, which is also known as sample selection, sampling plan, optimal experimental design or active learning in different scientific domains. In this component, we should select which sample designs we used to construct our surrogate-based models. The purpose of this component is to running a set of simulations according to a sampling strategy. The sampling strategy should disperse sampling points crossing the parameter space of selected simulation variables. The spreading pattern of sampling points in the parameter space should be uniform (Forrester and Keane, 2009).

Volumes of sampling techniques could be applied in order to construct surrogate-based models. Latin hypercube sampling is a common used sampling strategy in surrogate-based modeling, which conducts stratified sampling of in parameter space based on a multidimensional distribution (McKay et al., 2000). Different sampling strategies have been developed by improving the optimization of a spreading measure in standard Latin hypercube sampling (Johnson and Moore, 1990; Tang, 1993; Owen, 1994; Palmer and Tsui, 2001; Queipo et al., 2005). For example, in the work of Tang (1993), the author developed an orthogonal array (OA) based Latin hypercube sampling and compared the performance to standard Latin hypercube sampling. Morris and Mitchell (1995) illustrated the space filling maximim Latin hypercube sampling techniques, and the Matlab implementation was presented in the book of Forester et al. (2008). In addition, Forrester and Keane (2009) discussed an adaptive sampling in which sampling strategy could be recursively improved with regards to the analysis of previous sampling results.

The third component is model selection in which we select surrogate-based models to fit the sampling I/O data from experiments. Generally speaking, the surrogate-based modeling can be classified to two categories: parametric and non-parametric. Parametric surrogate-based models focus on representing the relationship between simulation variables and corresponding simulation output, for example polynomial regression and Kriging. Compared to parametric approach, non-parametric surrogate-based models do not have a fix number of parameters, such as radial basis functions. The number of parameters is determined based on the training data, not the spatiotemporal simulations. According to my research objective, I concentrate on parametric surrogate-based models in this dissertation.

Polynomial regression is a relatively straight forward and computationally efficient method in surrogate-based modeling (Kleijnen and Sargent, 2000; Gano et al., 2006; Draper and Smith, 2014). It is suitable for linear, multivariable relationship. On the other hand, Kriging is a statistically-based approach in spatial interpolation, original used in spatial statistics to predict spatial variables (Krige, 1951). The spatial interpolation is developed based on the first law of geography (Tobler, 1970, p236): "Everything is related to everything, but near things are more related than distant things." According to values of spatial variable at sample points, Kriging can estimate values of spatial variable across the study area.

Kriging approach consists of three main components: spatial trend, spatial autocorrelation, and random variation (Bolstad, 2005). The spatial trend indicates an increase/decrease of variable value along with directions. The spatial autocorrelation component represents the level of similarity of points near to each other. In particular, semiviarance is applied to measure the spatial autocorrelation at a given distance. Depending on the spatial autocorrelation, Kriging approach assigns an optimal set of weights for all selected driving variables in spatial interpolation (Cressie, 2015). Then, these weights can be applied in the following equation to estimate value of spatial variable at unknown location:

$$V_i = \sum_{n=1}^{k} (W_n V_n) \tag{7}$$

Where V_i is the estimated value of spatial variable at point i, W_n is the weight for each sample point n, and V_n is the value of spatial variable at sample points. Kriging surrogate-based models are more suitable to simulate non-linear relationship between multiple simulation variables and output (Jones et al., 1998; Van Beers and Kleijnen, 2001; Simpson et al., 2001; Santner, 2013).

The performance of surrogate-based models can be evaluated with alternative approaches in the component of model selection. Cross-validation and bootstrapping strategies are two common used approaches in different types of surrogate-based modeling (Queipo et al., 2005; Martin and Simpson, 2005; Forrester and Keane, 2009; Myers et al., 2016). (1) With k fold cross-validation strategy, we first divide the sampled I/O simulation data into k sub datasets. (2) Select one of the k sub datasets as validation dataset and other $k-1$ remaining datasets as training dataset. (3) We train our surrogate-based model with training dataset and evaluate its prediction performance with validation dataset. (4) Keep repeating step 2 and 3 until each sub dataset is selected as validation dataset once (Golub et al., 1979; Efron, 1983).

Compared to k fold cross-validation strategy, bootstrapping approach does not have to split the I/O data into sub datasets. It randomly selects subsamples in all sampling I/O simulation data as training data with replacement. And the rest of original sampled I/O simulation data are used as validation dataset (Hall, 1986; Efron, 1993). Based on the result of validation, we optimize the parameters used in surrogate-based models to minimize the prediction error represented by Root-Mean-Square error (RMSE). RMSE of each surrogate-based model is calculated with the following equation:

$$RMSE = \sqrt{\sum_{1}^{n}(\mathrm{y} - \mathrm{y}_{\mathrm{predicted}})^2/\mathrm{n}} \quad (8)$$

where y is the observed computing time of a computing job, while its predicted computing time is $y_{predicted}$. n is the total number of computing jobs. Represented by RMSE, it is easy for users to know how accurate the prediction result of computing performance and requirements of computing resources is for spatiotemporal simulations.

3.3 Significance of surrogate-based models

Geographers have developed spatiotemporal simulation models to better capture and understand underlying mechanisms of complex geographic phenomena. With rapid advances of computer science and data acquisition technology, complexity and data size in spatial analysis and modeling tend to exponentially increase, which result in expensive computational cost in the resolution of geographical spatial problems. Consequently, routing tasks in spatiotemporal simulations, for example model calibration, sensitive analysis, become impossible to complete due to incredibly computationally expensive processes with large amount of repetitive runs of model. Cyberinfrastructure is a promising way to provide high performance computing power, consisting of high-performance and parallel computing, massive data handling, and virtual organization (NSF, 2007).

In order to efficiently leverage the computing power of cyberinfrastructure, appropriately estimating the computing performance of spatiotemporal simulations is vitally important for guiding experiment designs. However, because of the complexity of spatiotemporal models, there is a huge challenge to estimate the computing performance of a parallel spatiotemporal simulation within cyberinfrastructure environments. To efficiently overcome this challenge, I develop a surrogate-based approach to assess the computational performance of a parallel spatiotemporal simulation within cyberinfrastructure environments. In this work, I incorporate the surrogate-based modeling and cyberinfrastructure together to solve computationally intensive issue in a spatiotemporal simulation.

Surrogate-based modeling approach has been quite widely studied and employed in different scientific domains, handling computational intensity design problems (Wang, and Shan, 2007). However, this approach has been rarely introduced into geography domain to facilitate spatial analysis and modeling. To fill this gap, I propose a surrogate-based methodology to tackle computational burdens in spatiotemporal simulations within cyberinfrastructure environments. Our methodology in this dissertation has the following novelties: first, I firstly highlight the important role of surrogate-based models in the study of complex geographic phenomena; second, the methodology illustrates how to fit and validate a surrogate-based model within a spatiotemporal simulation; third, I couple surrogate-based models with parallel computing to aid the efficiently utilization of cyberinfrastructure.

3.4 Designs for spatiotemporal simulations

Alignment with other scientific domain such as engineering design, we conduct volumes of simulations to solve real world problems in geography (e.g. global climate change, food crisis, and deforestation). However, spatiotemporal simulations for those real world problems always bring us overwhelming computing challenges when we have a very large study area. The objective of this work is to develop a generalized methodology to support parallel spatiotemporal simulations by blending surrogate-based modeling with cyberinfrastructure. Through the support of this methodology, computing performance of parallel spatiotemporal simulations within cyberinfrastructure environments can be appropriately predicted to guide the design of computationally intensive tasks such as model calibration, sensitive analysis, and model optimization.

In this section, I will illustrate the design of a sequential procedure for the construction of a surrogate-based model to estimate the computing performance in spatiotemporal simulations. Generally speaking, I develop spatiotemporal simulations to help us understand the underlying mechanisms of complex geographic phenomena. Within spatiotemporal simulations, a set of model variables are designed to present heterogeneous spatiotemporal characteristics/content of units in study area, for example input data size, coverage, population, time range and so on. Some of these spatial characteristics/content variables can directly impact the computing intensity of the spatiotemporal simulation.

We can conduct spatiotemporal simulations with sampled combination of variables values and generated the I/O (input/output) simulation data. While input is values of spatial characteristics/content variables, output is the associated computing performance (e.g. computing time). Based on values of spatial characteristics/content variables used in spatiotemporal simulations and their computing performance using cyberinfrastructure, a surrogate-based model can be built to represent the relationship between them. This surrogate-based model is able to help users manage computing resources and computing cost (time and money) by constructing the response surface of computing performance for the whole parameter space of variables.

I design a sequential procedure to build a surrogate-based model to capture the relationship between spatial characteristics/content variables and computing performance. The sequential procedure consists of five main parts to build a surrogate model (please see Figure 4): sample selection, identification of driving variables, surrogate model construction, optimization, and response surface generation as follows:

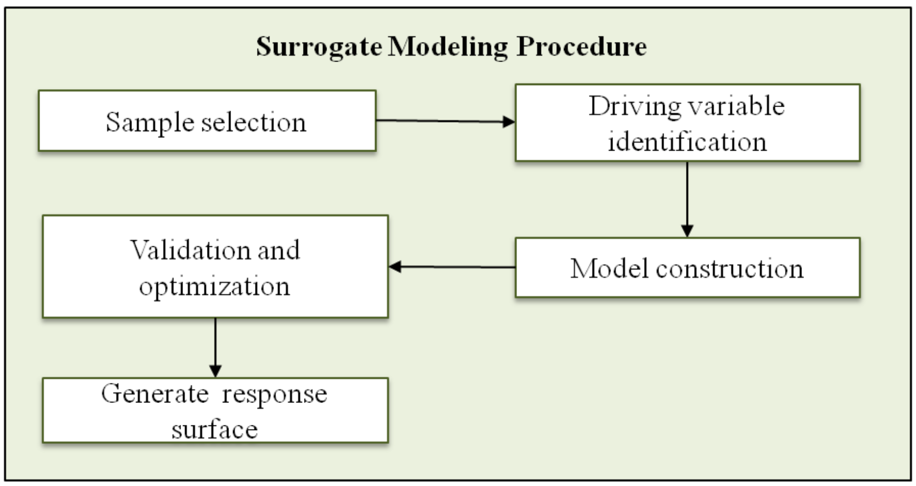

Figure 4. Sequential procedure of surrogate-based modeling

(1) Start with sample selection to generate I/O simulation data. According to parameter space characteristics of spatial characteristics/content variables, we can select appropriate sampling strategy such like random sampling or systematic sampling. Based on the sampling strategy applied, we conduct the spatiotemporal simulation with sampled combinations of spatial characteristics/content variable values in parameter space. Corresponding computing performance for each combination of characteristics/content variable values is recorded as the sampling I/O simulation data based on the computing

resources employed. In order to obtain statistical robust sampling data, Monte Carlo approach can be applied to repeat spatiotemporal simulation run of each parameter combination for multiple times. Then the average computing performance can be calculated and used in I/O simulation data. In this way, we can greatly reduce the uncertainty in I/O simulation data due to computing resource hardware.

(2) Conduct the identification of driving variables. The driving variables in surrogate-based modeling are spatial characteristics/content variables which have significant influence on the computing performance of spatiotemporal simulations (e.g. computing time). According to I/O simulation data, we can first explore the relationship between different spatial characteristic/content variables and computing performance by means of calculation of the correlation. Based on the correlation of spatial characteristic/content variables, select these variables that have high negative/positive correlation with computing performance as the driving variables.

(3) Construct a surrogate-based model according to the I/O simulation data and driving variables identified. Choose a type of surrogate-based model and fit the surrogate-based model to the sampled I/O simulation data from the previous step. Surrogate-based modeling per se is a data driven, black-box, approaching approach. Therefore, with advances in data analysis and data mining technology, volumes of methods are well studied and compared for their applications in surrogate-based modeling, such as polynomial regression, radial basis functions, Kriging, and machine learning methods (Queipo et al., 2005; Forrester and Keane, 2009; Keleijnen, 2009). According to different mechanisms, each type of surrogate-based models presents its own

advantage and disadvantage in terms of their computational effectiveness and prediction performance. In this dissertation, I construct surrogate-based models to represent the relationship between spatial characteristic/content variables and computing performance with two common used approaches: polynomial regression and Kriging.

(4) Validate and optimize the surrogate-based model. In this step, we evaluate the prediction performance of surrogate-based model constructed in previous step and optimize its parameters to minimize the error between real and estimated values of computing performance. Base on the sampled I/O simulation data, we can choose one of common used model validation strategies to assess the prediction performance of our surrogate-based model, such as k fold cross-validation or bootstrapping approaches.

(5) Generate the response surface. At last, a response surface of computing performance is generated for the corresponding parameter space in spatiotemporal simulations.

In order to improve the computational efficiency, this sequential procedure of the construction of surrogate-based models can be automated and wrapped into scientific workflow (discussed in detail in Chapter 4). With support of scientific workflow, this sequential procedure can be implemented crossing computing platforms (such as Window and Linux platforms) within cyberinfrastructure. And the modularity of this sequential procedure is substantially improved. Therefore, we are able to integrate this sequential procedure in our computational framework to estimate the computing performance of spatiotemporal simulations within cyberinfrastructure environments.

CHAPTER 4: COMPUTATIONAL FRAMEWORK

4.1 Research objective 1 methodology

4.1.1 Computational framework of spatiotemporal simulation

For the efficient use of cyberinfrastructure, it is critically important to appropriately estimate the computing performance and computing resource requirements of parallel spatiotemporal simulation. However, the heterogeneity of computing recourses increases the complex of the estimation of computing performance (Wang and Armstrong, 2008). With regard to the first research objective, I propose a computational framework to facilitate the application of cyberinfrastructure in spatiotemporal simulation. With support from this computational framework, the computation performance of routing tasks within spatiotemporal simulation (e.g., calibration, validation, and what-if scenario analysis) can be predicted based on computing resource employed.

The computational framework of a spatiotemporal framework consists of three components: spatiotemporal simulation, cyberinfrastructure, and a surrogate model (see Figure 4). Instead of exact processes, a surrogate model is a data driven method which focuses on the relationship between input and output. For the real-world spatiotemporal simulation, it is impossible to go through all parameters to obtain the exact information of computing performance and requirements of computing recourses. Taking advantage of the surrogate model, our computational framework can represent the computing performance, measured by computing performance metrics such as speedup and computing time, with a response surface. Instead of the whole parameter space, the construction of this response surface only requires a limited number of sample parameters and associated information of computing performance. The response surface represents the computing performance for the whole parameter space in spatiotemporal simulation with cyberinfrastructure.

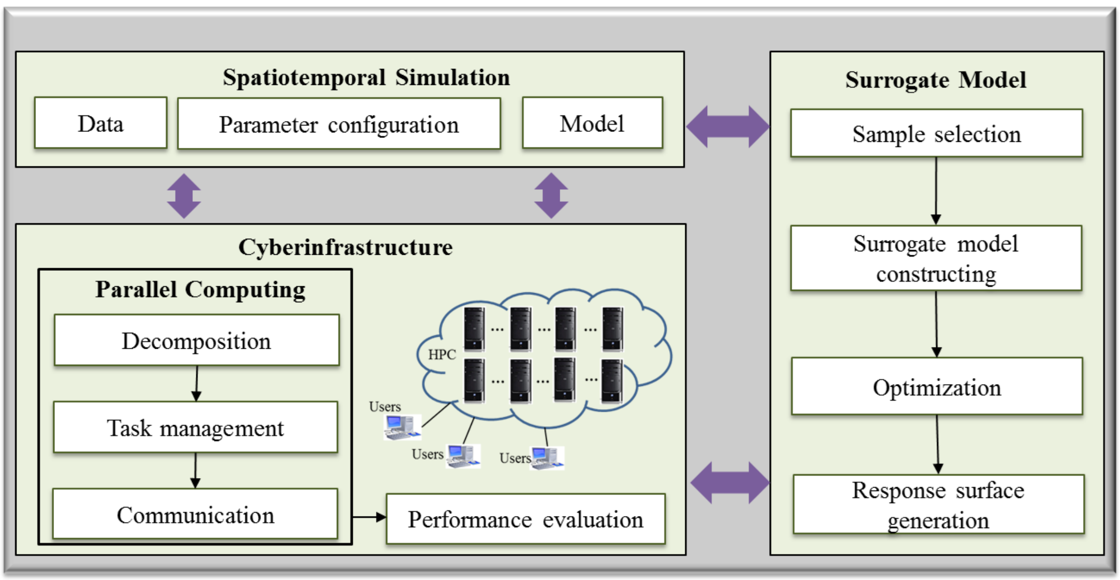

Figure 5. Surrogate-based computational framework of parallel spatiotemporal simulations

The component of cyberinfrastructure is in charge of handling the computation intensity using parallel computing. Based on the concurrency of spatiotemporal simulation and the computing resource used, we can decompose the whole computation task into a set of small computing tasks. The number of tasks should be large enough to achieve speedup with support of massive computing processors. When mapping computation tasks among computing processors, appropriate task scheduling mechanisms have to be designed to avoid extra communication in task management. Information update and synchronization have to be considered in the communication. Since communication in the simulation can introduce computation overhead, we should reduce communication among computation tasks as much as possible. At last, computing performance of the spatiotemporal simulation are estimated to provide the computing metrics for surrogate-based model, according to employed computing resources.

4.2 Research objective 2 methodology

In order to realize the computational tractability, I implement three main approaches in our computational framework: parallel computing, scientific workflow and automation.

4.2.1 Parallel computing

A set of parallel strategies have been applied to enable our computational framework the capability of big data handling and leveraging high performance computing in large-scale urban growth simulation. The parallel strategies have been implemented concentrating on three parts: decomposition, task management and communication.

