Spatial Analysis of Urbanization in the Salt Lake Valley: An Urban Ecosystem Perspective

John H. Lowry Jr.
Utah State University

Follow this and additional works at: https://digitalcommons.usu.edu/etd

Part of the Geography Commons, and the Urban Studies and Planning Commons

Recommended Citation
https://digitalcommons.usu.edu/etd/746

This Dissertation is brought to you for free and open access by the Graduate Studies at DigitalCommons@USU. It has been accepted for inclusion in All Graduate Theses and Dissertations by an authorized administrator of DigitalCommons@USU. For more information, please contact digitalcommons@usu.edu.
SPATIAL ANALYSIS OF URBANIZATION IN THE SALT LAKE VALLEY:
AN URBAN ECOSYSTEM PERSPECTIVE

by

John H. Lowry Jr.

A dissertation submitted in partial fulfillment
of the requirements for the degree
of
DOCTOR OF PHILOSOPHY
in
Human Dimensions of Ecosystem Science and Management

Approved:

R. Douglas Ramsey
Major Professor

Matthew Baker
Committee Member

Claudia Radel
Committee Member

Mevin Hooten
Committee Member

Richard Toth
Committee Member

Byron Burnham
Dean of Graduate Studies

UTAH STATE UNIVERSITY
Logan, Utah

2010
ABSTRACT

Spatial Analysis of Urbanization in the Salt Lake Valley: 
An Urban Ecosystem Perspective

By

John H. Lowry Jr., Doctor of Philosophy
Utah State University, 2010

Major Professor: R. Douglas Ramsey
Department: Environment and Society

Because urban areas comprise a variety of biotic (e.g. people, trees) and abiotic (e.g. streets, water) components that interact and are often interdependent upon one another, it is helpful to study urban areas as urban ecosystems.

Our goal in Chapter 2 is to measure and quantify the spatial and demographic structure of the urbanized portion of Salt Lake County, Utah. We use 18 metrics from four broad categories (density, centrality, accessibility, and neighborhood mix) to measure urban form for three age-based residential neighborhood types. Using analysis of variance (ANOVA) we test for differences in mean values for the 18 urban form metrics. We find measureable differences in the spatial and demographic characteristics of these neighborhoods, suggesting that the rate of urban sprawl in Salt Lake County has been holding steady, if not increasing, during the last 20 years.

Chapter 3 seeks to better understand how spatial heterogeneity in urban tree canopy is related to household characteristics, urban form, and the geophysical landscape
of residential neighborhoods. We consider neighborhood age a factor that moderates the relationship between these determinants of tree canopy, and the abundance of tree canopy observed. Using linear regression analysis with neighborhood age as interaction term, we assess the relationship between tree canopy and 15 determinants of tree canopy abundance at three neighborhood ages. We find that neighborhood age has a significant moderating effect on the relationship between several determinants of canopy cover and the abundance of canopy cover observed.

While the urban forest provides many benefits to human well-being, it also consumes considerable quantities of water. An important question in Chapter 4 is to determine whether a growing urban forest increases overall residential irrigation demand, decreases demand, or has no apparent effect. Using a water demand model borrowed from agronomy, we estimate irrigation water demand based on the area of three residential landscape types and climatic factors. We project future residential water demand by generating residential landscape scenarios based on predicted urban forest canopy growth. We find that urban forest growth has the effect of stabilizing or potentially decreasing overall residential irrigation water demand.
DEDICATION

John H. Lowry Sr.

December 28, 1934 – March 24, 1995
ACKNOWLEDGMENTS

No one accomplishes anything of any significance without the help of others. And so it is with this dissertation and my PhD at Utah State University. I would like to thank my committee for their patience and willingness to mentor me through this process. I thank Doug Ramsey for his encouragement and support. From the day I decided to pursue the PhD he has encouraged me and given me the latitude to work on it while working for him in the RS/GIS laboratory. I thank Mevin Hooten for teaching me spatial statistics, helping me with R-code, and helping me understand the utility of statistical analysis. I thank Matt Baker for encouraging me to think more deeply about the questions I ask, and for showing me how to write a better academic paper. I thank Claudia Radel for her helpful criticism of my writing, and for her guidance and advice on pursuing an academic career. I thank Dick Toth for sharing with me anecdotes and wisdom about life in general and also about bioregional planning.

As a nontraditional student working while studying I’ve been fortunate to learn from many teachers, mentors, and colleagues in my 11 years at USU. I thank Tom Edwards for being the first to encourage me to pursue the PhD. I thank Layne Coppock for introducing me to interdisciplinary scholarship and teaching me how to play squash—and for giving me encouragement in both. I thank Susan Durham for helping me—through numerous visits—to better understand regression analysis and for advice on the statistical analyses in Chapter 3. I thank Rogier Kjelgren and Joanna Endter-Wada for their interest in my work, and for their advice on ways to estimate water irrigation demand, which appears in Chapter 4. Others who have given me valuable
encouragement by simply showing interest in me personally and in my research include Mike Kuhns, Chris Luecke, and Mark Brunson among others.

Over the past 11 years I’ve had the privilege of working with well over 60 students and professional staff in the RS/GIS Lab. To simply say I have learned a lot from them would be an understatement. I’ve discovered that it’s impossible to be an expert in everything, and over the course of these years I’ve relied on the expertise of many individuals in the lab who have taught me about ecology, programming, field botany, GIS and remote sensing, and doing scholarly research. If it were not for their skills, knowledge, and dedication to their work, I would not have been able to pursue this PhD. I am indebted to, and thank, the many lab personnel—past and present—for their exemplary work at the RS/GIS Lab.

Finally, none of this would have been possible without the support and patience of my family, and in particular my wife, Brenda. Throughout she has encouraged me and supported me with patience as we plodded along waiting to turn the page to the next chapter in our lives. I guess the next chapter will be in Fiji. I thank her for her love, support, and patience. I also thank my sons Toby and Sam for being good kids, and hope that they will learn to experience the satisfaction that comes from working hard and dedicating themselves to worthy pursuits whether in academics or elsewhere.

John H. Lowry Jr.

22 July, 2010
## CONTENTS

<table>
<thead>
<tr>
<th>Chapter</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>INTRODUCTION</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>MEASURING URBAN SPRAWL: A SPATIAL AND DEMOGRAPHIC CHARACTERIZATION OF RESIDENTIAL NEIGHBORHOODS IN SALT LAKE COUNTY</td>
<td>5</td>
</tr>
<tr>
<td>3</td>
<td>DETERMINANTS OF URBAN TREE CANOPY IN RESIDENTIAL NEIGHBORHOODS: HOUSEHOLD CHARACTERISTICS, URBAN FORM, AND THE GEOPHYSICAL LANDSCAPE</td>
<td>37</td>
</tr>
<tr>
<td>4</td>
<td>PREDICTING URBAN FOREST GROWTH AND ITS IMPACT ON RESIDENTIAL LANDSCAPE WATER DEMAND IN A SEMIARID ENVIRONMENT</td>
<td>74</td>
</tr>
<tr>
<td>5</td>
<td>SUMMARY AND CONCLUSION</td>
<td>108</td>
</tr>
</tbody>
</table>

APPENDICES ........................................................................................................ 112

CURRICULUM VITA ................................................................................................. 152
**LIST OF TABLES**