Once spatial datasets are deployed in the framework, based on the spatial and computational characteristics of these datasets for example the shape of study area, an appropriate decomposition strategy, e.g., domain decomposition, control decomposition, and hybrid decomposition, can be designed to divide the datasets into a set of tasks. In our computational framework, the computation intensity due to I/O process could be substantially mitigated with control decomposition which divides various data input and output functions into a set of sub tasks running on different computing nodes. And the domain decomposition, for example column-wised decomposition, is employed in the urban growth simulation to decompose the whole study area into sub regions.

After the decomposition, all tasks wrapped up in decomposition are aggregated based on their estimated computation intensity. Taking advantage of statistical manner, computation intensity could be evaluated based on the relationship between the size of data and the computing performance metrics such as computing time. Then all tasks are assigned to available computing nodes from the master node. In order to minimize the parallel computing time decided by the longest computing time among computing nodes, load balancing strategies are implemented to make the work load as even as possible, either in static or dynamic ways.

With regard to heterogeneous computing architectures, there are two types of communication methods provided in our computational framework: message passing and shared memory. Within one computing node, multiple threads can communicate with each other by accessing the same memory to fetch information needed in urban growth simulation. In other hand, for communication among different computing nodes, message passing method is implemented with ghost zone strategy. Ghost zones on each computing node consist of cells not in current computing node but affecting the land cover transition of cells in current computing node. The status of the ghost zone cells are updated by corresponding computing nodes at the end of each time step. And the size of ghost zone depends on the land cover transition mechanism used in urban growth simulation.

4.2.2 Scientific workflow middleware

The spatiotemporal simulation of urban growth is a very complex process involving heterogeneous data and multiple processes. It is a big challenge to efficiently manage, execute and share complex urban growth simulation. To overcome this challenge, it is imperative to design a middleware to glue functionalities and computational infrastructure together in urban growth simulations. Scientific processes and heterogeneous data can automatically be connected and executed within scientific workflow. In this way, the reusability of functionalities of the spatiotemporal simulation can be greatly improved (Medeiros et al., 2005; Ludäscher et al., 2006; Taylor et al., 2014).

A scientific workflow middleware is implemented to integrate all processes and data crossing a heterogeneous computing platform in the proposed computational framework. Besides integrating and organizing processes, the scientific workflow also can enable the computational framework the capability of provenance. Thus, the information of input data, parameters used, intermediate results, and simulation result can be collected and recorded for multiple uses in the future such as validation and reproducibility.

To implement the scientific workflow, I apply scripting languages, including Windows batch scripting (https://en.wikibooks.org/wiki/Windows_Batch_Scripting), Linux shell scripting (http://www.freeos.com/guides/lsst/) and Python (https://www.python.org/). Taking advantage of scripting languages, we can organize data and execute processes in heterogeneous computing environments (e.g., network, desktop computers, and grid). Moreover, python allows us to better leverage functionality presented in most GIS software such as ArcGIS.

4.2.3 Automation

To make the computational framework as simple as possible, I implement the automation for spatial sampling and cross validation processes. A Monte Carlo based spatial sampling approach is developed to support the statistical analysis in the computational framework. Due to the limitation of computation capability, sample sizes are relatively small in traditional spatial sampling methods, which cannot fully reflect the heterogeneous spatial characteristics of a large scale study area. The spatial characteristics can be better discovered by spatial sampling with a large sample size. With the capability of sampling a massive number of sample points at different region, the spatial sampling approach allows researchers to achieve more robust and realistic statistical results for the purpose of calibration and validation in urban growth simulation. All the processes in the spatial sampling are automated and connected with each other. Also, spatial sampling can be conducted according to the spatial (e.g., which county), temporal (e.g., which year) and attribute (e.g., which land use type) criteria, which incorporates Monte Carlo testing in the computational framework.

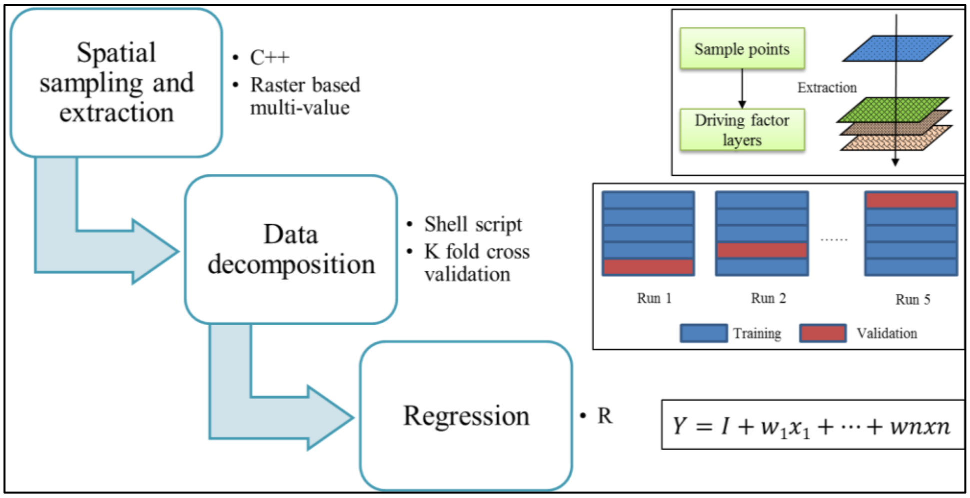

Figure 6. Workflow of spatial sampling and validation

Our spatial sampling approach is composed of three components: spatial sampling and extraction, data decomposition, and regression (see Figure 5). In spatial sampling and extraction, a set of sample points are selected with user specified sample size and sampling strategy such as random sampling and systematic sampling. Also, based on different research objectives, sample size can be different with a specific ratio among sub regions or land cover types. After the generation of sampling points, associated spatial and economic attribute values are extracted from corresponding GIS data layers, and then organized into tables in a database according to the spatial index (see Figure 6).

countyID	TransType	row	column	unchange	change	dp	d2stream	strDensity	d2road	d2city2	d2city6	d2city40	slope	dem
1	-1	3916	3791	1	0	46	201.246	0.00228108	210	3955.45	23586.4	38738.1	4.75892	225.374
1	-1	4025	3992	1	0	4	732.393	0.000945859	1440	4546.76	21816.7	39832.1	1.73775	246.301
1	-1	3880	3942	1	0	0	234.308	0.00176331	4353.72	5625.7	25036	42216.2	4.00224	228.478
1	-1	3748	3450	1	0	0	212.132	0.00142452	3624.42	14604.2	24299.1	39135.5	0.884344	276.578
1	-1	4159	3959	1	0	0	674.166	0.00181371	2002.6	5070.35	21613.3	36256.1	2.18532	233.981
1	-1	4260	3328	1	0	0	295.466	0.00225508	1682.41	9508.15	9508.15	23369.3	1.93517	236.856
1	-1	3704	3356	1	0	0	362.491	0.00145618	4729.45	12376.5	24457.1	39754.8	3.91318	260.394

Figure 7. An example of sampling table

Then, in the data decomposition part, *k* fold cross-validation approach is applied to create training and validation datasets for the validation purpose with the following steps: 1) Shuffle all records in the sampling table; 2) Divide the sampling table into *k* sub files (*k* is specified by users); 3) Select a single file as validation data, and merge the remaining *k-1* files as training data; 4) Repeat step 3 for *k* times to make sure each sub file is used exactly once as the validation data.

Based on the training and validation dataset created in the second step, statistical approaches, for example logistic regression, can be applied to analyze the relationship between converted land cover cells and associated spatial and economic characteristics. Furthermore, significance testing can be employed to identify the significant driving factors of land cover change for each regional unit (for example county), and determine the contribution of each driving factor, representing with a weight matrix. The regression equation can be derived with statistics software such as R.

4.3 Research objective 3 methodology

4.3.1 ABM urban growth simulation

To evaluate the proposed computational framework in this work, a large-scale ABM integrated model is developed to simulate the land use and land cover change in urban growth. The ABM integrated model simulates land cover transition mechanisms, and predicts spatial pattern and allocation of various land cover types in urban growth for the future. The ABM integrated model comprises of three interacting components: demand, CA (cellular automata) and ABM (agent-based modeling) components (see Figure 7).

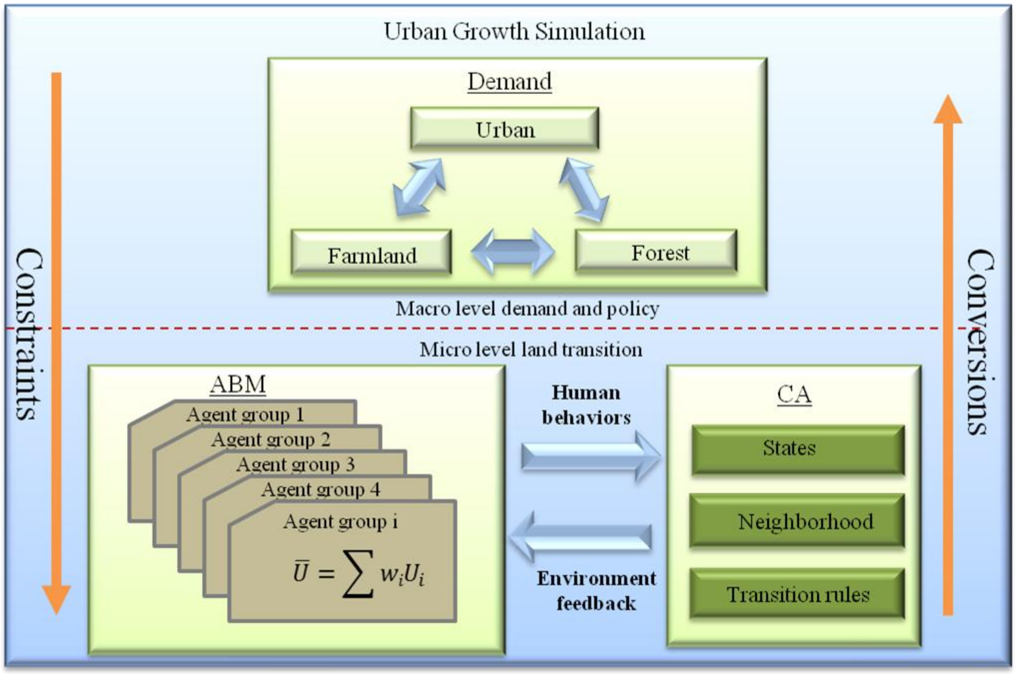

Figure 8. Agent-based modeling integrated urban growth model

In demand component, land demand (for example, number of cells) of different land cover types (urban, farmland and forest) for each region unit (e.g. county) are projected and allocated within each time step. Thus, though controlling land demand of land cover types, the ABM integrated model implements spatiotemporal constrains for different regions in a top-down manner. With respect to these constrains, ABM and CA are in charge of determining the conversion of individual land cover cell from one land cover type to others in a bottom-up way. Specially, there are feedback loops existing between ABM and CA. Based on the environment represented by CA, ABM simulates heterogeneous human decision making process and behaviors in the urban growth process. On the other hand, with respect to the impact of human decision making from

ABM, CA simulates land cover change based on the spatial interactions among land cover cells.

For each region in the urban growth simulation, the model randomly selects a developer agent. The agent will have a suitability map of development based on its own decision rule derived from empirical data. Cells with high suitability values are more likely to be converted to urban area. Once one cell is converted, its neighborhood cells will be examined in a patch growth algorithm. The neighborhood cells will be converted to urban area if their suitability values are beyond a threshold value defined by users. The model will keep selecting new agents to develop urban area until the number of converted cells meets the demand. A model parameter development pressure is used to represent the urbanization level of each cell's neighborhood. The development pressure parameter is dynamically updated at each time step.

- Environment

The environment of urban growth simulation is represented by a CA model which is guided by the principle of self-organization. The landscape of urban growth simulation consists of a grid of cells. Each cell has a set of states representing associated attributes, for example land cover types. The neighborhood of each cell is determined by the neighborhood pattern, which could be, either exactly immediate neighborhoods (first- or second-order neighborhoods) or distance based neighborhoods, designed according to various simulation purpose. Based on the transition rule f, states of each cell in next time step S_t is changed based on current state values of itself S and its proximity to neighborhoods S_n, as well as the human decision H (see equation 9)

$$S_t = f(S + S_n + H) \tag{9}$$

The newly converted cells can dynamically impact the future land cover change by providing positive feedbacks on their neighborhood. At the end of each time step, the ABM integrated model can estimate the influences of newly converted cells by updating the development pressure values of cells. The development pressure representing the degree of urbanization in a neighborhood area is a dynamic spatial variable of the human decision making process in ABM. The more land development occurs in surrounding cells, the bigger the development pressure is. In additional, I assume that the influences of newly developed cells to undeveloped cells follow a distance-decayed manner. Therefore, weight matrix approach can be designed and applied in the calculation of development pressure for different sub regions. The CA model converts cells from one specific land cover type to others in the study region by iterations. In this way, spatial interactions among land cover cells at the local scale can give birth to specific spatial patterns at a larger scale.

- Agents

With regard to different roles in the human society, there are various types of individuals with heterogeneous preferences and human behaviors affecting the process of land use and land cover change in urban growth. For example, at a high level, local government plans and regulates land use and land cover transition behaviors with zoning regulations. At low levels, people within conservation groups would like to protect specific habitats of animal far away from urban area, while real estate agents prefer to make benefits by developing residential or commercial facilities where expand existing urban area.

Within ABM of our integrated model, each group of agents is in charge of one single specific land cover transition (e.g., from forest land to residential area). Even within the same group, agents' preferences could be very different, which represents the heterogeneity of agents in same agent group. For example, with the same purpose of converting forest to residential area, someone likes to live near downtown close to public facilities for example hospitals, while someone else prefers to build an individual house in a suburban area for more private space. Based on their own preferences and beliefs, agents convert a fixed amount of land cells to corresponding land cover type in each time step. Since the land use and land cover change simulation is conducted at the regional unit level (e.g., county) in the study area, the decision rules of different agent groups, representing their preferences and beliefs, are simulated with utility functions yielded from statistical approach, based on empirical data for each regional unit.

In the process of land cover transition, agents determine where and when a transition occurs based on their utility functions, which are a multi-criteria evaluation problem formed as the following general equation:

$$S_i^j = C_i \sum_{n=1}^{k} (X_n W_n) \tag{10}$$

where S_i^j is the suitability value of cell i for agent j, determined by the production of a set of environmental and economic factors $X = (X_1, X_2\dots\dots, X_k)$ and their corresponding weights W derived from statistical approaches, as well as a set of binary constraints C_i based on the zoning relations or land policies from local government.

The converting probability of each cell in study area is estimated based on current cell's suitability determined by the agents' utility function, which could be very different among agents. Instead of one single suitability map for the whole study area, an individual agent is only capable of estimating a limited number of cells by randomly adding a list of cells into their candidate pools, according to the decision-making theory of bounded rationality (Mason, 2005). With respect to the estimated suitability value, only cells whose values are over a threshold, customized for various development scenarios, can be kept and sorted in the candidate pool. Then, agents will stochastically pick cells based on their ranks in the pool. Therefore, the selected cell to be converted may not be the cell with the best suitability, but relatively good one within the candidate pool. After the selection, the selected cell is converted to the corresponding land cover type of agent.

CHAPTER 5: CASE STUDY

In this work, I apply our computational framework in a large-scale urban growth simulation to model urban growth in North Carolina from 1992 to 2001. The main target of this urban growth simulation focuses on the rural to urban transition of landscape taking advantage of a patch-growing algorithm. In the case study, I use our parallel approach of model calibration to automatically calibrate patch parameters. Based on the empirical data, I also build a surrogate-based model to estimate the computing performance according to the relationship between spatial characteristics/content and computing time in model calibration. Furthermore, the computing performance of our proposed framework is evaluated in this case study.

5.1 Study area and data

In order to estimate the computing performance, I apply our surrogate-based computational framework to model the land use and land cover change in the whole state of North Carolina, USA. North Carolina is located in the southeast part of United States (see Figure 8). Consisting of 100 counties, North Carolina has a total area of 139,390 km^2. Urban areas have rapidly grown in the past twenty years. According to the national land cover data from USGS, the percentage of urban area increased from 4.19 to 10.4 from 1996 to 2011, while the percentage of forest decreased from 64.76 to 56.13. Raleigh and Charlotte are the two largest metropolitan areas.