<table>
<thead>
<tr>
<th>Table</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2-1. List of urban form metrics and data sources</td>
<td>31</td>
</tr>
<tr>
<td>2-2. Comparison of urban form metrics for neighborhood types</td>
<td>32</td>
</tr>
<tr>
<td>3-1. Determinants of urban tree canopy</td>
<td>68</td>
</tr>
<tr>
<td>3-2. Estimated simple slopes for 15 explanatory variables</td>
<td>69</td>
</tr>
<tr>
<td>3-3. AIC information-theoretic multi-model analysis of eight candidate models from three theories explaining tree canopy abundance in residential neighborhoods</td>
<td>70</td>
</tr>
<tr>
<td>3-4. Computed variable importance for 16 determinants of urban tree canopy in residential neighborhoods using the sum of AIC weights (w_i) from analysis of (2^{16} - 1 = 65,535) models in which the explanatory variable appears</td>
<td>71</td>
</tr>
<tr>
<td>4-1. Relative evapotranspiration and precipitation data used to estimate residential water demand</td>
<td>98</td>
</tr>
<tr>
<td>A-1. Pearson’s test for correlation for 18 urban form metrics</td>
<td>114</td>
</tr>
<tr>
<td>A-2. Results for Moran’s I test for spatial dependence</td>
<td>115</td>
</tr>
<tr>
<td>B-1. Pearson’s moment correlations</td>
<td>123</td>
</tr>
<tr>
<td>B-2. Moran’s I test for spatial autocorrelation of model residuals</td>
<td>124</td>
</tr>
<tr>
<td>Figure</td>
<td>Page</td>
</tr>
<tr>
<td>--------</td>
<td>------</td>
</tr>
<tr>
<td>1-1. Conceptual model of human-integrated ecosystem (Grimm et al., 2000)</td>
<td>4</td>
</tr>
<tr>
<td>2-1. Study area (Salt Lake Valley) showing geographic distribution of three neighborhood types</td>
<td>33</td>
</tr>
<tr>
<td>2-2. Aerial photography showing neighborhoods representative of the three neighborhood types</td>
<td>34</td>
</tr>
<tr>
<td>2-3. Stripcharts for Density and Centrality metrics</td>
<td>35</td>
</tr>
<tr>
<td>2-4. Stripcharts for Accessibility and Neighborhood Mix metrics</td>
<td>36</td>
</tr>
<tr>
<td>3-1. Relative spatial distribution of urban canopy by neighborhood in Salt Lake County, Utah in 2006</td>
<td>72</td>
</tr>
<tr>
<td>3-2. Plots of simple slopes for 15 explanatory variables</td>
<td>73</td>
</tr>
<tr>
<td>4-1. Urban vegetation classification of 1-meter digital orthophotography used to attribute neighborhood polygons</td>
<td>99</td>
</tr>
<tr>
<td>4-2. Map of study area showing median age of housing stock by neighborhood</td>
<td>100</td>
</tr>
<tr>
<td>4-3. Plots showing the relationship of tree canopy to neighborhood age, and exposed turf grass cover to neighborhood age</td>
<td>101</td>
</tr>
<tr>
<td>4-4. Plot of predicted residential irrigation water demand compared to benchmark reported water use for 1975-2005</td>
<td>102</td>
</tr>
<tr>
<td>4-5. Maps of predictions of tree canopy, turf grass cover, and water demand for 2010 and 2040 under hypothetical scenario of no residential expansion</td>
<td>103</td>
</tr>
<tr>
<td>4-6. Estimated irrigation demand from 2010-2050 under hypothetical scenario of no residential expansion</td>
<td>104</td>
</tr>
<tr>
<td>4-7. Maps of predictions of tree canopy, turf grass cover, and water demand for 2010 and 2040 with simulated residential growth</td>
<td>105</td>
</tr>
<tr>
<td>4-8. Estimated water demand from 2010-2050 with simulated residential expansion</td>
<td>106</td>
</tr>
<tr>
<td>4-9. Plot of predictions and irrigation water demand for residential landscapes</td>
<td>107</td>
</tr>
</tbody>
</table>
5-1. Urban ecosystem conceptual model for Salt Lake Valley residential landscapes.................................................................110

A-1. Maps of urban form metrics 1-6.................................................................116
A-2. Maps of urban form metrics 7-12..............................................................117
A-3. Maps of urban form metrics 13-18............................................................118
B-1. Diagnostic plots for non-constant variance, normality, and leverage points....120
B-2. Partial residual plots..............................................................................121
B-3. Added variable plots..............................................................................122
B-4. Maps of determinants of urban tree canopy 1-6.................................123
B-5. Maps of determinants of urban tree canopy 7-12.................................124
B-6. Maps of determinants of urban tree canopy 13-15.................................125
CHAPTER 1
INTRODUCTION

The study of urban areas is increasingly becoming an important topic of academic research. At the turn of the 21st century, more than half of the world’s population lived in urban areas and the number of people moving to urban areas is expected to increase over the next 25 years (UNU/IAS, 2003). Tansley (1935) is credited with first using the term “ecosystem” to refer to a complex system of interaction and interdependence among biotic and abiotic components. Because urban areas comprise a variety of biotic (e.g. people, trees, etc.) and abiotic (e.g. buildings, streets, water, etc.) components that interact and are often interdependent upon one another, it is helpful to study urban areas as urban ecosystems.

Several conceptual models have been put forth as frameworks for integrating human and ecological processes (Pickett et al., 1997; Grimm et al., 2000; Pickett et al., 2001; Alberti et al., 2003; Redman et al., 2004). Common to all these models are linkages between socio-economic and biophysical variables driving human and natural ecosystem processes. The purpose of these conceptual models is to improve our understanding of human-environment interactions by presenting a framework for testing hypotheses regarding relationships among ecosystem processes, patterns, and functions.

Fig. 1-1 presents a human-ecosystem integrated model developed by Grimm et al. (2000). The authors suggest that to fully understand urban ecosystems, more research must be directed to the study of “patterns of human activities” (Grimm et al., 2000). According to the Grimm model, patterns of human activities, such as land use and management, have a reciprocal relationship with socioeconomic drivers. That is,
socioeconomic variables influence land use patterns, and land use patterns, in turn, influence socioeconomic variables. Chapter 2 of this research addresses the spatial and demographic patterns of residential neighborhoods. The aim is to quantitatively measure residential neighborhoods that were built during different time periods, and to assess whether significant differences in their spatial structure that can be attributed to the era during which they were built. Quantifying the pattern, or form, of the urban landscape is a first step toward gaining a better understanding of how abiotic components (e.g. built environment) of the urban ecosystem relate to the system’s biotic components (urban vegetation).

Chapter 3 investigates how socio-demographic and physical (built urban form and geophysical landscape) characteristics of neighborhoods are related to urban tree canopy. Grimm et al. (2000) suggest that a one-way relationship exists between patterns and processes of ecosystems and patterns of human activities (Fig. 1-1). The focus of chapter 3 is to demonstrate that this relationship is indeed reciprocal. That is, not only do patterns and processes of biophysical components on the landscape influence patterns of human activity, but patterns of human activity also influence biophysical patterns and processes.

The interplay between landscape patterns of human activities (i.e. where people live) and patterns and processes of biotic components of the system, becomes particularly relevant as we try to better understand how the ecosystem functions. The focus of chapter 4 is to better understand how spatial heterogeneity of urban vegetation in residential neighborhoods influences the demand for irrigation water. This chapter examines how the age of neighborhoods is related to the amount of tree canopy present in
a neighborhood, and how the composition of vegetation in residential neighborhoods
influences the demand for irrigation water in those neighborhoods and across the urban
landscape.

References

Alberti, M., Marzluff, J. M., Shulenberger, E., Bradley, G., Ryan, C., Zumbrunnen, C.,
2003. Integrating humans into ecology: opportunities and challenges for studying

long-term studies of urban ecological systems. BioScience 50 (7), 571-584.

Pickett, S. T., Burch, W. R. J., Dalton, S. E., Foresman, T. W., Grove, J. M., Rowntree,
R., 1997. A conceptual framework for the study of human ecosystems in urban
areas. Urban Ecosystems 1, 185-199.

C., Costanza, R., 2001. Urban ecological systems: linking terrestrial, ecological,
physical, and socioeconomic components of metropolitan areas. Annual Review

Redman, C. L., Grove, J. M., Kuby, L. H., 2004. Integrating social science into the Long-
Term Ecological Research (LTER) network: social dimensions of ecological
change and ecological dimensions of social change. Ecosystems 7, 161-171.

Tansley, A. G., 1935. The use and abuse of vegetational concepts and terms. Ecology 16,
284-307.

Nations Institute of Advanced Studies (UNU/IAS), Tokyo, Japan.
Fig. 1-1. Conceptual model of a human-integrated ecosystem (Grimm et al., 2000).
CHAPTER 2
MEASURING URBAN SPRAWL: A SPATIAL AND DEMOGRAPHIC CHARACTERIZATION OF RESIDENTIAL NEIGHBORHOODS IN SALT LAKE COUNTY, UTAH

Introduction

Over the course of the last century the spatial structure of cities in the U.S. has experienced unprecedented change (Jackson, 1985; Garreau, 1991). Cities at the beginning of the century were relatively compact and densely populated. Transportation was primarily by foot, wagon, or trolley (Jackson, 1985). By the end of the twentieth century the defining characteristic of most U.S. cities was, and still is, a heavy reliance on the automobile for transportation (Jackson, 1985; Batty et al., 2003). The spatial structure of many modern cities is therefore characterized by low-density, sprawling residential, commercial, and industrial development (Garreau, 1991; Ewing et al., 2002).

Quantitative Measures of Urban Form

Often the motivation to measure the social and spatial structure of cities, or “urban form,” is a desire to better understand urban sprawl. Many have pointed out that defining urban sprawl is not easy and we must first learn how to objectively measure it before we can properly define it (Galster et al., 2001; Torrens and Alberti, 2001; Ewing et al., 2002; Clifton et al., 2008; Frenkel and Ashkenazi, 2008; Torrens, 2008). One of the most comprehensive and detailed studies of the spatial dimensions of urban sprawl is that of Galster et al. (2001) who suggest eight conceptually distinct spatial measures of urban land use that characterize sprawl: density, continuity, concentration, clustering,
nuclearity, mixed uses, and proximity. Each of these indicators was used by Galster et al. (2001) to quantitatively measure the amount of sprawl for 13 U.S. Urban Areas. In another comprehensive study aimed at measuring urban sprawl and its impact, Ewing et al. (2002) used principal components analysis to group 22 sprawl variables into one of several factors, or sprawl types: density, land use mix, degree of centering, or street accessibility. Factor scores for each of the four sprawl indices were derived to produce “sprawl rankings” for 83 U.S. cities.

The ability to quantitatively measure urban form is useful because it allows us to objectively assess how urban form changes through time. In a study of Portland, Oregon’s urban form, Song and Knaap (2004) use a variety of metrics to assess urban form for three time periods of development. Portland is widely recognized as a city with progressive ideals in regard to urban planning, and in particular controlling sprawl (Song and Knapp, 2004; Envision Utah, 2009). In their study, Song and Knaap (2004) wanted to determine whether spatial measures of urban form offer empirical evidence that urban sprawl is decreasing in Portland. They found that starting in the 1990s several measures of sprawl had changed—for example, on average, newer residential neighborhoods are better connected and more pedestrian friendly.

The 1990s sets itself apart as an important time period in the social, political, and economic development of American cities. It was in the mid 1990s that the American Planning Association (APA) began promoting the concept of “Smart Growth” as a strategy for controlling urban sprawl (Knapp, 2005). Key principles of the “Smart Growth” movement include encouraging mixed land uses, developing walkable neighborhoods, promoting public transportation, and fostering communities with a strong
sense of place (Knapp, 2005). Related to Smart Growth, are the ideals espoused by the philosophy of “New Urbanism” (Leccese et al., 2000). Proponents of New Urbanism advocate an urban form reminiscent of residential development before World War II—in other words, compact, pedestrian friendly neighborhoods, mixed land uses, and easy access to public transit and activity centers (Leccese et al., 2000). In her study of urban form in Austin, Texas, Weston (2002) measured urban form to assess the feasibility of retrofitting current residential neighborhoods to New Urbanist ideals. Using spatial metrics of land use mix, street connectivity, housing types, and amount of open space, she measured the spatial dimensions of urban form for eight Austin neighborhoods. She concluded that if planners with New Urbanist ideals hope to encourage re-development of existing neighborhoods they must first know, in a quantitative fashion, how far off those neighborhoods are from the ideal.

Study Area and Envision Utah

Approximately 1 million people live in the 16 cities that comprise Salt Lake County, Utah (U.S. Census, 2009). Nearly all of the county’s population resides in a valley bounded by the Oquirrh Mountains to the west, Wasatch Range to the east, and Great Salt Lake to the north. In many respects urban expansion is limited by a natural growth boundary created by the physical features of the landscape. The areal extent of the valley bottom is approximately 800 km² (308 mi²). While a significant portion of the valley bottom remains undeveloped as agriculture and rangeland, the county has witnessed tremendous growth in the last 30 years. Between 1977 and 2006 approximately 140 km² (54 mi²) converted from agricultural production to suburban land
uses (GIS calculation using UDNR-DWR (2006) data), with population growth acting as the major driver of land use change.

Interest in growth issues for Salt Lake County and the Greater Wasatch Area, of which Salt Lake County is a part, can be traced to 1988 with the formation of the non-profit *Coalition for Utah’s Future* (Envision Utah, 2009). Comprised of business leaders and community members, the Coalition was interested in multiple issues pertaining to growth, including looking at ways to encourage economic growth throughout the state. By 1995 the Coalition began to become concerned about how growth would affect Utah’s quality of life and formed the *Quality Growth Steering Committee* which included business leaders and a representative from the Governor’s Office of Planning and Budget (Envision Utah, 2009). Out of the Quality Growth Steering Committee, *Envision Utah*, a public/private partnership was formed with a mission to “guide the development of a broadly and publicly supported Quality Growth Strategy—a vision to protect Utah’s environment, economic strength, and quality of life” (Envision Utah, 2009).

To fulfill its mission Envision Utah began by researching quality/smart growth strategies underway in California, Denver, and Portland. Recognizing Utah’s unique political climate where local control is revered, Envision Utah chose to adopt a strategy that focused on involving the public and local government leaders rather than attempt to create a governmental regional planning organization like Portland’s *Metro 2040* (Envision Utah, 2009). The strategy chosen by Envision Utah involved organizing workshops aimed at presenting different growth scenarios to local leaders and community members. Growth scenarios are based on surveys administered to workshop participants and presented as alternative growth maps created within a geographic information
system. The strategy therefore aims to educate the public and local leaders about land use planning and growth options, with the idea that an informed public and electorate will make better decisions about growth if they can visualize different options presented as maps (Envision Utah, 2009).