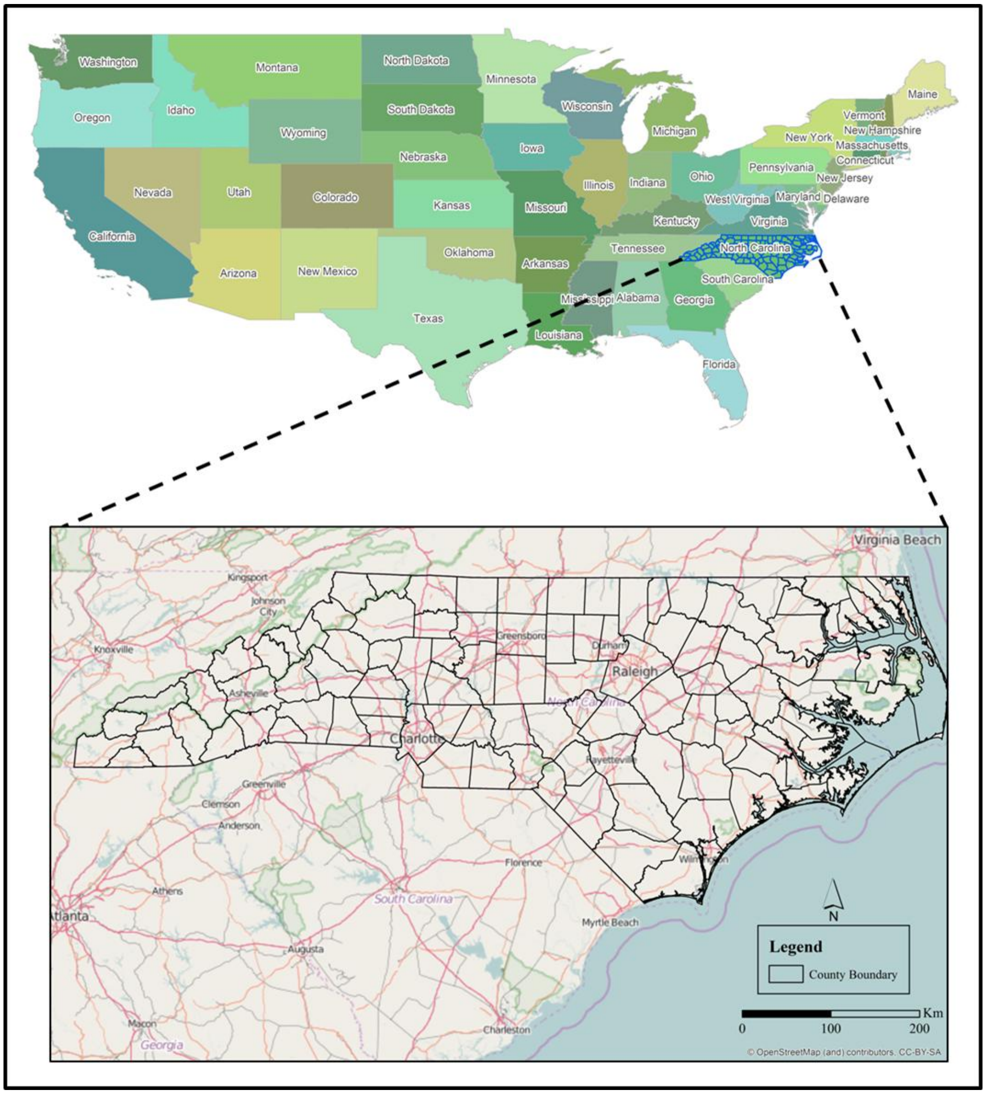

Figure 9. Map of North Carolina

Based on requirements of the urban growth simulation, I have collected social economic and environmental data for North Carolina, organized in the following table (see Table1). The population and road network data are from TIGER (Topologically Integrated Geographic Encoding and Referencing) census data (https://www.census.gov/geo/maps-data/data/tiger-line.html). I have downloaded county boundary shapefile of North Carolina from NCDOT (North Carolina Department of Transportation) and shapefile of streams from NHD (National Hydrology Dataset; https://nhd.usgs.gov/). NLCD (National Land Cover Databased; https://www.mrlc.gov) data are used to analyze land cover change with a spatial resolution of 30 m by 30 m. The shapefile of cities are from ESRI GIS database.

Table 1. Table of data used in this dissertation

Data	Spatial Resolution	Temporal Resolution	Source
Population	County	1992, 2006, 2011	Census
Boundaries	County	Present	NCDOT
Land use	30 m $\times$ 30 m	1992, 2001, 2006, 2011	USGS
Road	Road	2013	Census
Hydrography	Stream	2015	NHD
City	City	Present	ESRI

5.2 Parallel calibration of urban growth simulation

I designed a scientific workflow approach of spatiotemporal model calibration in our computational framework to leverage the computing power of cyberinfrastructure (see Figure 9). The approach can automatically generate and divide the combinations of model parameters into a set of subsets each computing on a single processor. The approach consists of six main steps for each subset of parameter combinations: parameter adjust, simulation run, region extraction, pattern analysis, performance evaluation, and create parameter file. These six steps are executed in an iterative way. First, based on the order of parameter combinations, I choose a combination of parameters to configure the spatiotemporal model. Then, run the simulation with the selected parameter combination and extract simulated landscape for each analysis unit (defined by user, e.g., county) in pattern analysis.

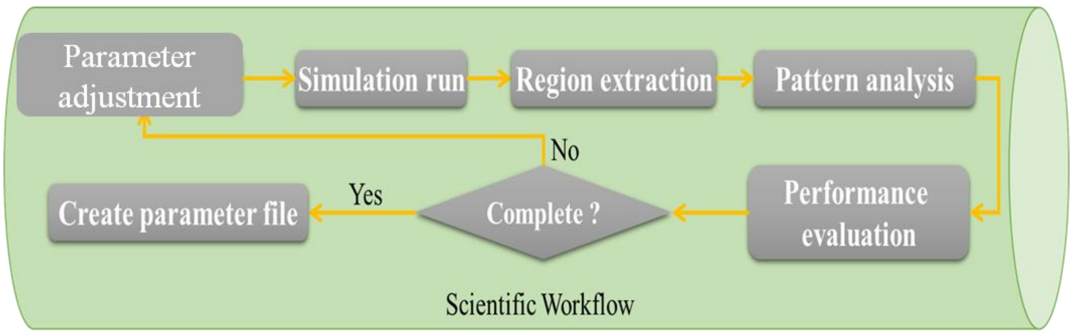

Figure 10. Workflow of urban growth model calibration

The following step is pattern analysis which conducts spatial analysis for the extracted simulated landscape pattern. The smallest component in our landscape pattern analysis is a patch consisting of contiguous pixels with the same land cover category in the simulated landscape. The spatial characteristics of the simulated landscape can be quantitatively represented by a set of landscape metrics based on the associated attributes of patches, such as size and shape (Turner et al. 2001). With respect to varied research questions, we can choose different landscape metrics (e.g., patch size, shape index, contrast) at three scales: patch, class, and landscape scales. Through these landscape metrics, we can quantitatively analyze the landscape patterns and structures (Kupfer

2012). After deriving the landscape metrics for simulated and observed data, we can evaluate the performance of the current parameter combination by estimating the difference between simulated and observe landscape patterns.

To quantify the difference, we need to derive histograms of landscape metrics for each analysis unit. And the difference between simulated and observed landscape patterns can be indicated by the MSE (Mean Square Error) of two histograms. MSE is calculated with the following equation:

$$MSE = \sum_{1}^{\mathrm{n}} (\mathrm{y}_{\mathrm{simulated}} - \mathrm{y}_{\mathrm{observed}})^2 / \mathrm{n} \tag{11}$$

where $y_{simulated}$ is simulated value, while its observed value is $y_{observed}$. n is the total number of bins in the histogram.

Smaller MSE indicates that the simulated landscape pattern is closer to the observed landscape pattern. In other words, the smaller the MSE is, the better the simulation performance of the parameter combination does. The approach will go through all the parameter combinations in the subset and store their MSEs. Once the approach completes the calculation for all subsets of parameter combinations, the MSEs of parameter combinations will be collected together and sorted. The best parameter combination with smallest MSE will be selected and recorded by the approach. The whole process is encapsulated into a scientific workflow which can automatically organize, manage, execute these steps, as well as recording the parameter, input and output for each step.

5.3 Implementation

I illustrate the implementation of our parallel framework of simulation calibration in this section. I use ESRI ArcGIS (version 10.4) to process the spatial dataset required by our framework, and visualize the result of simulation. To support spatial statistical analysis, I employ an open-source software package R (version 3.12). The latest version Fragstats (version 4.2) is the landscape metric program integrated into parallel framework of simulation calibration. I use Python language to organize and assemble GIS data and spatial operations together in scientific workflow.

In terms of high performance computing resource, I used one Linux computing cluster of Copperhead and one Windows computing cluster of Sapphire at University of North Carolina. The Linux has 59 computing nodes with 944 computing cores. The windows computing cluster composes of 20 computing nodes. Each computing node has two CPUs (Intel Core 2 Duo CPU with 3.00GHz of clock rate), associated with 4GB of memory. I installed Windows Server 2012 as the operating system on each node of the HPC cluster. For task scheduling purpose, HPC Pack 2012 Cluster Manager supports the management and allocation of tasks among window computing cluster.

CHAPTER 6: EXPERIMENTS

6.1. Hypotheses

In this work, I aim at investigating how to build surrogate-based approach to represent the relationship between spatial characteristics/content and computing intensity, which, in turn, can be applied in our computational framework to provide the estimation of computing performance of parallel spatiotemporal simulations. This estimation of computing performance can guide the design and implementation of parallel strategies of parallel spatiotemporal simulations to efficiently utilize the high performance computing power of cyberinfrastructure. Therefore, I probe the application of surrogate-based approach in overcoming computationally intensive challenge of spatiotemporal model calibration by means of appropriately predicting the computing performance within the context of urban growth in North Carolina.

The relationship of spatial characteristics/content and computing intensity should be captured and simulated by surrogate-based approach in parallel spatiotemporal simulation**s**. Regarding this, I identify three specific hypotheses to explore how proposed surrogate-based computational framework overcomes the computationally intensive challenge in parallel spatiotemporal simulations:

Hypothesis 1: Computing intensity will be correlated with spatial characteristics/content in spatiotemporal simulations. We usually apply spatiotemporal simulations to mimic the underlying mechanism of urban growth. In particular, various model variables in spatiotemporal simulations are designed to represent corresponding spatial characteristics/content (such as area of study region, land use configuration, amount of non-urban to urban conversion) of urban growth phenomenon. These spatial characteristics/content variables play an important role in affecting the computing intensity of parallel spatiotemporal simulations in a specific study region. As a result, there will be a significant relationship between computing intensity and model variables used in simulation runs. The utility of surrogate-based approach can capture this relationship between computing intensity and model variables. Through surrogate-based approaches, we can predict the computing performance based on model variables used in spatiotemporal simulations.

Hypothesis 2: Sample size and the type of surrogate-based approaches will impact the prediction ability for computing intensity of spatiotemporal simulations. The prediction ability of surrogate-based approach can be reflected by the prediction accuracy that is measured by the difference between estimated and real computing performance in simulation runs. Depending on the relationship between computing intensity and spatial characteristics/content, various sample sizes and types of surrogate-based approaches will have different prediction ability in parallel spatiotemporal simulations. Consequently, the selection of sample size and the type of surrogate-based approaches will critically affect the prediction performance for computing intensity in proposed computational framework.

Hypothesis 3: The application of surrogate-based approaches will improve the computing performance of parallel spatiotemporal simulations. With support of appropriate surrogate-based approach, our computational framework can accurately predict the computing performance (e.g., speedup, efficiency, and scalability) of each computing task within a parallel spatiotemporal simulation. Thus, the information of computing performance can guide the design and implementation of parallel strategies, such as load balancing and task scheduling, to more efficiently leverage the high performance computing power of cyberinfrastructure in parallel spatiotemporal simulations.

In sum, these three specific hypotheses are designed to help us examine 1) whether or not surrogate-based approaches can predict the computational performance of parallel spatiotemporal simulations; 2) how sample size and whether or not the type of surrogate-based approaches can affect the prediction accuracy of computing performance; 3) how to efficiently utilize the cyberinfrastructure to facilitate parallel spatiotemporal simulations by incorporating the appropriate surrogate-based approach in our computational framework.

6.2. Experiments

In this section, I designed three experiments to test three hypotheses in section 6.1 respectively. I applied proposed computational framework to support the calibration of a large-scale spatiotemporal model which simulates the urban growth of North Carolina from 1992 to 2001. Empirical spatial datasets were employed in the model calibration. In terms of computing intensity, each analysis unit (a single county in our case) is differentiated with others according to three variables of spatial characteristics/content in parallel model calibration: 1) the variable of number of cells indicates the area of current analysis unit; 2) the variable of urban percentage can reflect the configuration of land use pattern; 3) the variable of demand represents total amount of land conversion from non-urban to urban land cover at each time step (one year in our case).

These three model variables can work together to impact the spatial characteristics in analysis unit. The spatial characteristics mainly focus on three parts: spatial dependency, stationarity, and isotropy. In terms of number of cells and demand, if the number of cells is very close to the demand, new development patches will be more close to each other. Because there is little open space area to be converted, new development patches will overlap with each other. Therefore, it will result in more spatial dependency in this analysis unit. Otherwise, there is less spatial dependency. Also, the larger the area of analysis unit is, the more spatial heterogeneity there exists. For the urban percentage variable, the existing urban area has spatial impact to its neighborhood, represented by development pressure in our urban growth simulation. Thus, cells close to existing urban area are more likely to be converted to new development patches, which also can change the spatial dependency in analysis unit.

According to my proposed surrogated-based approach, values of model variables used and associated computing intensity are recorded for in total 8,100 simulation runs of model calibration. These values of model variables and associated computing intensity are used to construct surrogate-based models and examine three hypotheses within three corresponding experiments.

6.2.1 Relationship between computing performance and spatial characteristics/content

A patch-based urban growth model (proposed in section 4.3.1) is implemented to simulate the urban growth in North Carolina in this case study. Before employing an urban growth model, we need to calibrate the model based on empirical land cover change data to make it as close as possible to the reality. In this urban growth model, patch parameters of the patch growth algorithm should be calibrated. Because combinations of these patch parameters control the size and shape of patches developed in urban growth simulations. I applied proposed parallel calibration approach (described in section 5.2) to calibrate these patch parameters based on the empirical data. Instead of a single global value at aggregated level, I calibrate these patch parameters at a very fine scale (county level). In other words, a county is the smallest analysis unit in landscape pattern analysis. Each county will have its own unique set of patch parameters in the patch-growing algorithm to represent the heterogeneity in large-scale urban growth.

Three patch parameters need to be calibrated for the urban growth simulation, including discount factor, patch mean, and patch range. The parameter of discount factor controls the size of new developed patches in the simulation, which has a range [0, 1] with an interval of 0.01. The larger the discount factor, the more land cells developed in a single patch. While the parameter of patch mean represents the shape of patch, patch range restricts the fluctuation range of the patch mean. Patch mean and patch range work together to determine the compactness of a new developed patch. Both patch mean and patch range are calibrated from 0 to 1 with an interval of 0.1. When patch mean approaches to 1, the patch shape will be compact. Otherwise, the shape of patch tends to be a linear shape.

I used 1996 and 2001 NLCD land cover data (downloaded from USGS: http://www.mrlc.gov/) as observed data. Based on the NLCD data, I generated the observed land cover metrics on patch level for each county. Since the discount factor is only related to the size of patch, I first calibrated the discount factor of each county with landscape metrics of patch size. Figure 11 shows the calibrated discount factor of each county. Patch mean and patch can affect each other in patch growth process. Therefore, we have to calibrate them together. For each combination of these two parameters, I run the urban growth simulation from 1996 to 2001, resulting in 8,100 simulation runs. Then, landscape pattern analysis is conducted by comparing new developed patches of simulated result with patches of observed data for the year of 2001, with respect to the landscape metrics of parameter-area ratio.

The calibration results of patch mean and patch range are represented by Figure 12 and 13. Figure 14 represents the simulation result of urban growth in the year 2001.From these calibration results, we can see that the urban percentage of a county (see Figure 10) has a substantial impact on the calibration of patch growth parameters in this county. The correlation coefficient between discount factor and urban percentage for 100 counties is -0.4. In other words, the larger the urban percentage is, the smaller the discount factor is, and the smaller size the new developed patch is. Also, we can observe that patch means of metropolitan area, such as Charlotte, Raleigh-Durham, Asheville, and Wilmington, tend to be relatively small. Compared to counties with low urban percentage, counties with high urban percentage have relatively smaller open space but higher demand for the new urban patches. Consequently, it happens more in counties with high urban percentage that new developed patches overlap with each over, resulting in the smaller size and more compact shape observed in these counties.

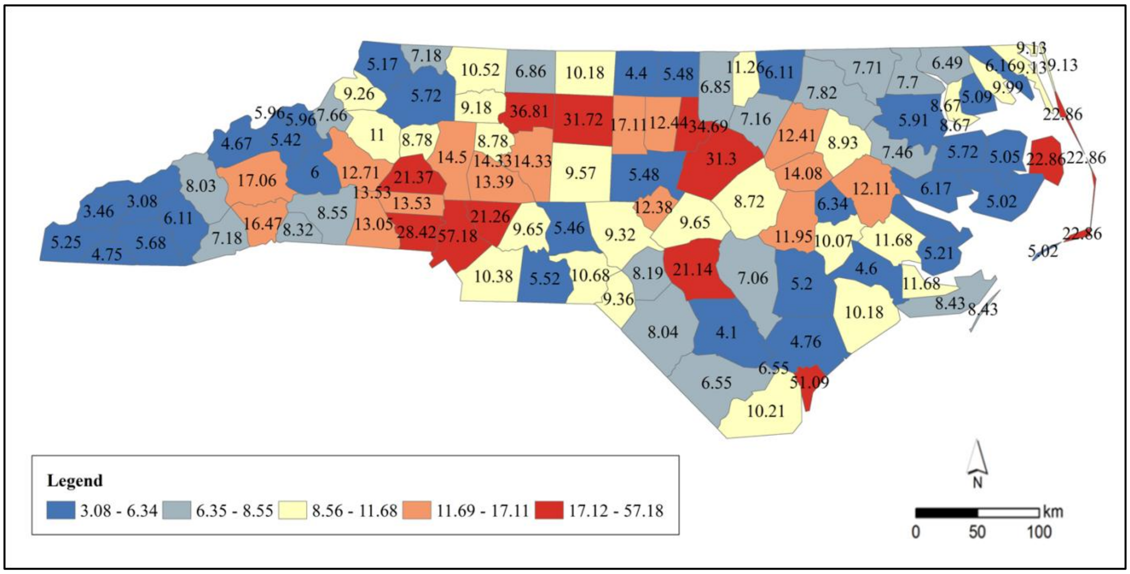

Figure 11. Urban percentage of NC for year 2001

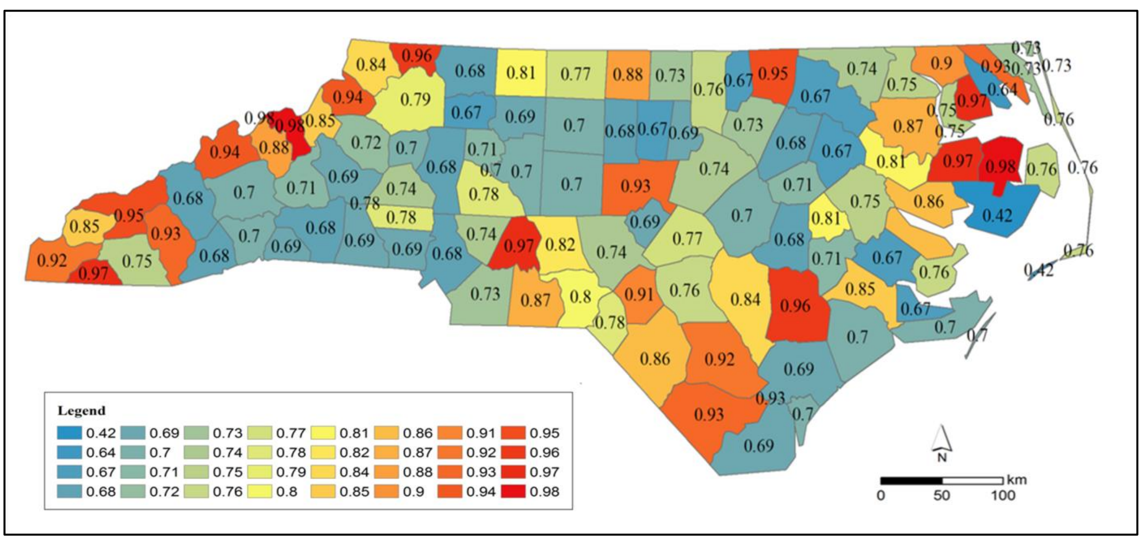

Figure 12. Calibrated discount factor of each county in NC

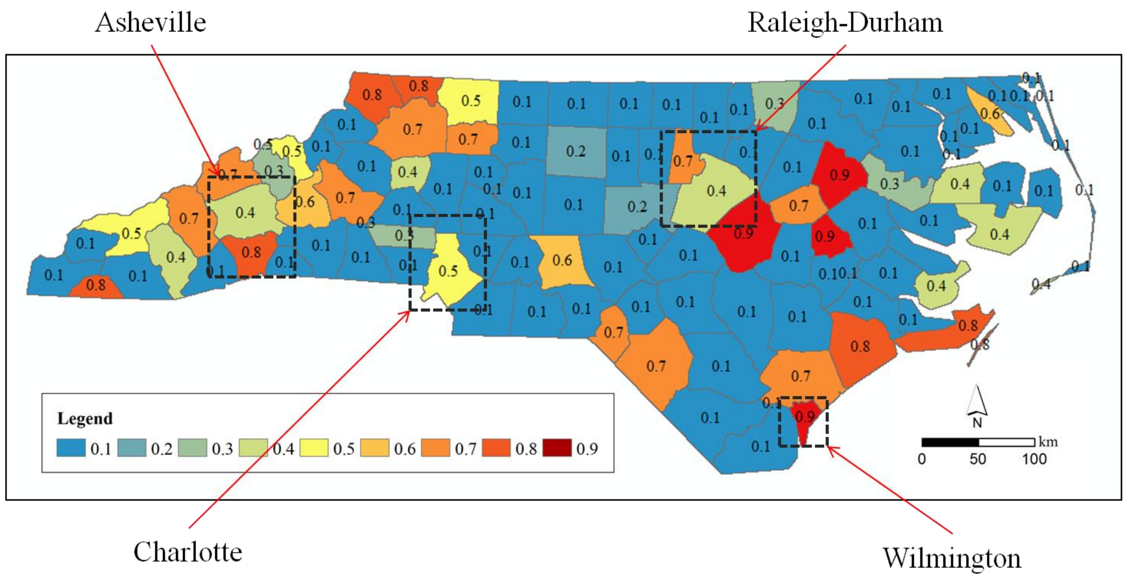

Figure 13. Calibrated patch mean of each county in NC

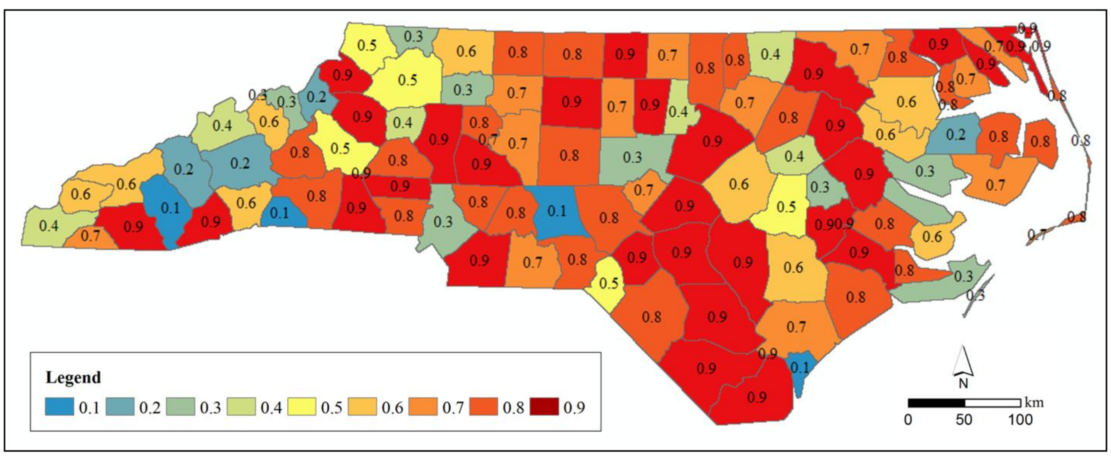

Figure 14. Calibrated patch range of each county in NC

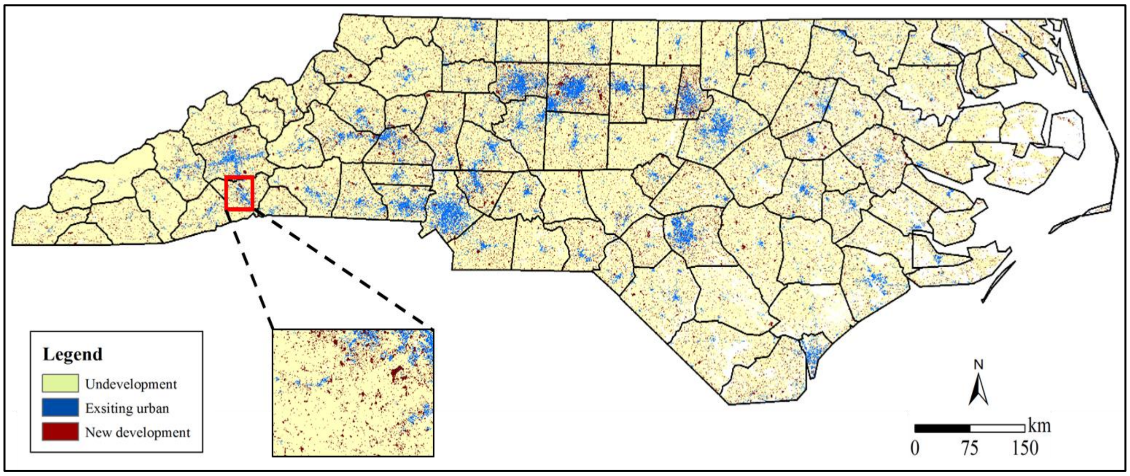

Figure 15. Simulated urban growth in NC for the year 2001

In order to probe the relationship between computing intensity and spatial characteristics/content, I recorded values of three model variables used (number of cells, urban percentage, and demand) and associated computing time of 8,100 simulation runs. There are in total 100 counties in North Carolina. Each county has 81 (9 × 9) simulation runs for 81 combinations of patch mean and patch range. For each county, average computing time of 81 simulation runs is calculated to estimate the relationship between computing intensity and spatial characteristics/content. While Figure 15 is the scatter plot representing the relationship, Table 2 demonstrates correlation coefficients between corresponding spatial characteristics/content variables and computing time.

Since the variable of number of cells directly determines the size of input data to be processed, it has the largest correlation value of 0.937 with average computing time. Following number of cells, the variable of demand also has a very strong correlation with a value of 0.502. Compared with number of cells and demand, the variable of urban percentage has a relatively weaker correlation with average computing time (0.206). Reported by experiment result, all three model variables are positively correlated with average computing time. Therefore, the result of this experiment suggests that we can accept the hypothesis 1, that is, computing intensity will be correlated with spatial characteristics/content in spatiotemporal simulations.

Table 2. Correlation of average computing time and model variables

	Correlation
Number of cells	0.937
Urban percentage	0.206
Demand	0.502

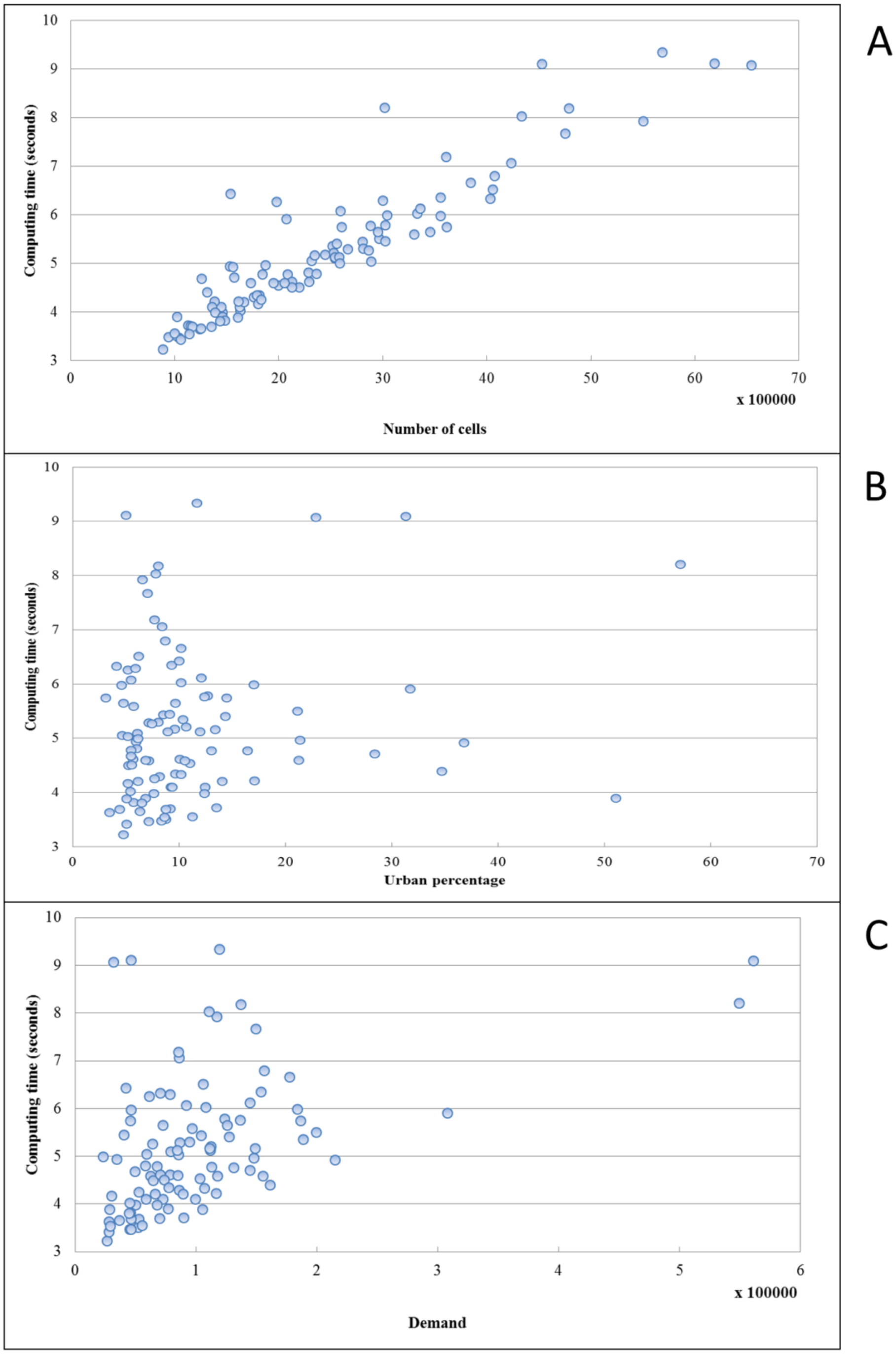

Figure 16. Scatter plot of computing time and model variables of urban growth (A: number of cells; B: urban percentage; C: demand)

6.2.2 Impacts of sample size and types of surrogate-based approach

Result of the previous experiment demonstrates that there is an existing correlation between computational intensity and spatial characteristics/content. Therefore, in this experiment, I investigate how sample size and type of surrogate-based approaches can affect the estimation of computing performance when we build surrogate-based models. I applied two common used types of surrogate-based models (polynomial regression and Kriging) to capture the relationship between computational intensity and spatial characteristics/content, and make comparison in terms of prediction ability for computing performance.

The analysis unit of model calibration is a county in this experiment. In other words, with support of surrogate-based approaches, we can estimate the computing performance of a county based on the relationship between computing performance and spatial characteristics/content variables of this county. Similar to the previous experiment, the spatial characteristics/content variables of each county in the urban growth model include number of cells, urban percentage, and demand.

In order to examine the impact of sample size and the type of surrogate models, I applied empirical approach to build regression and Kriging surrogate-based models with different sample size in two treatment groups. For each group, there are in total nine treatments set up by varying sample size from 10%- 90% of the total 8,100 simulation runs, increasing with an interval of 10%. For each treatment, I conducted 100 random sampling using bootstrapping strategies with same sample size. With the same sampling dataset, each sampling randomly selects samples from the total population as training dataset to train the surrogate-based model. With respect to the accuracy of prediction, I applied each surrogate model to predict the computing time of all computing jobs. Based on the real computing time of all computing jobs, RMSE (root-mean-square error) of each surrogate-based model is calculated with the RMSE equation in Chapter 3. The average RMSE based on 100 sampling within each treatment was calculated and used as an indicator of the prediction ability of current sample size. Since RMSE indicates the difference between real and estimated computing time, thus, the smaller the RMSE is, the better the prediction performance is.

Based on the average computing time obtained and associated values of spatial characteristics/content variables in sampling dataset, I constructed the linear regression surrogate-based model as follow:

$$\mathrm{T}_i = \mathrm{I} + \sum_{\mathrm{j}=1}^{\mathrm{k}} \mathrm{w}_j\, \mathrm{x}_j \tag{12}$$

where T_i is the estimated computing time of computing job *i*. *I* is the intercept. *k* indicates the total number of spatial characteristics/content variables. *x* are spatial characteristics/content variables, while *w* are the associated weights of corresponding variables.

Based on the relationship between sampled values of model variables and their computing time, Kriging can interpolate the computing time for all parameter space. Since I repeated the same spatiotemporal simulation for different variable combinations, I choose Ordinary Kriging approach which assumes the underlying mechanism is stationary. The previous experiment result indicated that number of cells and demand can significantly affect the computing time in each county. Therefore, I used Ordinary Kriging approach to build a surrogate-based model of computing time based on variables of number of cells and demand.

The Ordinary Kriging surrogate-based model generates a response surface to estimate the computing time for different combinations of number of cells and demand (see Figure 16). Compared to linear regression, the Ordinary Kriging surrogate-based model can better visualize the relationship between input and output with the response surface. Based on Figure 16, we can observe a clear increasing pattern of computing time with the increase of number of cells and demand.

In terms of implementation, I use R software to conduct experiment for linear regression treatment group. For Kriging treatment group, I build a scientific workflow using Python to conduct sampling, construct Kriging models, predict computing performance, and calculate RMSEs. For each sampling, this workflow first randomly selects sample points from sample dataset. Through ArcPy package in python, a function of “KrigingModelOrdinary” of ArcGIS software is triggered to conduct spatial interpolation and predict the computing performance based on selected sample points. A spherical model is applied to calculate semivarigram in Kriging. Then, RMSE of this sampling is calculated based on observed and predicted computing performance.

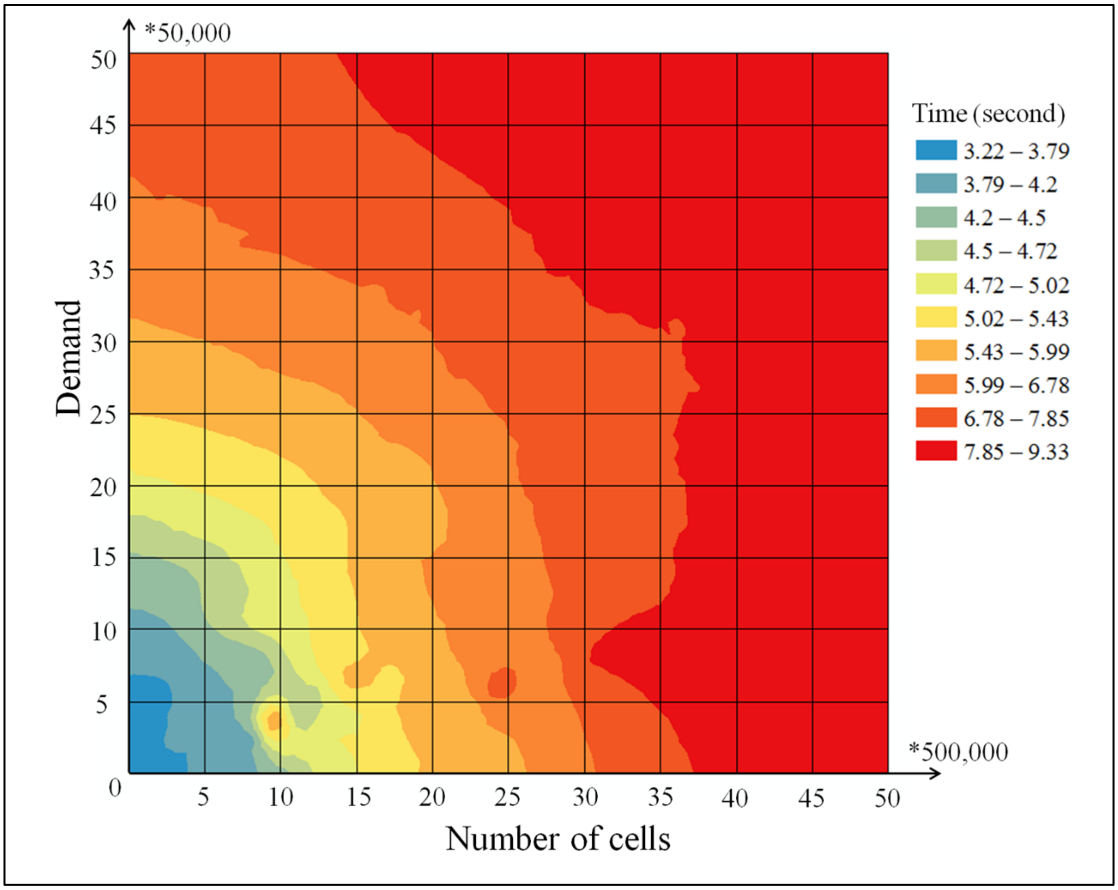

Figure 17. An example of response surface of computing time in Kriging surrogate-based model

Table 3 reports the average RMSE of surrogate-based models in two treatment groups with different sample size. In order to present the trend and make a comparison, I plotted average RMSEs for all nine treatments in Figure 17 and Figure 18. Based on Table 3 and Figure 17 and 18, we can see the average RMSE trends to be decreased for both linear regression and Kriging surrogate-based models, when increasing the sample size. In other words, the increase of the sample size leads the improvement of prediction performance of surrogate-based models. In terms of the trend, two types of surrogate-based models present different patterns: the average RMSE of linear regression sharply jumped from 0.39 to 0.31 when we increased sample size from 10% to 30%. Then the decrease trend of RMSE becomes relatively stable around 0.30. Compared to linear regression type, Kriging surrogate-based model has a better prediction performance in all treatments. And the slope of decrease trend turns to be larger when the sample size beyond 40%, which means the prediction performance of Kriging is more sensitive than linear regression in our case.