**Study Objectives**

This paper presents a study of urban form by measuring the spatial and demographic characteristics of different residential neighborhood types in Salt Lake County, Utah. Our first objective is to quantitatively measure observed spatial and demographic characteristics of residential neighborhoods. To this end we have adopted many of the quantitative metrics suggested by others (Torrens and Alberti, 2000; Galster et al., 2001; Ewing et al., 2002; Weston, 2002; Torrens, 2008). Following Ewing et al. (2002) we group these metrics into four categories: *Density, Centrality, Accessibility,* and *Neighborhood Mix.*

Our second objective is to determine whether the spatial and demographic form of present-day neighborhoods differs according to the time period during which they were developed. Contemporary residential neighborhoods reflect social and cultural values of the past as much as they reflect where people choose to live today. Urban infrastructure is largely a relict of the time period during which the land was developed. Where people live today and how they choose to develop new neighborhoods reflects contemporary urban planning and policy, as well as current social and economic conditions. By examining the spatial form of residential neighborhoods today we can learn something about how historical events, policies, and culture shaped the physical form of the urban
landscape, and make an assessment of contemporary social values with regards to residential development.

Three neighborhood types are identified: neighborhoods developed in the pre-suburban era (1945 and before), suburban era (after 1945 to 1989), and late-suburban era (1990 and after). By many accounts (Mieszkowski and Mills, 1993; Lindsrom and Bartling, 2003) the pre-suburban era ended at the close of World War II—the economy was strong, returning servicemen and women received government loans to build new homes, and automobiles were affordable. Arguably, suburbanization continues today. However as noted previously, the 1990s mark a significant time period during which professionals and laypersons became more actively involved in addressing urban growth issues than at any time period before (Knapp, 2005). In Portland, Song and Knaap (2004) found that efforts initiated in the 1990s to curb sprawling growth appear to be having an effect in that city. The 1990s marked a similar turning point in Utah and the Salt Lake County area (Envision Utah, 2009). This research aims to determine whether there is evidence suggesting empirical differences in urban form between neighborhoods developed during suburban and late-suburban eras, as well as determine whether there are differences (or similarities) between suburban or late-suburban neighborhoods and neighborhoods developed before the end of World War II.

Methods

To meet our first objective it was necessary to spatially define what constitutes a “neighborhood” and then generate the various metrics of urban form for each neighborhood in the study area. This was accomplished using a geographic information
system (GIS) and readily available U.S. Census data and county GIS datasets. For our second objective—determining whether urban form metrics differ by neighborhood type (Pre-Suburban, Suburban, Late-Suburban) we use analysis of variance (ANOVA) to compare the urban form metric means across the three groups.

When fitting linear models, it is important to assess and account for spatial dependence in model residuals. To do this we first assessed the model residuals (assuming independent errors) using the Moran’s I statistic, and then used a Simultaneous Autoregressive (SAR) model to account for the spatial dependence, if present. In what follows, we describe our metrics of urban form and their data sources, the neighborhoods (sample observations) and neighborhood groups, and our implementation of ANOVA using SAR.

*Defining Residential Neighborhoods*

What constitutes a neighborhood has been a matter of debate among urban planners and scholars for much of the last century. Neighborhoods have been defined variously as spatial units with relatively homogenous social and economic characteristics, residential areas with little through traffic, or as catchment areas for the local elementary school (Lynch, 1984). Because neighborhoods are typically considered the basic building block of urban form (Lynch, 1984) we sought a spatial unit that met these general conditions of relative spatial and demographic homogeneity. Looking to the U.S. Census we considered the *census block*, the *block group*, and the *census tract* for our neighborhood unit. In the geographic hierarchy of the U.S. Census, census blocks are subdivisions of block groups and census tracts are aggregates of block groups. Census
tracts are considered to be fairly homogenous areas with similar social, economic, and demographic characteristics (Peters and MacDonald, 2004). However, the number of people per census tract ranges from 1,500 to 8,000 with 4,000 considered the optimum number by the Census Bureau (Peters and MacDonald, 2004). Desiring a smaller spatial unit, we decided the block group, with 1,500 people considered the optimum size by the Census Bureau, to be more appropriate for our study.

Faced with a similar decision, Song and Knaap (2004) computed several measures of urban form at each of the three spatial units provided by the U.S. Census and plotted the data against the median age of residential neighborhoods. They found that data for neighborhoods at the census tract level provided less information about urban form over time than data for neighborhoods at the block group level (Song and Knapp, 2004). They also found that information gained at the block level was minimal compared to the block group. This helped confirm our choice of the block group as a reasonable spatial unit with which to define neighborhoods.

Census units in the U.S. Census do not distinguish between land uses. In other words, a census block group or census tract may contain land uses besides residential. Because the focus of our study was on residential neighborhoods, and many of our metrics are based on the spatial configuration of urban infrastructure such as residential roads, we identified “neighborhoods” as the residential portion of the census block group spatial unit. To create our “neighborhoods” GIS dataset, we selected residential parcels (single family) from the county parcels GIS dataset, dissolved them to create residential land use polygons, and intersected those polygons with the census block group dataset to
create a dataset with 542 neighborhoods defined as the *residential portion of each census block group* (Fig 2-1).

**Neighborhood Groups**

We assigned all residential neighborhoods to one of three neighborhood types based on the era during which they were developed. Neighborhood types were *Pre-Suburban* (developed in 1945 or before), *Suburban* (developed after 1945 through 1990) and *Late-Suburban* (developed after 1990). Neighborhood age was determined by calculating the median age of all residential buildings in the neighborhood using county parcel data (Fig. 2-1). Figure 2-2 presents examples of the three neighborhood types depicted at the same map scale with high resolution aerial orthophotography.

**Data Sources and Metrics of Urban Form**

As demonstrated in the literature (Torrens and Alberti, 2000; Galster et al., 2001; Ewing et al., 2002; Weston, 2002; Clifton et al., 2008), a wide variety of spatial and demographic metrics have been posited as good measures of urban form. Most would agree that urban form is really a multi-dimensional phenomenon (Ewing et al., 2002) and cannot be measured by any single metric. The literature suggests that urban form can be characterized by common measures that closely align with the four categories suggested by Ewing et al. (2002). *Density* for example is widely regarded a good measure of urban form (Torrens and Alberti, 2000; Galster et al., 2001; Ewing et al., 2002). Density however can be measured in different ways. Therefore, rather than reduce multiple density metrics to a single index, as was done by Ewing et al. (2002) we chose to use
several different density metrics, each of which characterizes a different aspect of density as it relates to urban form.

Below, and in Table 2-1 we briefly describe 18 metrics of urban form grouped into four categories: *Density, Centrality, Accessibility, and Neighborhood Mix*. As noted in Table 2-1 many of these metrics have been suggested by others in the literature. Ultimately our choice of metrics was based on our satisfaction that each of these metrics offers a valuable measure of urban form, and could be derived using available data. Because of some concern for redundancy among the indices we assessed their correlation using a Pearson’s test for correlation (Appendix A, Table A-1).