In this experiment, we apply *t*-test to statistically test if two approaches are significantly different from each other. For each treatment, there are 100 values of RMSEs for each approach as the population of the *t*-test. The null hypothesis is that there is no difference in RMSE between two treatment groups. The significance level is 0.05. Table 4 reports *p* values of *t*-tests for two treatment groups with different sample size. Based on table 4, for treatment 1 and 2, the *p* values are bigger than the value of significance level. Therefore, we fail to reject the null hypothesis there is no difference between two treatment groups. Conversely, for other seven treatments, we observed that the *p* values are much smaller than the significance level. Therefore, we reject the null hypothesis there is no difference between two treatment groups. In other words, the differences between two treatment groups are significant in treatment 3-9 with *p*-values smaller than significance level.

In summary, surrogate-based models with both linear regression and Ordinary Kriging can provide us very good estimation of computing performance in terms of RMSE. In comparison with linear regression, Kriging surrogate-based model can better predict the computing performance based on spatial characteristics/content variables.

With support of a response surface, Kriging surrogate-based model has an advantage in visualizing the relationship between model input and output. In addition, the larger the sample size is, the higher the prediction accuracy is, as a result, the more the cost is (since more simulation runs needed to obtain spatial characteristics/content variables and associated computing performance). Therefore, in order to determine the sample size, we should first explore how predict performance varies according to the change of sample size when building a surrogate-based model to predict the computing performance. Then we can determine the sample size based on the research objective and tradeoff between prediction accuracy and computing cost. We should choose the appropriate model type according to the sample size and their corresponding prediction performance. In this experiment, the comparison of different sample sizes and types of surrogate-based approaches suggests the acceptance of hypothesis 2, i.e., sample size and the type of surrogate-based approaches will impact the prediction ability for computing intensity of spatiotemporal simulations.

Table 3. Average root-mean-square errors of surrogate-based models for different sample size

Treatment	Sample size (percent of population)	RMSE (Linear regression)	RMSE (Kriging)
1	10	0.388	0.360
2	20	0.340	0.334
3	30	0.324	0.304
4	40	0.310	0.294
5	50	0.307	0.257
6	60	0.303	0.235
7	70	0.301	0.196
8	80	0.300	0.158
9	90	0.299	0.113

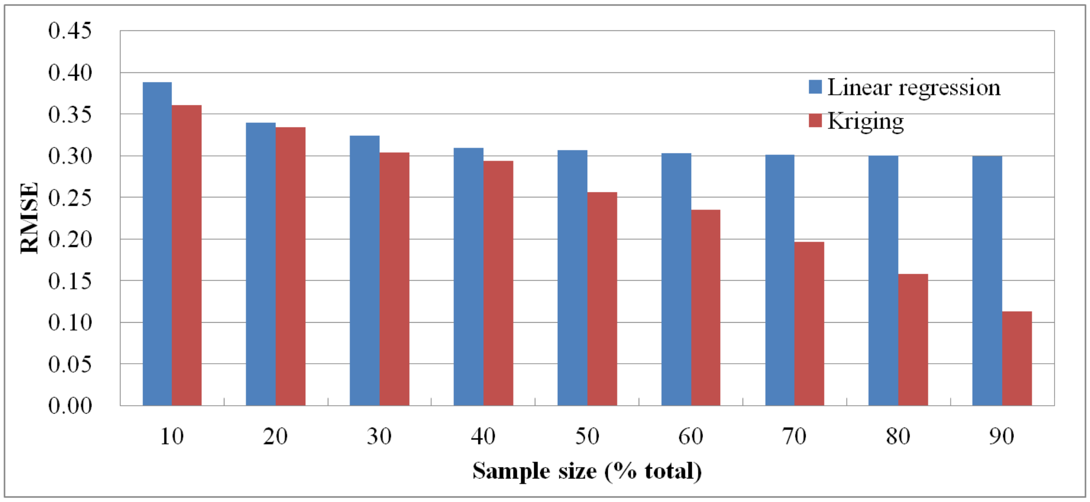

Figure 18. Average root-mean-square error of surrogate-based models using different sample size

Table 4. *P*-value of *t*-test for different sample size

Treatment	Sample size (percent of population)	*P*-value
1	10	0.152185
2	20	0.587881
3	30	0.000268
4	40	0.002291
5	50	9.11E-15
6	60	2.75E-19
7	70	5.91E-31
8	80	5.95E-38
9	90	9.52E-53

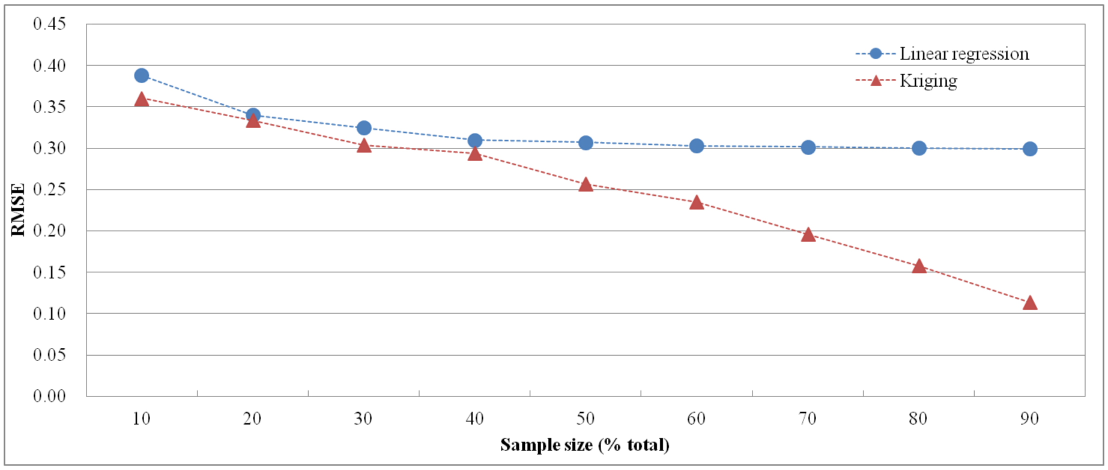

Figure 19. Average root-mean-square errors for linear regression and Kriging surrogate-based models

6.2.3 Impact of surrogate-based approach on computing performance

In the third experiment, we construct a Kriging surrogate-based model to predict the computing performance based on input of spatiotemporal simulations. The Kriging model applies a spherical model to calculate semivarigram with a set of optimized parameters (Number of lags: 12; Lag size: 65,049.62; Nugget: 0; Range: 520,396.99; Anisotropy: NO; Partial sill: 3.51). The computing performance predicted is used to guide the design and implementation of parallel strategies in parallel urban growth simulation. The real computing performance (speed up and efficiency) of our computational framework can be evaluated with corresponding metrics. Based on the evaluation of computing performance, thus, we can examine hypothesis 3 (i.e., the application of surrogate-based approach will improve the computing performance of parallel spatiotemporal simulations). Since the analysis unit is a county, I applied a domain decomposition strategy to divide the whole computation of calibration into a set of individual computing jobs based on county boundary. Consequently, I aggregated 8,100 simulations into 100 individual computing jobs in total. Each computing job calibrates patch parameters for each county by comparing the difference between observed and simulated patches for all combinations of parameters (e.g., 81 combinations in total for patch shape calibration). One single CPU takes 42,049.94 seconds (around 11.68 hours) to complete these 100 computing jobs. Our surrogate-based computational framework decreases the total computing time from 42,049.94 seconds to 2,719 seconds (about 45 minutes) with 18 CPUs (the upper limit of our windows cluster). We can see our parallel approach can provide a significant speedup of 15.517 with the efficiency of 86.2%.

In order to investigate the impact of surrogate-based approach on the load balancing of our parallel computational framework, I set up two treatment groups (with and without surrogate-based approach group) for the purpose of comparison. Both two treatment groups processed the same 100 individual computing jobs. The first group randomly allocated 100 computing jobs to CPUs to be used with even number of computing jobs. For the second treatment group, I applied the Kriging surrogate-based approach to estimate the computing time of each county based on values of spatial characteristics/content variables. According to the estimated computing time for each computing jobs, load balancing strategy was designed and implemented for the task scheduling process in group two to make the total estimated computing time more even among CPUs available. For each treatment group, I designed 9 treatments by varying the number of CPUs to be used in each treatment from 2 to 18, with an interval of 2. Speedup and efficiency were calculated and reported for both treatment groups.

I first compare the real and estimated computing performance in each treatment with load balancing (computing time, speedup, and efficiency; see Table 4). Figure 19 is the plot of real and estimated speedup from treatment 1 to 9. Indicated by Table 4 and Figure 19, we can see that the trend of real computing performance is well predicted and captured by our surrogate-based approach. Table 5 reports the results of computing time, speedup, and efficiency for the parallel calibration approach with and without surrogate-based approach groups. Figure 20 and 21 illustrate the trend of speedup and efficiency for both treatment groups. From Table 5 and Figure 20 and 21, we can observe that the speedups are increased and the efficiencies are dropped for both groups when we increase the number of CPUs used. Compared to the group without surrogate-based approach, the group with surrogate-based approach has better speedups and efficiencies in all 9 treatments. More important, the speedup of the group with surrogate-based approach is more close to the linear speedup (the theoretical upper limit of speedup; see Wilkinson and Allen, 2004). The range of efficiency of group without surrogate model is from 84.3% to 56.5%. The application of surrogate-based approach for load balancing improved the efficiency to the range of 99.2% to 81.7%. Since there is not any information of computing performance of each computing job, random allocation has a high possibility to put computing jobs with large computational intensity to the same CPU, which leads to the downgrade of computing performance of parallel spatiotemporal simulations. With the support of surrogate-based approach, our load balancing strategy can more evenly allocate computing jobs based on estimated computing performance, resulting in very good efficiencies over 81% in all treatments. Therefore, the results in this experiment suggest that we can accept the hypothesis 3, that is, the application of surrogate-based approaches will improve the computing performance of parallel spatiotemporal simulation.

Table 5. Estimated and real computing performance of parallel urban growth model calibration

Treatments	#CPUs	Time	Estimated Speedup	Efficiency	Time	Real Speedup	Efficiency
T1	2	21,213	1.989	0.994	21,274	1.983	0.992
T2	4	10,745	3.926	0.982	10,846	3.890	0.972
T3	6	7,336	5.751	0.959	7,292	5.786	0.964
T4	8	5,627	7.499	0.937	5,829	7.238	0.905
T5	10	4,453	9.475	0.948	4,628	9.117	0.912
T6	12	3,969	10.631	0.886	3,937	10.717	0.893
T7	14	3,525	11.971	0.855	3,521	11.984	0.856
T8	16	3,103	13.598	0.850	3,227	13.076	0.817
T9	18	2,732	15.446	0.858	2,719	15.517	0.862

Table 6. Comparison of computing performance of two treatment groups (Group 1: without surrogate-based load balancing; Group 2: with surrogate-based load balancing)

Treatments	#CPUs	Group 1			Group 2		
		Time	Speedup	Efficiency	Time	Speedup	Efficiency
T1	2	24,939	1.686	0.843	21,274	1.983	0.992
T2	4	14,161	2.969	0.742	10,846	3.890	0.972
T3	6	9,545	4.405	0.734	7,292	5.786	0.964
T4	8	7,619	5.519	0.690	5,829	7.238	0.905
T5	10	6,649	6.324	0.632	4,628	9.117	0.912
T6	12	5,559	7.565	0.630	3,937	10.717	0.893
T7	14	5,312	7.916	0.565	3,521	11.984	0.856
T8	16	4,451	9.447	0.590	3,227	13.076	0.817
T9	18	3,449	12.192	0.677	2,719	15.517	0.862

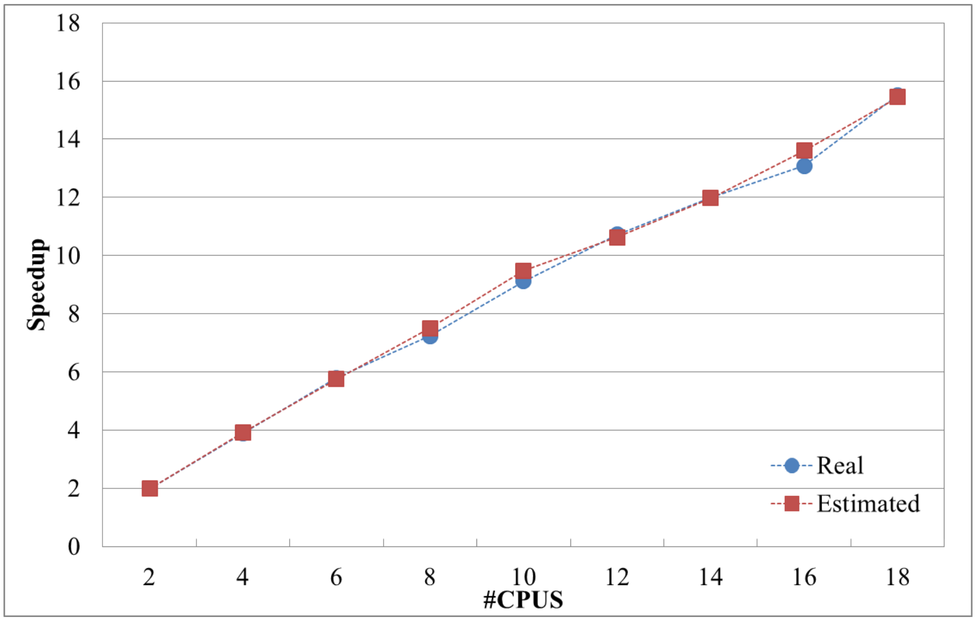

Figure 20. Plot of real and estimated speedup of parallel urban growth model calibration

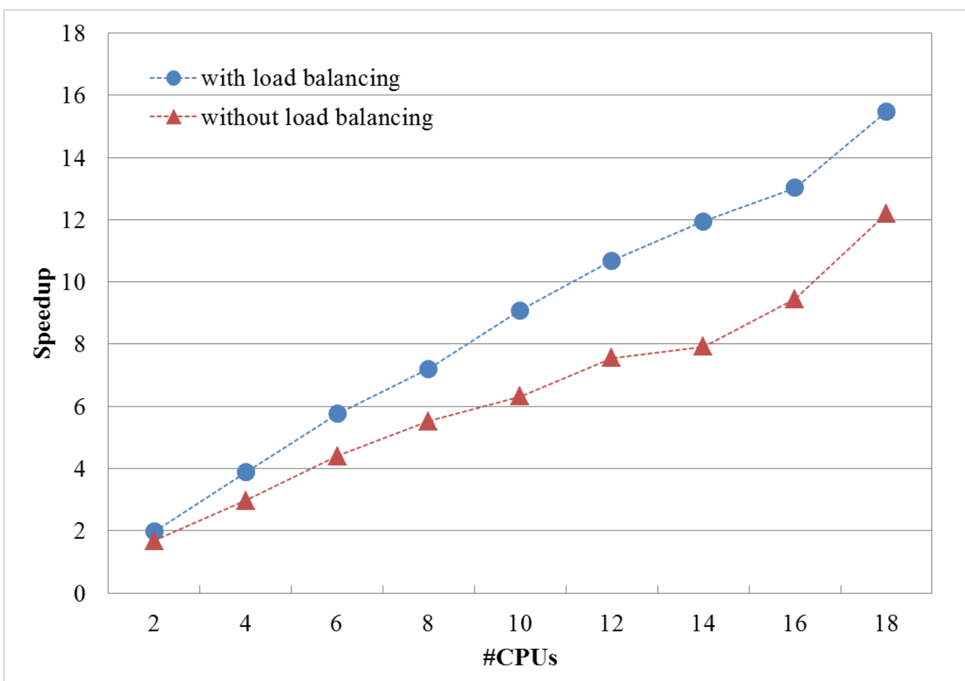

Figure 21. Plot of speedup of with surrogate-based model group (blue line) and without surrogate-based model group (red line)

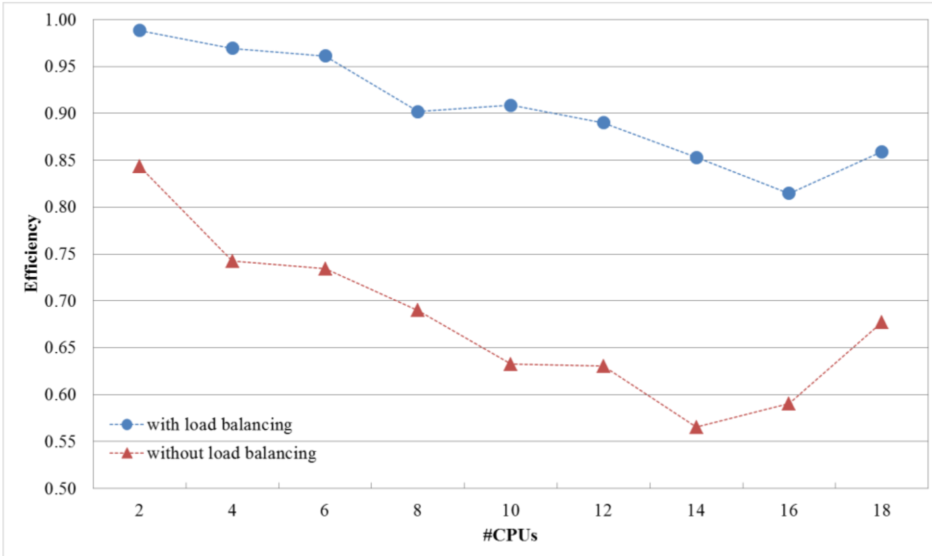

Figure 22. Plot of efficiency of with surrogate model group (blue line) and without surrogate model group (red line)

6.3 Concluding Discussion

In this study, the effect of surrogate-based approaches on more effective utilization of cyberinfrastructure is investigated by applying surrogate-based computational framework in the resolution of computationally intensive challenge in parallel spatiotemporal simulations, within the context of urban growth. According to experiment results, we can observe that the computing intensity of each analysis unit is highly correlated with its spatial characteristic/content variables used in spatiotemporal simulations, which is the foundation of surrogate-based approaches in this study. Without detailed knowledge of underlying mechanism and implementation of spatiotemporal simulations, surrogate-based approaches can apply empirical or heuristic methods to simulate this relationship between inputs (model variables) and output results (computing performance).