*Density Metrics*

Density is perhaps the most intuitive measure of urban form, and density metrics provide a great deal of useful information about urban and suburban living. For example, living in low-density neighborhoods increases dependence on automobiles, with potentially adverse health and environmental effects (Johnson, 2001; Ewing et al., 2002). Although population density is the most common measure of urban density, *housing density* is considered a better measure of the physical condition of land use, and therefore urban form (Galster et al., 2001; Theobald, 2002). We found these two metrics to be highly correlated ($\rho = 0.9$) but decided to include both in our study because of the common use of population density in other studies (Ewing et al., 2002). Perhaps a better measure of density as a physical condition of land use is median size of residential lots (Ewing et al., 2002; Song and Knapp, 2004) which we obtained from a GIS dataset of residential parcels. Measures of the size of homes and number of people living in homes
shed some light on societal values and affluence of neighborhoods. These data were available through the U.S. Census as the *median number of rooms per housing unit*, and *population/number of housing units*.

**Centrality Metrics**

Centrality is a measure of the strength of activity centers, such as the central business district or other commercial centers on the urban landscape (Ewing et al., 2002). These are important because they measure the separation between where people live and where they must go for common daily activities (Song and Knapp, 2004). In addition to a metric of *mean distance of neighborhoods to commercial centers* we include *mean distance to bus stops*, *mean distance to neighborhood parks*, and *mean distance to K-12 schools* as other metrics of centrality. These were computed using available county GIS datasets and the Euclidean distance function within the GIS (mean distance (as a zonal measurement) to activity center).

**Accessibility Metrics**

Accessibility refers to street design and the connectivity of street networks. Critics of sprawl contend that neighborhoods with winding streets, large residential blocks, and long cul-de-sacs are not pedestrian friendly (Ewing et al., 2002). Smart Growth principles specifically address these urban landscape forms and encourage polices and planning incentives to discourage this form of development. Metrics we include were obtained from street centerline and parcel GIS datasets. We use the *ratio of streets to intersections* and *ratio of cul-de-sacs to thru-streets* as measures of network connectivity (Weston, 2002; Song and Knapp, 2004). Additional metrics we include are:
median size (perimeter) of residential blocks and median length of cul-de-sacs in residential neighborhoods, obtained from parcel and street centerline GIS datasets.

**Neighborhood Mix Metrics**

Neighborhood mix refers to measures of spatial and demographic heterogeneity. Neighborhoods built during the pre-suburban era are often thought to comprise a more heterogeneous mix of land uses by including commercial “corner stores” interspersed within residential neighborhoods (Weston, 2002). Zoning restrictions, more common after World War II, tend to encourage segregating residential subdivisions from commercial activities. Subdivision development also encouraged (either intentionally or unintentionally) social and economic segregation. We include three metrics addressing different aspects of spatial heterogeneity of land uses. Borrowing landscape metrics common to the discipline of landscape ecology (McGarigal et al., 2002) we calculate metrics for *land use contiguity* (Juxtaposition and Interspersion (IJI), or distribution of adjacent land uses), *land use richness* (Patch Richness (PR), or number of different land use types per 100 hectares), and *land use diversity* (Simpsons Diversity Index (SIDI), or spatial evenness of land uses) from a parcel GIS dataset with attributes for land use type. Measures of demographic heterogeneity in the literature are the ratio of home-renters to home-owners (Torrens, 2008) and the proportion of people working outside the city of residence (Glaser et al., 2001; Ewing et al., 2002) both of which were available through the U.S. Census.
Statistical Approach

Analysis of variance (ANOVA) is a statistical method used to test for significant differences between or among the means of two or more groups. It can also be considered a special case of the general linear model (Hamilton, 1992; Lattin et al., 2003). In this case the model is viewed as the effect of a factor variable on a continuous dependent variable, with the factor variable encoded as dummy variables in the regression equation:

\[ Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \epsilon_i \]

For our study, \( Y_i \) represents the urban form metric \( i \), a continuous variable, and \( X_{i1} \) and \( X_{i2} \) are indicator variables for the Pre-Suburban and Suburban groups. Pre-Suburban neighborhoods are encoded with a 1 for variable \( X_1 \) and 0 for variable \( X_2 \), with the opposite applying to Suburban neighborhoods. Late-Suburban neighborhoods are considered the omitted group, and are encoded 0 in both the \( X_1 \) and \( X_2 \) variables.

An advantage to viewing ANOVA as a regression model is that it allows for the estimation of regression coefficients that can be used to make inference (e.g., confidence intervals and hypothesis tests) concerning group means (Hamilton, 1992). Here, the mean for the omitted group (Late-Suburban) equals the intercept \( \beta_0 \), while the mean for the Pre-Suburban group is \( \beta_0 + \beta_1 \), and the mean for the Suburban group is \( \beta_0 + \beta_2 \) (Hamilton, 1992). When making simultaneous inference concerning more than two groups, p-values tend to be exaggerated and should be corrected to ensure the desired
level of significance is achieved. For this study we corrected for multiple testing using the Bonferroni correction method (Dalgaard, 2002).

The assumptions for ANOVA are the same as general linear models using an Ordinary Least Squares (OLS) method for fitting. Conventional inference concerning unknown parameters (e.g. regression coefficients, group means) relies on the assumption that model errors are normally distributed, independent, and identically distributed (Hamilton, 1992). With spatial data, observed errors are frequently spatially autocorrelated. In our study for example, we could expect two adjacent neighborhoods to be similar, even after accounting for a particular measure, such as housing density. These additional forms of dependence often manifest themselves in the model residuals, thus violating a key assumption of the conventional independent error model and resulting in a biased estimation of regression coefficients (Bailey and Gatrell, 1995; Bivand et al., 2008).

The Simultaneous Autoregressive (SAR) model is one of several spatial regression models that accounts for spatial dependence, and makes use of it to improve parameter estimation (Bailey and Gatrell, 1995; Bivand et al., 2008). The SAR model accounts for spatial dependence by linking each observation with its neighbors, thus, a key component of the model is specification of spatial neighbors. Spatial neighbors can be specified as those polygons that share contiguous boundaries, the nearest centroid, or centroids within a certain distance of one another (Bivand et al., 2008). Once the SAR model has been specified, and parameters are estimated, model residuals are tested to determine whether spatial dependence has been adequately accounted for (Bailey and
Gatrell, 1995). A common test for spatial dependence is the Moran’s I test for spatial autocorrelation (Wong and Lee, 2005).

Before fitting a SAR model, however, it is common to determine whether SAR is necessary by first fitting a conventional OLS model, and checking the residuals for spatial dependence (Bailey and Gatrell, 1995; Bivand et al., 2008). If they are not independent, proceeding with a SAR model is warranted. For our study we regressed each of the 18 urban form metrics to the factor variable for neighborhood type using a conventional linear regression model, and tested the residuals for spatial dependence using Moran’s I (Appendix A, Table A-2). Values for the Moran’s I statistic range from -1 indicating high negative spatial dependence, or regularity in spatial dispersion, to 1 indicating high positive spatial dependence or clustering (Wong and Lee, 2005). Less extreme Moran’s I values generally indicate a random spatial pattern. Residuals for all 18 initial models exhibited high positive spatial dependence, thus the SAR model was warranted.

The choice of neighbor specification for a SAR model is often determined by a theory-based assumption of how observation units interact on the landscape (Bivand et al., 2008). In our study this means how neighborhoods interact within one another—for example, is the density of one neighborhood most correlated to the single nearest neighborhood, all adjoining neighborhoods, or all the neighborhoods that are within a given distance? Lacking theory-based evidence to support one assumption over another, we chose the most parsimonious approach. For our SAR models, we defined neighbors as the nearest neighbor based on distance between neighborhood centroids. Fitting SAR
models in this manner produced models with negligible spatial dependence in the
residuals as noted by the Moran’s I statistic (Appendix A, Table A-2).

**Results**

Table 2-2 compares population means of the three neighborhood types for the 18
urban form metrics. Population means are estimated using the methods described above,
and a z-test is used to test for equality of means. Larger z-scores indicate greater
difference between two group means, with statistically significant differences denoted at
\( \alpha = 0.10, 0.05, \) and 0.01 levels by asterisks.

Figs. 3 and 4 present strip charts produced using Monte-Carlo simulations of 1000
observations based on the estimated coefficients and residual variances from the SAR
model, for each of the 18 urban metrics. Error bars on the charts mark positions 1 and 2
standard deviations from the mean, which is marked with an X. The strip charts are
useful because they visually graph differences in means presented in Table 2-2, and they
convey the variability of urban form metrics by neighborhood era. As an example, we
note that the while there is little difference in means for *Median Block Perimeter Size*
between the Suburban era and the Late-Suburban era, the variance around the mean in the
Late-Suburban era is nearly double the variance of the Suburban era neighborhoods—in
other words, there is a greater variety of neighborhood block sizes in Late-Suburban
neighborhoods (1990 and after) than in neighborhoods built during the Suburban era
Discussion

Comparison of Pre-Suburban (Pre WWII) to Suburban Neighborhoods

A comparison of urban form metrics for neighborhoods developed during the Pre-Suburban era to neighborhoods of the Suburban era lends empirical evidence to the hypothesis that the spatial structure of urban neighborhoods experienced significant changes following World War II. While not explicitly examined in this study, the implication is that these differences can be attributed to differences in economic conditions, societal values, and available technology (namely transportation) during the two time periods.

We found significant differences ($\alpha = 0.05$) in mean values for 11 of the 18 urban form metrics. Perhaps the most striking differences are in the density and accessibility metrics, both of which are measures of the impact of the automobile on residential development. Median lot size increased from 0.05 hectares (0.13 acres) in the Pre-Suburban era to 0.11 hectares (0.25 acres) during the Suburban era. The ratio of through streets to cul-de-sacs provides a measure of street connectivity, and changes significantly from 21.70 through streets to every cul-de-sac in the Pre-Suburban neighborhoods to 11.08 through streets per cul-de-sac in Suburban era neighborhoods. In other words, the use of cul-de-sacs in neighborhood design during the Suburban era nearly doubled.

Other metrics not as explicitly tied to automobile use are metrics for centrality and neighborhood mix. It should also be noted that these metrics measure characteristics of urban form in the present-day for neighborhoods developed at different time periods in the past. For example, the centrality metric mean distance to commercial areas
(COMMEAN) is the mean distance of neighborhoods to present-day commercial areas. This type of metric, while not a measure of how the neighborhood was developed, is useful because it provides an assessment of urban form across the landscape as it is today, and in the case of distance to commercial areas, tells us how much travel is required of people living in older (Pre-Suburban) neighborhoods as opposed to middle aged (Suburban) neighborhoods. What is striking about the centrality metrics is that we do not see much difference between the two neighborhood types. And in particular for distance to parks and K-12 schools we find no significant difference. This is likely attributed to the fact that schools and parks were built roughly at the same time period during which the neighborhoods were developed.

Neighborhood mix metrics, like centrality metrics, tell us something about how the neighborhoods fit into the spatial and demographic fabric of the urban landscape today. Here we find significant differences in the metric for land use contiguity (IJI). Higher IJI values indicate “more even” interspersion and juxtaposition, while lower values indicate “less even” interspersion and juxtaposition, or clumping (McGarigal et al., 2002). The differences we find between the Pre-Suburban and Suburban era neighborhoods suggest a more heterogeneous mix of land uses in the Pre-Suburban era neighborhoods than the Suburban era neighborhoods. Again, because this is based on present-day land uses, the more heterogeneous nature of Pre-Suburban neighborhoods may be a relict of how the neighborhoods were developed (i.e. non-residential land uses juxtaposed to residential land uses) or current zoning policies in older neighborhoods allowing mixed uses, or a combination of the two. Finally, the two neighborhood mix metrics, proportion of people working outside their city of residence, and the ratio of
renters to owners, tell us something about the present-day demographic characteristics of these neighborhoods. For proportion of people working outside their city of residence a mean value of 0.5 would indicate an even mix, or heterogeneous composition of workers. For the ratio of renters to owners, an even mix of renters to owners would have a mean value of 1.0. Our findings indicate that the proportion of people working outside their city of residence is significantly higher in the Suburban era neighborhoods than the Pre-Suburban neighborhoods, and the ratio of renters to owners is nearly three times higher in older (Pre-Suburban) neighborhoods than in middle aged (Suburban era) neighborhoods. This is evidence that more people living in suburban neighborhoods own the homes they live in and commute to their workplaces.

Comparison of Suburban to Late-Suburban (1990 and after) Neighborhoods

Perhaps of greatest interest is the comparison of Suburban era and Late-Suburban era neighborhoods, as this provides some indication of how well principles of Smart Growth are succeeding in metropolitan Salt Lake County. Beginning with the density metrics we find that in general, the trend toward lower density residential housing continues in Late Suburban era neighborhoods. Where the median lot size from 1945 to 1989 was 0.10 hectares (0.25 acres), the median lot size is now 0.12 hectares (0.30 acres). Population density also continues to decrease, from 44.20 people/sq. km. to 35.25 people/sq. km. However, the number of people per home is slightly higher at 3.09 people/home in Late-Suburban neighborhoods compared to 2.91 people/home in Suburban era neighborhoods. Two metrics that are not significantly different are the median number of rooms per single family home, and housing density, as measured by
housing units per square kilometer. The former suggests the size of single family homes remains the same in both neighborhood types, and the latter seems to contradict the finding that median lot sizes are increasing. This apparent contradiction could be explained by the way housing density was calculated within the GIS, or by the fact that housing density has indeed decreased (from 16.14 houses/sq. km. to 12.93 houses/sq. km) but the difference is not found to be statistically significant at \( \alpha = 0.05 \). In general, our analyses suggest that housing and population density is lower in Late-Suburban era neighborhoods than Suburban era neighborhoods, and that having multiple metrics for measuring density provides a more complete description of density as a dimension of urban form.

The accessibility metrics provide information about residential design and street connectivity, with the assumption that greater connectivity lends itself to more walkable neighborhoods. Here we find that two of the four metrics are significantly different between neighborhoods developed during the two time periods. Street connectivity, as measured by the ratio of streets (links) to intersections (nodes) has decreased from 1.92 to 1.80, and the average length of cul-de-sacs has increased from 68.24 meters to 80.61 meters. The other two metrics suggest there has been no significant change in the size of residential blocks, and the number of through streets per cul-de-sac.

When we examine the centrality and neighborhood mix metrics we find additional information about present-day spatial and demographic characteristics of these two neighborhood types. Notably for the centrality metrics, we find no significant difference for distance to commercial areas, or distance to neighborhood parks. What we do find is a significant difference between the Suburban and Late-Suburban neighborhoods for
distance to K-12 schools and distance to public transportation (bus stops). One possible explanation is that a time-lag exists between the time a new neighborhood is developed and when schools are built and bus stops put in place. Given this explanation, services such as schools and bus stops will eventually be as accessible in the Late-Suburban neighborhoods as they presently are in Suburban era and Pre-Suburban era neighborhoods. An alternative explanation is that Late-Suburban neighborhoods, and the people living in them, are even more automobile-oriented than their Suburban neighborhood counterparts. People living in these neighborhoods are willing to drive longer distances (or assume the tax burden for school buses) required for school access, and do not value public transportation enough to demand access to it in their neighborhoods.