In particular, I focus on investigating that how to use empirical approach to capture the relationship, and then build surrogate-based models to predict the computing performance based on inputs of spatiotemporal simulations. Both sample size and the type of surrogate-based models can substantially affect the prediction ability for computing performance. Sample size should be determined with respect to the prediction accuracy required and the consideration of planned budget. On the other hand, different types of surrogate-based models have disadvantages and advantages according to a specific research objective based on their own mechanisms, which will result in significant differences in budget planning and implementation. Consequently, the selection of sample size and type of surrogate-based models, it turns out, plays an important role in the establishment of an appropriate surrogate-based approach to predict computing performance. Therefore, when dealing with large amount of spatiotemporal data or computationally intensive spatiotemporal simulations, we should choose the appropriate sample size and type of surrogate-based approach using model evaluation approaches, based on the characteristics of spatiotemporal data or the relationship of input and output within a spatiotemporal simulation.

The measures of computing performance in experiments provide the quantitative evaluation of the impact of surrogate-based approaches on accelerating parallel spatiotemporal simulation. In particular, the comparison of computing performance with and without the support of surrogate-based approach proves that the application of surrogate-based approach can substantially improve the efficient utilization of cyberinfrastructure, reflected by the superior speedup, efficiency, and scalability.

Regarding this, it is necessary to design and implement our parallel strategies with support of surrogate-based approach in parallel spatiotemporal simulations.

Generally speaking, results of experiments support hypotheses stated in this dissertation and suggest that our surrogate-based computational framework provides substantial support for accelerating spatiotemporal simulations by efficaciously leveraging state-of-the-art cyberinfrastructure.

CHAPTER 7: CONCLUSIONS

The study of complex geographic phenomena has been significantly improved by applying cyberinfrastructure in spatiotemporal simulations (Armstrong, 2000; Wang, 2010; Scheutz and Harris, 2012). The computing intensity of parallel simulations is hard to be estimated due to the complex simulation mechanisms. Therefore, it has been recognized that the efficient utilization of cyberinfrastructure is still challenging (Wang and Armstrong, 2009; Parry and Bithell, 2012; Shook et al, 2013). In this dissertation, I discuss the inadequacy of estimation and validation of computing intensity in existing work. Additionally, the existing methods cannot be generally applied in parallel simulations with different mechanisms. To overcome these limitations, I present a surrogate-based computational framework which integrates surrogate-based approaches, spatiotemporal simulations, and cyberinfrastructures. This general framework can facilitate the efficient utilization of cyberinfrastructure based on the accurate estimation of computing intensity, which makes substantial contribution on the application of cyberinfrastructure in solving complex geographic problems.

I concentrate on applying surrogate-based approach to capture and simulate the relationship between input spatial characteristics/content variables and output computing performance in a black-box and approaching way. Therefore, researchers, without background about high performance and parallel computing, are able to efficiently predict the computing performance of their spatiotemporal simulations. The estimated computing performance, in turn, can guide the parallelization of spatiotemporal simulations on cyberinfrastructure. Consequently, computationally intensive tasks such as model calibration, sensitive analysis, and model optimization can be better designed and implemented with respect to the estimated computing performance. Furthermore, research budget (time and funding) can be well planned.

I start this dissertation with discussing the significance of spatiotemporal simulations in the investigation of global complex geographic phenomena. In order to solve these phenomena at large scale, we need to take advantage of GIS, spatiotemporal simulations, and cyberinfrastructure. In Chapter 2, I discuss the existing work and literatures related to coupled human and natural systems, main stream spatial simulation modeling approaches, and the application of high performance and parallel computing in spatial data analysis and spatial modeling. CyberGIS holds great promise to provide great support in overcoming computational challenges for spatiotemporal simulations. With computational support of cyberinfrastructure, spatiotemporal simulations can mimic the underlying mechanisms of these complex geographic phenomena at large scale. However, the study of the efficient utilization of cyberinfrastructure within spatiotemporal simulations is inadequate. A key computational challenge exists in the integration of GIS, spatiotemporal simulations, and cyberinfrastructure, which is the appropriate estimation of computing performance.

With regards to this key computational challenge, I introduce surrogate-based approaches in Chapter 3. Surrogate-based approaches, which can appropriate estimate the computing performance of spatiotemporal simulations within cyberinfrastructure, are built based on the relationship between spatial characteristics/content and computing performance. Although surrogate-based models have been extensively applied to tackle computationally intensive issue in engineering design, they have been rarely used in spatial analysis and modeling domain. Through the literature review of surrogate-based models in other scientific domains, I explore how to apply surrogate-based models to support spatiotemporal simulations within cyberinfrastructure environments. I design a generalized sequential procedure which illustrates how to build and validate a surrogate-based model within a spatiotemporal simulation. With support of this sequential procedure, we can construct a surrogate-based model to aid the efficiently utilization of cyberinfrastructure in the study of complex geographic phenomena.

Proposed in Chapter 4, a surrogate-based computational framework integrates spatiotemporal simulations, cyberinfrastructure, and surrogate-based models as three main components. Compared to past work, the main contribution of the proposed computational framework is that it highlights the important role of surrogate-based models in parallel spatiotemporal simulations. The application of surrogate-based models helps us gain significant insight into the estimation of computing performance within parallel spatiotemporal simulations, which is the key computational challenge to efficient utilization of cyberinfrastructure. Moreover, in order to effectively manage, execute, and share complex spatiotemporal simulations, a scientific workflow middleware is designed to glue functionalities and computational infrastructure within the surrogate-based computational framework. Scientific workflow can wrap up all main components, resulting in the automation of heterogeneous spatial data processing and execution of multiple processes in spatiotemporal simulations. Therefore, it is much easier for researchers without computational background to apply the proposed computational framework in their own spatiotemporal simulations.

I use the spatiotemporal simulation of urban growth as our case study in Chapter 5. A parallel model calibration workflow is designed to tackle the most computationally intensive task in our simulation: model calibration. With support of the parallel approach of model calibration, we can greatly simplify the complex and computationally intensive calibration by assembling those tedious processes into workflow. The utility of our surrogate-based computational framework is examined in Chapter 6. We can observe that the spatial characteristics/content can significantly affect the computing intensity of spatiotemporal simulations. The experiment results indicate that surrogate-based approaches play a very important role in improving the computing performance of spatiotemporal simulations within cyberinfrastructure environments.

In sum, surrogate-based models can appropriate predict the computing performance by means of capturing the relationship between spatial characteristics/content and computing intensity. Regarding this, the surrogate-based computational framework can efficiently leverage the high performance computing power of cyberinfrastructure in spatiotemporal simulations. This framework substantially facilitates the solution of complex geographic problems at large scale.

Since very few existing work are related to the synthesis of surrogate-based approach and parallel spatiotemporal simulations, there are many aspects along this direction open for further study. Based on this work, we can concentrate further study on the following aspects: 1) development of various sampling strategies according to different characteristics of big spatial data (e.g., size and heterogeneity) to construct surrogate-based models ; 2) comparison of alternative types of surrogate-based model in terms of their prediction ability and computational effectiveness; 3) investigation of heuristic approaches for the establishment of surrogate-based models; 4) improvement of the model evaluation approach of surrogate-based models; 5) exploration of the visualization of response surface of surrogate-based model, for example the utilization of self-organizing map; 6) integration of space-time GIS and urban growth study .

REFERENCES

Abbott, C. A., Berry, M. W., Comiskey, E. J., Gross, L. J., & Hank-Kwang, L. (1997). Parallel individual-based modeling of Everglades deer ecology. Computational Science & Engineering, IEEE, 4(4), 60-78. doi: 10.1109/99.641610

Abbott, M. D., & Stanley, R. S. (1999). Modeling groundwater recharge and flow in an upland fractured bedrock aquifer. System Dynamics Review, 15(2), 163-184.

Acosta-Michlik, L., & Espaldon, V. (2008). Assessing vulnerability of selected farming communities in the Philippines based on a behavioural model of agent's adaptation to global environmental change. Global Environmental Change, 18(4), 554-563.

An, L. (2012). Modeling human decisions in coupled human and natural systems: review of agent-based models. Ecological modelling, 229, 25-36.

An, L., Linderman, M., Qi, J., Shortridge, A., & Liu, J. (2005). Exploring Complexity in a Human–Environment System: An Agent-Based Spatial Model for Multidisciplinary and Multiscale Integration. Annals of the association of American Geographers, 95(1), 54-79. doi: 10.1111/j.1467-8306.2005.00450.x

An, L., & Liu, J. (2010). Long-term effects of family planning and other determinants of fertility on population and environment: agent-based modeling evidence from Wolong Nature Reserve, China. Population and Environment, 31(6), 427-459.

Armbrust, M., Fox, A., Griffith, R., Joseph, A. D., Katz, R., Konwinski, A., . . . Stoica, I. (2010). A view of cloud computing. Communications of the ACM, 53(4), 50-58.

Armstrong, M. P. (2000). Geography and computational science. Annals of the association of American Geographers, 90(1), 146-156.

Armstrong, M. P., & Densham, P. J. (1992). Domain decomposition for parallel processing of spatial problems. Computers, Environment and Urban Systems, 16(6), 497-513.

Axelrod, R. M., Axelrod, R., & Cohen, M. D. (2000). Harnessing complexity: Organizational implications of a scientific frontier: Basic Books.

Baker, W. (1989). A review of models of landscape change. Landscape Ecology, 2(2), 111-133. doi: 10.1007/BF00137155

Batty, M. (1995). New ways of looking at cities. Nature, 377, 574.

Batty, M. (2007). Cities and complexity: understanding cities with cellular automata, agent-based models, and fractals: The MIT press.

Batty, M., Crooks, A. T., See, L. M., & Heppenstall, A. J. (2012). Perspectives on agent-based models and geographical systems Agent-Based Models of Geographical Systems (pp. 1-15): Springer.

Batty, M., & Longley, P. A. (1994). Fractal cities: a geometry of form and function: Academic Press.

Batty, M., & Xie, Y. (1994). From cells to cities. Environment and planning B, 21, s31-s31.

Batty, M., & Xie, Y. (1994). Urban analysis in a GIS environment: population density modelling using ARC/INFO. Spatial analysis and GIS, 189-220.

Batty, M., Xie, Y., & Sun, Z. (1999). Modeling urban dynamics through GIS-based cellular automata. Computers, Environment and Urban Systems, 23, 205-233. doi: 10.1016/S0198-9715(99)00015-0

Batty, M. (1991). Generating urban forms from diffusive growth. Environment and planning A, 23, 511-544.

Bell, E., & Hinojosa, R. (1977). Markov analysis of land use change: continuous time and stationary processes. Socio-Economic Planning Sciences, 11(1), 13-17.

Bell, E. J. (1974). Markov analysis of land use change—an application of stochastic processes to remotely sensed data. Socio-Economic Planning Sciences, 8(6), 311-316.

Bennett, D. A., & Tang, W. (2006). Modelling adaptive, spatially aware, and mobile agents: Elk migration in Yellowstone. International journal of geographical information science, 20(9), 1039-1066.

Berry, M., & Minser, K. (1997). Distributed Land-Cover Change Simulation Using PVM and MPI. Paper presented at the Land use modeling workshop.

Bertalanffy, L. v. (1968). General system theory: Foundations, development, applications: Braziller. New York.

Bestul, T. (1989). A general technique for creating SIMD algorithms on parallel pointer-based quadtrees: University of Maryland USA.

Bockstael, N. E., & Irwin, E. G. (2000). Economics and the land use-environment link. The international yearbook of environmental and resource economics, 2001, 1-54.

Bourne, L. S. (1971). Physical adjustment processes and land use succession: a conceptual review and central city example. Economic Geography, 1-15.

Bousquet, F., & Le Page, C. (2004). Multi-agent simulations and ecosystem management: a review. Ecological modelling, 176(3), 313-332.

Brown, D. G., Page, S., Riolo, R., Zellner, M., & Rand, W. (2005). Path dependence and the validation of agent - based spatial models of land use. International journal of geographical information science, 19(2), 153-174.

Brown, D. G., Page, S. E., Riolo, R., & Rand, W. (2004). Agent-based and analytical modeling to evaluate the effectiveness of greenbelts. Environmental Modelling & Software, 19(12), 1097-1109.

Brown, D. G., Pijanowski, B. C., & Duh, J. (2000). Modeling the relationships between land use and land cover on private lands in the Upper Midwest, USA. Journal of Environmental Management, 59(4), 247-263.

Brown, D. G., Pijanowski, B. C., & Duh, J. D. (2000). Modeling the relationships between land use and land cover on private lands in the Upper Midwest, USA. Journal of Environmental Management, 59(4), 247-263. doi: http://dx.doi.org/10.1006/jema.2000.0369

Brown, D. G., & Robinson, D. T. (2006). Effects of heterogeneity in residential preferences on an agent-based model of urban sprawl. Ecology and society, 11(1), 46.

Bolstad, P. (2005). GIS Fundamentals: A First Text on Geographic Information Systems. Eider Press.

Burnham, B. O. (1973). Markov intertemporal land use simulation model. Southern Journal of Agricultural Economics, 5(1), 253-258.

Cecchini, A. (1996). Urban modelling by means of cellular automata: generalised urban automata with the help on-line (AUGH) model. Environment and planning B, 23, 721-732.

Chang, Y., Hong, F., & Lee, M. (2008). A system dynamic based DSS for sustainable coral reef management in Kenting coastal zone, Taiwan. Ecological modelling, 211(1), 153-168.

Clarke, K., Hoppen, S., & Gaydos, L. (1997). A self-modifying cellular automaton model of historical. Environ Plan B, 24, 247-261.

Clarke, K. C. (2008). A decade of cellular urban modeling with SLEUTH: Unresolved issues and problems. Ch, 3, 47-60.

Clarke, K. C., & Gaydos, L. J. (1998). Loose-coupling a cellular automaton model and GIS: long-term urban growth prediction for San Francisco and Washington/Baltimore. International journal of geographical information science, 12(7), 699-714.

Clarke, K. C., Gazulis, N., Dietzel, C., & Goldstein, N. C. (2007). A decade of SLEUTHing: Lessons learned from applications of a cellular automaton land use change model. Classics in IJGIS: twenty years of the international journal of geographical information science and systems, 413-427.

Clarke, K. C., Hoppen, S., & Gaydos, L. (1996). Methods and techniques for rigorous calibration of a cellular automaton model of urban growth. Paper presented at the Third International Conference/Workshop on Integrating GIS and Environmental Modeling, Santa Fe, New Mexico.

Costanza, R., & Maxwell, T. (1991). Spatial ecosystem modelling using parallel processors. Ecological modelling, 58(1–4), 159-183. doi: http://dx.doi.org/10.1016/0304-3800(91)90034-X

Couclelis, H. (1985). Cellular worlds: a framework for modeling micro-macro dynamics. Environment and planning A, 17(5), 585-596.

Couclelis, H. (1989). Macrostructure and microbehavior in a metropolitan area. Environment and planning B, 16(2), 141-154.

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

Crawford, T. W., Messina, J. P., Manson, S. M., & O'Sullivan, D. (2005). Complexity science, complex systems, and land-use research. Environment and Planning B: Planning and Design, 32(6), 792-798.

Cressie, N. (2015). Statistics for spatial data. John Wiley & Sons.

Deadman, P., & Gimblett, R. H. (1994). A role for goal-oriented autonomous agents in modeling people-environment interactions in forest recreation. Mathematical and Computer Modelling, 20(8), 121-133.