Examining the neighborhood mix metrics we find that the proportion of people working outside their city of residence, and the ratio of renters to owners are not significantly different between the two neighborhood types. Residents in Suburban neighborhoods commute just as much as residents in Late-Suburban neighborhoods (though distances may be different), and the ratio of renters to owners are not significantly different. However, significant differences are observed for the land use contiguity and land use richness metrics; though it should be noted however that these metrics operate irrespective of land use class and are simply measures of juxtaposition and interspersion of different land uses in the former, and the number of different land use classes in the latter. So the observation that greater land use heterogeneity exists in Late-Suburban neighborhoods is not qualitatively the same as the heterogeneity observed in the Pre-Suburban neighborhoods. It is most likely that land uses in Late-Suburban
neighborhoods are more heterogeneous than Suburban neighborhoods because of the interspersion of vacant and agricultural land. These differences suggest a detection of more discontinuous residential development, often referred to as “leap-frog” sprawl, in Late-Suburban neighborhoods.

_Comparison of Pre-Suburban (Pre WWII) to Late-Suburban neighborhoods_

A comparison of Pre-Suburban neighborhoods to Late-Suburban neighborhoods is relevant in light of the ideals espoused by New Urbanist philosophy; namely a return to an urban form reminiscent of the pre World War II era. Our results show that across nearly all metrics, urban form in Late-Suburban neighborhoods is significantly different from Pre-Suburban (Pre-World War II) era neighborhoods. A few metrics did not show a significant difference. We found no significant difference in the _distance to neighborhood parks_ between all three neighborhood types. It is notable that _median cul-de-sac length_ in the Late-Suburban era is about the same as the Pre-Suburban era. However there are approximately three times as many cul-de-sacs in Late-Suburban neighborhoods compared to Pre-Suburban neighborhoods. Finally, measures of neighborhood mix as measured by _land use contiguity_ and _land use diversity_ (spatial evenness of land uses) are not statistically different, though as noted previously there are likely qualitative differences not captured by the metrics we used.

**Conclusion**

Our analysis reveals significant differences in the spatial and demographic characteristics of the three residential neighborhood types we studied. The spatial structure of neighborhood types is very much a relict of the time period during which...
they were developed. The differences we report between Pre-Suburban and Suburban neighborhoods are not surprising given the widely recognized changes that occurred in U.S. cities across these two time periods (Mieszkowski and Mills, 1993; Lindsrom and Bartling, 2003). Of greater interest and more immediate concern is what we find when we compare Suburban and Late-Suburban era neighborhoods.

Beginning in the 1990s, concern for uncontrolled urban growth reached the consciousness of professionals and laypersons alike. In a study similar to ours Song and Knaap (2004) report that the city of Portland, OR appears to be winning the battle against urban sprawl. While they do not explicitly correlate the policies of Portland’s growth management program to their observations of an urban form more characteristic of Smart Growth, they suggest that the changes observed after 1990 could very well be a result of such policies.

Our findings suggest that Salt Lake County is not doing as well in the battle against sprawling urban growth. Many factors might explain the lack of shift toward “smarter growth.” Geographically the county has room to grow. As of 2006, 165 km² (64 mi²) or 20% of the valley bottom was still in agricultural production (GIS calculation with UDWR (2006) data) and in prime position for residential expansion. Utah’s population is young and growing, fueling demand for additional homes (U.S. Dept. of HUD, 2009). Generally the political climate is conservative, favoring local control of land use planning rather than planning on a regional scale (Envision Utah, 2009). The strategies employed by Envision Utah are not enforceable regulations, but are rather guiding principles, offered to local decision makers based on information obtained from the public through informal workshops. Finally, Utah experienced high economic growth...
during the 1990s and early 2000s (U.S. Dept. of HUD, 2009) as did much of the nation. These combined factors greatly influence the spatial and demographic pattern of residential neighborhoods observed in this study.

References


UDNR-DWR, 2006. GIS dataset of water-related land uses. GIS dataset of water-related land uses., Utah Department of Natural Resources, Division of Water Resources, Salt Lake City, UT.


Table 2-1. List of urban form metrics and data sources.

<table>
<thead>
<tr>
<th>DENSITY METRICS</th>
<th>Data Source(s)</th>
<th>Name</th>
<th>Reference(s)</th>
<th>Measurement of</th>
</tr>
</thead>
<tbody>
<tr>
<td>Median Single Family Res. Lot Size</td>
<td>County Parcels</td>
<td>MEDLOTACRE</td>
<td>Song and Knaap 2004; Ewing et al. 2002</td>
<td>Lot size</td>
</tr>
<tr>
<td>Housing Density (units/sq. km.)</td>
<td>Census Data &amp; County Parcels</td>
<td>HUTPERSQKM</td>
<td>Song and Knaap 2004; Galster et al. 2001; Theobald 2002</td>
<td>Housing density</td>
</tr>
<tr>
<td>Median Number of Rooms</td>
<td>Census Data</td>
<td>MEDNURMS</td>
<td>Song and Knaap 2004</td>
<td>Home size or floor space</td>
</tr>
<tr>
<td>Population Density (pop./sq. km.)</td>
<td>Census &amp; County Parcels</td>
<td>POPPERSQKM</td>
<td>Ewing et al. 2002</td>
<td>Residential Population Density</td>
</tr>
<tr>
<td>People/Housing Unit</td>
<td>Census Data</td>
<td>POPPERHU</td>
<td>Song and Knaap 2004</td>
<td>Living density</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>CENTRALITY METRICS</th>
<th>Data Source(s)</th>
<th>Name</th>
<th>Reference(s)</th>
<th>Measurement of</th>
</tr>
</thead>
<tbody>
<tr>
<td>Distance to Commercial Zone</td>
<td>County Parcels Centroids (commercial) &amp; residential parcels</td>
<td>COMMEAN</td>
<td>Galster et al., 2002; Song and Knaap 2004</td>
<td>Strength/Access to Activity Center</td>
</tr>
<tr>
<td>Distance to Neighborhood and City Parks</td>
<td>County Parks Dataset</td>
<td>PRKMEAN</td>
<td>Song and Knaap 2004</td>
<td>Strength/Access to Activity Center</td>
</tr>
<tr>
<td>Distance to K-12 Schools</td>
<td>County K-12 Schools Point Dataset</td>
<td>SCHMEAN</td>
<td>Song and Knaap 2004</td>
<td>Strength/Access to Activity Center</td>
</tr>
<tr>
<td>Distance to Bus Stops (public transportation)</td>
<td>Bus Stops Point Dataset</td>
<td>BUSMEAN</td>
<td>Song and Knaap 2004; Weston2002</td>
<td>Strength/Access to Activity Center</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>ACCESSIBILITY METRICS</th>
<th>Data Source(s)</th>
<th>Name</th>
<th>Reference(s)</th>
<th>Measurement of</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ratio of Streets to Intersections</td>
<td>County Streets &amp; County Parcels</td>
<td>R_LINK_NOD</td>
<td>Song and Knaap 2004; Weston 2002</td>
<td>Street design &amp; connectivity</td>
</tr>
<tr>
<td>Median Perimeter of Residential Blocks</td>
<td>County Parcels</td>
<td>MED_BLK_PE</td>
<td>Song and Knaap 2004</td>
<td>Neighborhood design</td>
</tr>
<tr>
<td>Ratio of Streets to Cul-de-sacs</td>
<td>County Streets &amp; County Parcels</td>
<td>R_LINKS_CUL</td>
<td>Weston 2002</td>
<td>Street design &amp; connectivity</td>
</tr>
<tr>
<td>Median Length of Cul-de-sacs</td>
<td>County Streets &amp; County Parcels</td>
<td>MED_CUL_LE</td>
<td>Song and Knaap 2004</td>
<td>Street design &amp; connectivity</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>NEIGHBORHOOD MIX METRICS</th>
<th>Data Source(s)</th>
<th>Name</th>
<th>Reference(s)</th>
<th>Measurement of</th>
</tr>
</thead>
<tbody>
<tr>
<td>Land use Contiguity</td>
<td>County Parcels (all land uses)</td>
<td>UJ</td>
<td>Torrens &amp; Alberti 2000</td>
<td>Clumped or dissectedness</td>
</tr>
<tr>
<td>Land use Richness</td>
<td>County Parcels (all land uses)</td>
<td>PR</td>
<td>Number of land uses</td>
<td></td>
</tr>
<tr>
<td>Land use Diversity</td>
<td>County Parcels (all land uses)</td>
<td>SIDI</td>
<td>Weston 2002</td>
<td>Relative evenness of land uses</td>
</tr>
<tr>
<td>Proportion of People Working Outside City of Residence</td>
<td>2000 Census</td>
<td>PRPWRKOUT</td>
<td>Ewing et al. 2002; Glaeser, 2001</td>
<td>Balance of jobs to residences</td>
</tr>
<tr>
<td>Ratio of Renters to Owners</td>
<td>Census Data</td>
<td>R_RENT_OWN</td>
<td>Torrens &amp; Alberti 2000</td>
<td>Balance of renters to owners</td>
</tr>
</tbody>
</table>