Deadman, P., Robinson, D., Moran, E., & Brondizio, E. (2004). Colonist household decisionmaking and land-use change in the Amazon Rainforest: an agent-based simulation. Environment and planning B, 31, 693-710.

Deissenberg, C., van der Hoog, S., & Dawid, H. (2008). EURACE: A massively parallel agent-based model of the European economy. Applied Mathematics and Computation, 204(2), 541-552. doi: http://dx.doi.org/10.1016/j.amc.2008.05.116

Dietzel, C., & Clarke, K. C. (2007). Toward optimal calibration of the SLEUTH land use change model. Transactions in GIS, 11(1), 29-45.

Ding, Y., & Densham, P. J. (1996). Spatial strategies for parallel spatial modelling. International Journal of Geographical Information Systems, 10(6), 669-698. doi: 10.1080/02693799608902104

Draper, N. R., & Smith, H. (2014). Applied regression analysis. John Wiley & Sons.

Drewett, J. (1969). A stochastic model of the land conversion process: An interim report. Regional Studies, 3(3), 269-280.

Dyson, B., & Chang, N.-B. (2005). Forecasting municipal solid waste generation in a fast-growing urban region with system dynamics modeling. Waste management, 25(7), 669-679.

Efron, B. (1983). Estimating the error rate of a prediction rule: improvement on cross-validation. Journal of the American Statistical Association, 78(382), 316-331.

Efron, B., & Tibshirani, R. J. (1994). An introduction to the bootstrap. CRC press.

Epstein, J. M., & Axtell, R. (1996). Growing artificial societies: social science from the bottom up, Brookings Institution Press.

Epstein, J. M. (1999). Agent‐based computational models and generative social science. Complexity, 4(5), 41-60.

Erra, U., Frola, B., Scarano, V., & Couzin, I. (2009). An efficient GPU implementation for large scale individual-based simulation of collective behavior. Paper presented at the High Performance Computational Systems Biology, 2009. HIBI'09. International Workshop on.

Evans, T. P., & Kelley, H. (2004). Multi-scale analysis of a household level agent-based model of landcover change. Journal of Environmental Management, 72(1), 57-72.

Evans, T. P., Sun, W., & Kelley, H. (2006). Spatially explicit experiments for the exploration of land - use decision - making dynamics. International journal of geographical information science, 20(9), 1013-1037.

Farmer, J. D., & Foley, D. (2009). The economy needs agent-based modelling. Nature, 460(7256), 685-686.

Foley, J. A., DeFries, R., Asner, G. P., Barford, C., Bonan, G., Carpenter, S. R., . . . Gibbs, H. K. (2005). Global consequences of land use. science, 309(5734), 570-574.

Ford, A. (1996). Testing the snake river explorer. System Dynamics Review, 12(4), 305-329.

Ford, A. (2000). Modeling the environment: an introduction to system dynamics modeling of environmental systems. International Journal of Sustainability in Higher Education, 1(1).

Forrester, J. W. (1970). Urban dynamics. IMR; Industrial Management Review (pre-1986), 11(3), 67.

Forrester, A., Sobester, A., & Keane, A. (2008). Engineering design via surrogate modelling: a practical guide. John Wiley & Sons.

Franklin, W. R., Narayanaswami, C., Kankanhalli, M., Sun, D., Zhou, M.-C., & Wu, P. Y. (1989). Uniform grids: A technique for intersection detection on serial and parallel machines. Paper presented at the Proceedings of Auto Carto 9: Ninth International Symposium on Computer-Assisted Cartography.

Gahegan, M. (2003). Is inductive machine learning just another wild goose (or might it lay the golden egg)? International journal of geographical information science, 17(1), 69-92.

Gano, S. E., Kim, H., & Brown, D. E. (2006, September). Comparison of three surrogate modeling techniques: Datascape, kriging, and second order regression. In Proceedings of the 11th AIAA/ISSMO Multidisciplinary Analysis and Optimization Conference, AIAA-2006-7048, Portsmouth, Virginia.

Gimblett, H. R. (2002). Integrating geographic information systems and agent-based modeling techniques for simulating social and ecological processes (Vol. 168): Oxford University Press New York.

Gilbert, N., & Troitzsch, K. (2005). Simulation for the social scientist. McGraw-Hill Education (UK).

Goel, T., Vaidyanathan, R., Haftka, R. T., Shyy, W., Queipo, N. V., & Tucker, K. (2007). Response surface approximation of Pareto optimal front in multi-objective optimization. *Computer methods in applied mechanics and engineering*, *196*(4), 879-893.

Golub, G. H., Heath, M., & Wahba, G. (1979). Generalized cross-validation as a method for choosing a good ridge parameter. Technometrics, 21(2), 215-223.

Golledge, R., & Stimson, R. (1997). Spatial behaviour. Guilford, London.

Gong, Z., Tang, W., Bennett, D. A., & Thill, J.-C. (2013). Parallel agent-based simulation of individual-level spatial interactions within a multicore computing environment. International journal of geographical information science, 27(6), 1152-1170.

Gong, Z., Tang, W., & Thill, J.-C. (2012). Parallelization of ensemble neural networks for spatial land-use modeling. Paper presented at the Proceedings of the 5th International Workshop on Location-Based Social Networks, Redondo Beach, California.

Goodchild, M. F. (1992). Geographical information science. International Journal of Geographical Information Systems, 6(1), 31-45. doi: 10.1080/02693799208901893

Goodchild, M. F. (2001, September). A geographer looks at spatial information theory. In International Conference on Spatial Information Theory (pp. 1-13). Springer Berlin Heidelberg.

Goodchild, M. F. (2003). Geographic information science and systems for environmental management. Annual Review of Environment and Resources, 28(1), 493-519.

Goodchild, M. F. (2004). GIScience, geography, form, and process. Annals of the association of American Geographers, 94(4), 709-714.

Grant, W. E., Pedersen, E. K., & Mar ń, S. L. (1997). Ecology and natural resource management: systems analysis and simulation: John Wiley & Sons.

Greenberg, H. J. (1973). The generalized penalty-function/surrogate model. Operations Research, 21(1), 162-178.

Grimm, V. (1999). Ten years of individual-based modelling in ecology: what have we learned and what could we learn in the future? Ecological modelling, 115(2), 129-148.

Guan, Q., & Clarke, K. C. (2010). A general-purpose parallel raster processing programming library test application using a geographic cellular automata model. International journal of geographical information science, 24(5), 695-722. doi: 10.1080/13658810902984228

Guan, Q., Shi, X., Huang, M., & Lai, C. (2015). A hybrid parallel cellular automata model for urban growth simulation over GPU/CPU heterogeneous architectures. International journal of geographical information science, 1-21. doi: 10.1080/13658816.2015.1039538

Guo, H., Liu, L., Huang, G., Fuller, G., Zou, R., & Yin, Y. (2001). A system dynamics approach for regional environmental planning and management: a study for the Lake Erhai Basin. Journal of Environmental Management, 61(1), 93-111.

Han, J., Hayashi, Y., Cao, X., & Imura, H. (2009). Application of an integrated system dynamics and cellular automata model for urban growth assessment: A case study of Shanghai, China. Landscape and urban planning, 91(3), 133-141. doi: http://dx.doi.org/10.1016/j.landurbplan.2008.12.002

Hazen, B. C., & Berry, M. W. (1997). The simulation of land-cover change using a distributed computing environment. Simulation Practice and Theory, 5(6), 489-514.

Hall, P. (1986). On the bootstrap and confidence intervals. The Annals of Statistics, 1431-1452.

Herold, M., Goldstein, N. C., & Clarke, K. C. (2003). The spatiotemporal form of urban growth: measurement, analysis and modeling. Remote sensing of Environment, 86(3), 286-302.

Holland, J. H. (1992). Adaptation in natural and artificial systems: an introductory analysis with applications to biology, control, and artificial intelligence: MIT press.

Holland, J. H. (1995). Hidden order: How adaptation builds complexity: Basic Books.

Holland, J. H. (2000). Emergence: From chaos to order: Oxford University Press.

Holland, J. H. (2006). Studying complex adaptive systems. Journal of Systems Science and Complexity, 19(1), 1-8.

Hopkins, S., Healey, R., & Waugh, T. (1992). Algorithm scalability for line intersection detection in parallel polygon overlay. Bresnahan et al.[BCC92], 210-218.

Hu, Y., Cai, X., & DuPont, B. (2015). Design of a web-based application of the coupled multi-agent system model and environmental model for watershed management analysis using Hadoop. Environmental Modelling & Software, 70(0), 149-162. doi: http://dx.doi.org/10.1016/j.envsoft.2015.04.011

Itami, R. M. (1988). Cellular worlds, models for dynamic conceptions of landscape. Landscape Architecture, 78(5), 52-57.

Itami, R. M. (1994). Simulating spatial dynamics: cellular automata theory. Landscape and urban planning, 30, 27-47. doi: 10.1016/0169-2046(94)90065-5

Jahan, S. (1986). The determination of stability and similarity of Markovian land use change processes: a theoretical and empirical analysis. Socio-Economic Planning Sciences, 20(4), 243-251.

Johnson, M. E., Moore, L. M., & Ylvisaker, D. (1990). Minimax and maximin distance designs. Journal of statistical planning and inference, 26(2), 131-148.

Jones, D. R., Schonlau, M., & Welch, W. J. (1998). Efficient global optimization of expensive black-box functions. Journal of Global optimization, 13(4), 455-492.

Jenerette, G. D., & Wu, J. (2001). Analysis and simulation of land-use change in the central Arizona – Phoenix region, USA. Landscape Ecology, 16(7), 611-626. doi: 10.1023/A:1013170528551

Kleijnen, J. P. (2009). Kriging metamodeling in simulation: A review. *European Journal of Operational Research*, *192*(3), 707-716.

Kleijnen, J. P., & Sargent, R. G. (2000). A methodology for fitting and validating metamodels in simulation. *European Journal of Operational Research*, *120*(1), 14-29.

Kim, I.-H., Tsou, M.-H., & Feng, C.-C. (2015). Design and implementation strategy of a parallel agent-based Schelling model. Computers, Environment and Urban Systems, 49, 30-41.

Krige, D. G. (1951) A statistical approach to some basic mine valuation problems on the Witwatersrand. Journal of the Chemical, Metal, and Mining Society of South Africa, 52, 119-139

Lambin, E. F. (1997). Modelling and monitoring land-cover change processes in tropical regions. Progress in physical geography, 21(3), 375-393.

Lambin, E. F., Rounsevell, M., & Geist, H. (2000). Are agricultural land-use models able to predict changes in land-use intensity? Agriculture, Ecosystems & Environment, 82(1), 321-331.

Lan, Z., Taylor, V. E., & Bryan, G. (2002). A novel dynamic load balancing scheme for parallel systems. Journal of Parallel and Distributed Computing, 62(12), 1763-1781. doi: http://dx.doi.org/10.1016/S0743-7315(02)00008-4

Lansing, J. S. (2003). Complex adaptive systems. Annual review of anthropology, 32(1), 183-204.

Li, B. (1992). Opportunities and challenges of parallel processing in spatial data analysis: initial experiments with data parallel map analysis. Paper presented at the GIS LIS-INTERNATIONAL CONFERENCE-.

Li, F., Liang, J., Clarke, K., Li, M., Liu, Y., & Huang, Q. (2015). Urban land growth in eastern China: a general analytical framework based on the role of urban micro-agents' adaptive behavior. Regional Environmental Change, 15(4), 695-707.

Li, X., & Liu, X. (2008). Embedding sustainable development strategies in agent‐based models for use as a planning tool. International journal of geographical information science, 22(1), 21-45.

Li, X., & Yeh, A. G.-O. (2000). Modelling sustainable urban development by the integration of constrained cellular automata and GIS. International journal of geographical information science, 14, 131-152. doi: 10.1080/136588100240886

Li, X., & Yeh, A. G.-O. (2002). Urban simulation using principal components analysis and cellular automata for land-use planning. Photogrammetric engineering and remote sensing, 68(4), 341-352.

Li, X., Zhang, X., Yeh, A., & Liu, X. (2010). Parallel cellular automata for large-scale urban simulation using load-balancing techniques. International journal of geographical information science, 24(6), 803-820. doi: 10.1080/13658810903107464

Liu, J., Dietz, T., Carpenter, S. R., Alberti, M., Folke, C., Moran, E., . . . Lubchenco, J. (2007). Complexity of coupled human and natural systems. science, 317(5844), 1513-1516.

Liu, X., Ou, J., Li, X., & Ai, B. (2013). Combining system dynamics and hybrid particle swarm optimization for land use allocation. Ecological modelling, 257(0), 11-24. doi: http://dx.doi.org/10.1016/j.ecolmodel.2013.02.027

Liu, Y., Lv, X., Qin, X., Guo, H., Yu, Y., Wang, J., & Mao, G. (2007). An integrated GIS-based analysis system for land-use management of lake areas in urban fringe. Landscape and urban planning, 82(4), 233-246.

Loibl, W., & Toetzer, T. (2003). Modeling growth and densification processes in suburban regions—simulation of landscape transition with spatial agents. Environmental Modelling & Software, 18(6), 553-563.

Lubowski, R. N., Plantinga, A. J., & Stavins, R. N. (2008). What drives land-use change in the United States? A national analysis of landowner decisions. Land Economics, 84(4), 529-550.

Ludäscher, B., Altintas, I., Berkley, C., Higgins, D., Jaeger, E., Jones, M. B., . . . Zhao, Y. (2006). Scientific workflow management and the Kepler system. Concurrency and Computation: Practice and Experience, 18(10), 1039-1065.

Mahiny, A., & Gholamalifard, M. (2007). Dynamic spatial modeling of urban growth through cellular automata in a GIS environment.

Manson, S. (2006). Land use in the southern Yucatán peninsular region of Mexico: Scenarios of population and institutional change. Computers, Environment and Urban Systems, 30(3), 230-253.

Manson, S. M. (2001). Simplifying complexity: a review of complexity theory. Geoforum, 32(3), 405-414.

Manson, S. M. (2005). Agent-based modeling and genetic programming for modeling land change in the Southern Yucatán Peninsular Region of Mexico. Agriculture, Ecosystems & Environment, 111(1–4), 47-62. doi: http://dx.doi.org/10.1016/j.agee.2005.04.024

Manson, S. M. (2006). Bounded rationality in agent‐based models: experiments with evolutionary programs. International journal of geographical information science, 20(9), 991-1012.

Manson, S. M., & Evans, T. (2007). Agent-based modeling of deforestation in southern Yucatán, Mexico, and reforestation in the Midwest United States. Proceedings of the National Academy of Sciences, 104(52), 20678-20683.

Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C., & Byers, A. H. (2011). Big data: The next frontier for innovation, competition, and productivity.

Marurngsith, W. (2014). Computing Platforms for Large-Scale Multi-Agent Simulations: The Niche for Heterogeneous Systems Intelligent Data Engineering and Automated Learning–IDEAL 2014 (pp. 424-432): Springer.

Mashayekhi, A. N. (1990). Rangelands destruction under population growth: the case of Iran. System Dynamics Review, 6(2), 167-193.

Martin, J. D., & Simpson, T. W. (2005). Use of kriging models to approximate deterministic computer models. *AIAA journal*, *43*(4), 853-863.

Matthews, R. B., Gilbert, N. G., Roach, A., Polhill, J. G., & Gotts, N. M. (2007). Agent-based land-use models: a review of applications. Landscape Ecology, 22(10), 1447-1459.

McKay, M. D., Beckman, R. J., & Conover, W. J. (2000). A comparison of three methods for selecting values of input variables in the analysis of output from a computer code. Technometrics, 42(1), 55-61.

Medeiros, C. B., Perez-Alcazar, J., Digiampietri, L., Pastorello Jr, G. Z., Santanche, A., Torres, R., . . . Bacarin, E. (2005). WOODSS and the Web: annotating and reusing scientific workflows. ACM SIGMOD Record, 34(3), 18-23.

Meentemeyer, R. K., Tang, W., Dorning, M. A., Vogler, J. B., Cunniffe, N. J., & Shoemaker, D. A. (2013). FUTURES: multilevel simulations of emerging urban–rural landscape structure using a stochastic patch-growing algorithm. Annals of the association of American Geographers, 103(4), 785-807.

Millington, J., Romero-Calcerrada, R., Wainwright, J., & Perry, G. (2008). An agent-based model of Mediterranean agricultural land-use/cover change for examining wildfire risk. Journal of Artificial Societies and Social Simulation, 11(4), 4.

Mohapatra, P. K., Mandal, P., & Bora, M. C. (1994). Introduction to system dynamics modeling: University of Nevada Press.

Morris, M. D., & Mitchell, T. J. (1995). Exploratory designs for computational experiments. Journal of statistical planning and inference, 43(3), 381-402.

Muller, M., & Middleton, J. (1994). A Markov model of land-use change dynamics in the Niagara Region, Ontario, Canada. Landscape Ecology, 9(2), 151-157. doi: 10.1007/BF00124382

Myers, R. H., Montgomery, D. C., & Anderson-Cook, C. M. (2016). Response surface methodology: process and product optimization using designed experiments. John Wiley & Sons.

Nagel, K., & Rickert, M. (2001). Parallel implementation of the TRANSIMS micro-simulation. Parallel Computing, 27(12), 1611-1639. doi: http://dx.doi.org/10.1016/S0167-8191(01)00106-5

Naill, R. F., Belanger, S., Klinger, A., & Petersen, E. (1992). An analysis of the cost effectiveness of US energy policies to mitigate global warming. System Dynamics Review, 8(2), 111-128.

Nichols, J. A., Hallam, T. G., & Dimitrov, D. T. (2008). Parallel simulation of ecological structured communities: Computational needs, hardware capabilities, and nonlinear applications. Nonlinear Analysis: Theory, Methods & Applications, 69(3), 832-842.

NSF. (2007). Cyberinfrastructure Vision for 21st Century Discovery.

Nugala, V., Allan, S. J., & Haefner, J. W. (1998). Parallel implementations of individual-based models in biology: bulletin- and non-bulletin-board approaches. Biosystems, 45(2), 87-97. doi: http://dx.doi.org/10.1016/S0303-2647(97)00069-5

Openshaw, S. (2014). GeoComputation. Paper presented at the Geocomputation.

Openshaw, S., & Turton, I. (2005). High Performance Computing and the Art of Parallel Programming: An Introduction for Geographers, Social Scientists and Engineers: Routledge.

O'Sullivan, D. (2008). Geographical information science: agent-based models. Progress in Human Geography.

O'Sullivan, D., & Perry, G. L. (2013). Spatial simulation: exploring pattern and process: John Wiley & Sons.

Overmars, K. P., & Verburg, P. H. (2007). Comparison of a deductive and an inductive approach to specify land suitability in a spatially explicit land use model. Land use policy, 24(3), 584-599.

Owczarz, W., & Zlatev, Z. (2002). Parallel matrix computations in air pollution modelling. Parallel Computing, 28(2), 355-368. doi: http://dx.doi.org/10.1016/S0167-8191(01)00144-2

Queipo, N. V., Haftka, R. T., Shyy, W., Goel, T., Vaidyanathan, R., & Tucker, P. K. (2005). Surrogate-based analysis and optimization. Progress in aerospace sciences, 41(1), 1-28.

Palmer, K., & Tsui, K. L. (2001). A minimum bias Latin hypercube design. IIE Transactions, 33(9), 793-808.

Parker, D. C., Manson, S. M., Janssen, M. A., Hoffmann, M. J., & Deadman, P. (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.

Parker, D. C., & Meretsky, V. (2004). Measuring pattern outcomes in an agent-based model of edge-effect externalities using spatial metrics. Agriculture, Ecosystems & Environment, 101(2–3), 233-250. doi: http://dx.doi.org/10.1016/j.agee.2003.09.007

Parry, H. R., & Bithell, M. (2012). Large scale agent-based modelling: A review and guidelines for model scaling Agent-based models of geographical systems (pp. 271-308): Springer.

Parry, H. R., & Evans, A. J. (2008). A comparative analysis of parallel processing and super-individual methods for improving the computational performance of a large individual-based model. Ecological modelling, 214(2–4), 141-152. doi: http://dx.doi.org/10.1016/j.ecolmodel.2008.02.002

Phipps, M. J. (1992). From local to global: the lesson of cellular automata Individual-based models and approaches in ecology (pp. 165-187): Springer.

Pijanowski, B., Alexandridis, K., & Mueller, D. (2006). Modelling urbanization patterns in two diverse regions of the world. Journal of Land Use Science, 1(2-4), 83-108.

Pijanowski, B. C., Tayyebi, A., Doucette, J., Pekin, B. K., Braun, D., & Plourde, J. (2014). A big data urban growth simulation at a national scale: Configuring the GIS and neural network based Land Transformation Model to run in a High Performance Computing (HPC) environment. Environmental Modelling & Software, 51, 250-268.

Portugali, J., & Benenson, I. (1995). Artificial planning experience by means of a heuristic cell-space model: simulating international migration in the urban process. Environment and planning A, 27(10), 1647-1665.

Qu, W., & Barney, G. O. (1999). Projecting China's grain supply and demand using a new computer simulation model. China in the Global Economy Agriculture in

China and OECD Countries Past Policies and Future Challenges (OECD Proceedings): Past Policies and Future Challenges (OECD Proceedings), 227.

Qu, W., & Barney, G. O. (1999). Projecting China's grain supply and demand using a new computer simulation model. China in the Global Economy Agriculture in China and OECD Countries Past Policies and Future Challenges (OECD Proceedings): Past Policies and Future Challenges (OECD Proceedings), 227.

Railsback, S. F. (2001). Concepts from complex adaptive systems as a framework for individual-based modelling. Ecological modelling, 139(1), 47-62.

Randers, J. (1980). Elements of the system dynamics method: MIT press Cambridge, MA.

Reeves, H. W., & Zellner, M. L. (2010). Linking MODFLOW with an Agent ‐ Based Land ‐ Use Model to Support Decision Making. Groundwater, 48(5), 649-660.

Richardson, G. P., & Pugh III, A. I. (1981). Introduction to system dynamics modeling with DYNAMO: Productivity Press Inc.

Richmond, P., Coakley, S., & Romano, D. M. (2009). A high performance agent based modelling framework on graphics card hardware with CUDA. Paper presented at the Proceedings of The 8th International Conference on Autonomous Agents and Multiagent Systems-Volume 2.

Richmond, P., Walker, D., Coakley, S., & Romano, D. (2010). High performance cellular level agent-based simulation with FLAME for the GPU. Briefings in bioinformatics, 11(3), 334-347.

Rindfuss, R. R., Walsh, S. J., Turner, B., Fox, J., & Mishra, V. (2004). Developing a science of land change: challenges and methodological issues. Proceedings of the National Academy of Sciences of the United States of America, 101(39), 13976-13981.

Robinson, D. T., Brown, D. G., Parker, D. C., Schreinemachers, P., Janssen, M. A., Huigen, M., . . . Barnaud, C. (2007). Comparison of empirical methods for building agent-based models in land use science. Journal of Land Use Science, 2(1), 31-55. doi: 10.1080/17474230701201349

Robinson, V. B. (1978). Information theory and sequences of land use: an application. The Professional Geographer, 30(2), 174-179.

Sacks, J., Welch, W. J., Mitchell, T. J., & Wynn, H. P. (1989). Design and analysis of computer experiments. Statistical science, 409-423.

Santner, T. J., Williams, B. J., & Notz, W. I. (2013). The design and analysis of computer experiments. Springer Science & Business Media.

Santé, I., García, A. M., 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(2), 108-122.

Saysel, A. K., Barlas, Y., & Yenigün, O. (2002). Environmental sustainability in an agricultural development project: a system dynamics approach. Journal of Environmental Management, 64(3), 247-260.

Scheutz, M., & Harris, J. (2012). An overview of the SimWorld agent-based grid experimentation system. Large-Scale Computing Techniques for Complex System Simulations, 80, 59.

Shen, Q., Chen, Q., Tang, B.-s., Yeung, S., Hu, Y., & Cheung, G. (2009). A system dynamics model for the sustainable land use planning and development. Habitat International, 33(1), 15-25.

Shen, T.-y., Wang, W.-d., Min, H., Guo, Z.-c., Ling, X., & Yang, K.-z. (2007). Study on spatio-temporal system dynamic models of urban growth. Systems Engineering-Theory & Practice, 27(1), 10-17.

Shook, E., Wang, S., & Tang, W. (2013). A communication-aware framework for parallel spatially explicit agent-based models. International journal of geographical information science, 27(11), 2160-2181.

Simpson, T. W., Mauery, T. M., Korte, J. J., & Mistree, F. (2001). Kriging models for global approximation in simulation-based multidisciplinary design optimization. AIAA journal, 39(12), 2233-2241.

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

Soares-Filho, B. S., Nepstad, D. C., Curran, L. M., Cerqueira, G. C., Garcia, R. A., Ramos, C. A., . . . Schlesinger, P. (2006). Modelling conservation in the Amazon basin. Nature, 440(7083), 520-523.

Sohl, T., & Sayler, K. (2008). Using the FORE-SCE model to project land-cover change in the southeastern United States. Ecological modelling, 219(1), 49-65.

Sohl, T. L., Sayler, K. L., Drummond, M. A., & Loveland, T. R. (2007). The FORE-SCE model: a practical approach for projecting land cover change using scenario-based modeling. Journal of Land Use Science, 2(2), 103-126.

Solé, R., & Goodwin, B. (2008). Signs Of Life How Complexity Pervades Biology: How Complexity Pervades Biology: Basic books.

Sterman, J. D. (2000). Business dynamics: systems thinking and modeling for a complex world (Vol. 19): Irwin/McGraw-Hill Boston.

Stockwell, D. R., & Peterson, A. T. (2002). Effects of sample size on accuracy of species distribution models. Ecological modelling, 148(1), 1-13

Takeyama, M., & Couclelis, H. (1997). Map dynamics: integrating cellular automata and GIS through Geo-Algebra. International journal of geographical information science, 11(1), 73-91.

Tang, B. (1993). Orthogonal array-based Latin hypercubes. Journal of the American statistical association, 88(424), 1392-1397.

Tang, W., & Bennett, D. A. (2009). Parallel agent-based modelling of land-use opinion dynamics using Graphics Processing Units.

Tang, W., & Bennett, D. A. (2012). Reprint of: Parallel agent-based modeling of spatial opinion diffusion accelerated using graphics processing units. Ecological modelling, 229, 108-118. doi: 10.1016/j.ecolmodel.2012.02.003

Tang, W., Bennett, D. A., & Wang, S. (2011a). A parallel agent-based model of land use opinions. Journal of Land Use Science, 6(2-3), 121-135. doi: 10.1080/1747423x.2011.558597

Tang, W., & Jia, M. (2014). Global sensitivity analysis of a large agent-based model of spatial opinion exchange: A heterogeneous multi-GPU acceleration approach. Annals of the association of American Geographers, 104(3), 485-509.

Tang, W., & Wang, S. (2009). HPABM: A Hierarchical Parallel Simulation Framework for Spatially-explicit Agent-based Models. Transactions in GIS, 13(3), 315-333. doi: 10.1111/j.1467-9671.2009.01161.x

Tang, W., Wang, S., Bennett, D. A., & Liu, Y. (2011b). Agent-based modeling within a cyberinfrastructure environment: a service-oriented computing approach. International journal of geographical information science, 25(9), 1323-1346. doi: 10.1080/13658816.2011.585342

Taylor, I. J., Deelman, E., Gannon, D. B., & Shields, M. (2014). Workflows for e-Science: scientific workflows for grids: Springer Publishing Company, Incorporated.

Timm, I. J., & Pawlaszczyk, D. (2005). Large scale multiagent simulation on the grid. Paper presented at the Cluster Computing and the Grid, 2005. CCGrid 2005. IEEE International Symposium on.

Tobler, W. R. (1970). A computer movie simulating urban growth in the Detroit region. Economic geography, 46, 234-240.

Tobler, W. (1979). Cellular geography Philosophy in geography (pp. 379-386): Springer.

Toffoli, T., & Margolus, N. (1987). Cellular automata machines: a new environment for modeling: MIT press.

Turner, B., Lambin, E. F., & Reenberg, A. (2008). Land change science special feature: The emergence of land change science for global environmental change and sustainability (vol 104, pg 20666, 2007). Proceedings of the National Academy of Sciences of the United States of America, 105(7), 2751-2751.

Turner, M. (1987). Spatial simulation of landscape changes in Georgia: A comparison of 3 transition models. Landscape Ecology, 1(1), 29-36. doi: 10.1007/BF02275263

Turton, I., & Openshaw, S. (1998). High-performance computing and geography: developments, issues, and case studies. Environment and planning A, 30(10), 1839-1856.

Uziel, E., & Berry, M. W. (1995). Parallel Models of Animal Migration in Northern Yellowstone National Park. International Journal of High Performance Computing Applications, 9(4), 237-255. doi: 10.1177/109434209500900401

Valbuena, D., Bregt, A. K., McAlpine, C., Verburg, P. H., & Seabrook, L. (2010). An agent-based approach to explore the effect of voluntary mechanisms on land use change: A case in rural Queensland, Australia. Journal of Environmental Management, 91(12), 2615-2625.

Valbuena, D., Verburg, P., Bregt, A., & Ligtenberg, A. (2010). An agent-based approach to model land-use change at a regional scale. Landscape Ecology, 25(2), 185-199. doi: 10.1007/s10980-009-9380-6

Verburg, P. H., & Overmars, K. P. (2009). Combining top-down and bottom-up dynamics in land use modeling: exploring the future of abandoned farmlands in Europe with the Dyna-CLUE model. Landscape Ecology, 24(9), 1167-1181.

Verburg, P. H., & Overmars, K. P. (2009). Combining top-down and bottom-up dynamics in land use modeling: exploring the future of abandoned farmlands in Europe with the Dyna-CLUE model. Landscape Ecology, 24(9), 1167-1181.

Verburg, P. H., Schot, P. P., Dijst, M. J., & Veldkamp, A. (2004). Land use change modelling: current practice and research priorities. GeoJournal, 61(4), 309-324.

Verburg, P. H., Van De Steeg, J., Veldkamp, A., & Willemen, L. (2009). From land cover change to land function dynamics: a major challenge to improve land characterization. Journal of Environmental Management, 90(3), 1327-1335.

Vezjak, M., Savsek, T., & Stuhler, E. (1998). System dynamics of euthrophication processes in lakes. European Journal of operational research, 109(2), 442-451.

Vizayakumar, K., & Mohapatra, P. (1992). Modelling and simulation of environmental impacts of a coalfield: System dynamics approach. Journal of Environmental Systems, 22(1), 59-79.

Vizayakumar, K., & Mohapatra, P. K. (1992). Environmental impact analysis of a coalfield. Journal of Environmental Management, 34(2), 79-103.

Vlachos, D., Georgiadis, P., & Iakovou, E. (2007). A system dynamics model for dynamic capacity planning of remanufacturing in closed-loop supply chains. Computers & Operations Research, 34(2), 367-394.

Vokorokos, L. (2005, 13-16 April 2005). Parallel computer system utilization in geographic information systems. Paper presented at the Computational Cybernetics, 2005. ICCC 2005. IEEE 3rd International Conference on.

Wang, G. G., & Shan, S. (2007). Review of metamodeling techniques in support of engineering design optimization. *Journal of Mechanical design*, *129*(4), 370-380.

Waddell, P. (2002). UrbanSim: Modeling urban development for land use, transportation, and environmental planning. Journal of the American Planning Association, 68(3), 297-314.

Wang, D., Berry, M. W., Carr, E. A., & Gross, L. J. (2006). A Parallel Fish Landscape Model for Ecosystem Modeling. SIMULATION, 82(7), 451-465. doi: 10.1177/0037549706068826

Wang, S. (2010). A CyberGIS Framework for the Synthesis of Cyberinfrastructure, GIS, and Spatial Analysis. Annals of the association of American Geographers, 100(3), 535-557. doi: 10.1080/00045601003791243

Wang, S. (2010). CyberGIS: Empowering the synthesis of computational and spatial thinking. Paper presented at the National Science Foundation TeraGrid Workshop on Cyber-GIS, Washington, DC.

Wang, S., & Armstrong, M. P. (2003). A quadtree approach to domain decomposition for spatial interpolation in Grid computing environments. Parallel Computing, 29(10), 1481-1504. doi: http://dx.doi.org/10.1016/j.parco.2003.04.003

Wang, S., & Armstrong, M. P. (2009). A theoretical approach to the use of cyberinfrastructure in geographical analysis. International journal of geographical information science, 23(2), 169-193. doi: 10.1080/13658810801918509

Warren, K., Franklin, C., & Streeter, C. L. (1998). New directions in systems theory: Chaos and complexity. Social Work, 43(4), 357-372.

Weber, C., & Puissant, A. (2003). Urbanization pressure and modeling of urban growth: example of the Tunis Metropolitan Area. Remote sensing of Environment, 86(3), 341-352.

Weng, Q. (2002). Land use change analysis in the Zhujiang Delta of China using satellite remote sensing, GIS and stochastic modelling. Journal of Environmental Management, 64(3), 273-284.

White, R., & Engelen, G. (1993). Cellular automata and fractal urban form: a cellular modelling approach to the evolution of urban land-use patterns. Environment and planning A, 25(8), 1175-1199.

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

Wilkinson, B., & Allen, C. M. (2004). Parallel Programming: Techniques and Applications Using Networked Workstations and Parallel Computers (Second Edition). New Jersey: Peason Prentice Hall.

Wolfram, S. (1984). Cellular automata as models of complexity. Nature, 311(5985), 419-424.

Wolstenholme, E. F. (1983). Modelling national development programmes--An exercise in system description and qualitative analysis using system dynamics. Journal of the Operational Research Society, 1133-1148.

Wood, T. S., & Shelley, M. L. (1999). A dynamic model of bioavailability of metals in constructed wetland sediments. Ecological Engineering, 12(3), 231-252.

Wu, F. (1998). SimLand: a prototype to simulate land conversion through the integrated GIS and CA with AHP-derived transition rules. International journal of geographical information science, 12(1), 63-82.

Wu, F. (2002). Calibration of stochastic cellular automata: the application to rural-urban land conversions. International journal of geographical information science, 16(8), 795-818.

Wu, F., & Martin, D. (2002). Urban expansion simulation of Southeast England using population surface modelling and cellular automata. Environment and planning A, 34, 1855-1876. doi: 10.1068/a3520

Wu, F., & Webster, C. J. (1998). Simulation of land development through the integration of cellular automata and multicriteria evaluation. Environment and planning B, 25, 103-126.

Wu, J., Vankat, J. L., & Barlas, Y. (1993). Effects of patch connectivity and arrangement on animal metapopulation dynamics: a simulation study. Ecological modelling, 65(3), 221-254.

Wu, Y., Sklar, F. H., Gopu, K., & Rutchey, K. (1996). Fire simulations in the Everglades landscape using parallel programming. Ecological modelling, 93(1), 113-124.

Xiao, N., Bennett, D. A., & Armstrong, M. P. (2007). Interactive evolutionary approaches to multiobjective spatial decision making: A synthetic review. Computers, Environment and Urban Systems, 31(3), 232-252.

Yang, X., & Lo, C. (2003). Modelling urban growth and landscape changes in the Atlanta metropolitan area. International journal of geographical information science, 17(5), 463-488.

Yin, L., & Muller, B. (2007). Residential location and the biophysical environment: exurban development agents in a heterogeneous landscape. Environment and Planning B: Planning and Design, 34(2), 279.

Zikopoulos, P., & Eaton, C. (2011). Understanding big data: Analytics for enterprise class hadoop and streaming data: McGraw-Hill Osborne Media.