Note: Only residential parcels were used unless otherwise noted. Residential neighborhoods were identified as the residential portion of the Census Block Groups using Parcel GIS dataset to identify residential areas. All Census information was from the Estimated 2007 Census unless otherwise noted.
Table 2-2. Comparison of urban form metrics for neighborhood types. See Table 2-1 for descriptions of metric names.

Comparison of Population Means for Neighborhood Types based on Development Era

<table>
<thead>
<tr>
<th>Metric Name</th>
<th>Pre-Suburban</th>
<th>Suburban</th>
<th>Late Suburban</th>
<th>Pre-Suburban</th>
<th>Suburban</th>
<th>Late Suburban</th>
<th>Z-Score</th>
<th>Mean</th>
<th>Mean</th>
<th>Z-Score</th>
<th>Mean</th>
<th>Mean</th>
<th>Z-Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>MEDLOTACRE</td>
<td>0.13</td>
<td>0.25</td>
<td>8.778</td>
<td>***</td>
<td>0.25</td>
<td>0.30</td>
<td>-3.596</td>
<td>***</td>
<td>0.13</td>
<td>0.30</td>
<td>-9.442</td>
<td>***</td>
<td></td>
</tr>
<tr>
<td>MEDNURMS</td>
<td>5.15</td>
<td>6.50</td>
<td>7.097</td>
<td>***</td>
<td>6.50</td>
<td>6.60</td>
<td>-0.603</td>
<td></td>
<td>5.15</td>
<td>6.60</td>
<td>-6.005</td>
<td>***</td>
<td></td>
</tr>
<tr>
<td>POPPERSQKM</td>
<td>78.94</td>
<td>44.20</td>
<td>-9.526</td>
<td>***</td>
<td>44.20</td>
<td>35.25</td>
<td>2.504</td>
<td>**</td>
<td>78.94</td>
<td>35.26</td>
<td>9.264</td>
<td>***</td>
<td></td>
</tr>
<tr>
<td>POPPERHU</td>
<td>2.38</td>
<td>2.91</td>
<td>6.335</td>
<td>***</td>
<td>2.91</td>
<td>3.09</td>
<td>-2.448</td>
<td>**</td>
<td>2.38</td>
<td>3.09</td>
<td>-6.722</td>
<td>***</td>
<td></td>
</tr>
<tr>
<td>COMMEAN</td>
<td>3400.88</td>
<td>3681.01</td>
<td>2.810</td>
<td>**</td>
<td>3681.01</td>
<td>3837.95</td>
<td>-2.130</td>
<td>***</td>
<td>3400.88</td>
<td>3837.95</td>
<td>-3.537</td>
<td>***</td>
<td></td>
</tr>
<tr>
<td>PRKMEAN</td>
<td>628.74</td>
<td>701.60</td>
<td>1.572</td>
<td></td>
<td>701.60</td>
<td>731.55</td>
<td>-0.785</td>
<td></td>
<td>628.74</td>
<td>731.55</td>
<td>-1.774</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SCHMEAN</td>
<td>2113.99</td>
<td>2136.25</td>
<td>0.177</td>
<td></td>
<td>2136.25</td>
<td>2779.44</td>
<td>-5.612</td>
<td>***</td>
<td>2113.99</td>
<td>2779.44</td>
<td>-4.151</td>
<td>***</td>
<td></td>
</tr>
<tr>
<td>BUSMEAN</td>
<td>322.08</td>
<td>411.31</td>
<td>2.215</td>
<td>*</td>
<td>411.31</td>
<td>530.06</td>
<td>-3.806</td>
<td>***</td>
<td>322.08</td>
<td>530.07</td>
<td>-4.183</td>
<td>***</td>
<td></td>
</tr>
<tr>
<td>R_LINK_NOD</td>
<td>2.24</td>
<td>1.92</td>
<td>-7.212</td>
<td>***</td>
<td>1.92</td>
<td>1.80</td>
<td>2.697</td>
<td>**</td>
<td>2.24</td>
<td>1.81</td>
<td>7.614</td>
<td>***</td>
<td></td>
</tr>
<tr>
<td>MED_BLK_PE</td>
<td>356.69</td>
<td>593.49</td>
<td>8.137</td>
<td>***</td>
<td>593.49</td>
<td>588.42</td>
<td>0.167</td>
<td></td>
<td>356.69</td>
<td>588.43</td>
<td>-6.064</td>
<td>***</td>
<td></td>
</tr>
<tr>
<td>R_LINK_CUL</td>
<td>21.70</td>
<td>11.08</td>
<td>-5.732</td>
<td>**</td>
<td>11.08</td>
<td>7.09</td>
<td>2.083</td>
<td></td>
<td>21.70</td>
<td>7.09</td>
<td>-6.015</td>
<td>***</td>
<td></td>
</tr>
<tr>
<td>MED_CUL_LE</td>
<td>76.63</td>
<td>68.24</td>
<td>-1.968</td>
<td></td>
<td>68.24</td>
<td>80.61</td>
<td>-2.757</td>
<td>**</td>
<td>76.63</td>
<td>80.62</td>
<td>-0.709</td>
<td></td>
<td></td>
</tr>
<tr>
<td>IJI</td>
<td>52.17</td>
<td>44.06</td>
<td>-4.059</td>
<td>***</td>
<td>44.06</td>
<td>51.95</td>
<td>-3.880</td>
<td>***</td>
<td>52.17</td>
<td>51.96</td>
<td>0.082</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PR</td>
<td>5.32</td>
<td>5.12</td>
<td>-1.801</td>
<td></td>
<td>5.12</td>
<td>5.67</td>
<td>-4.814</td>
<td>***</td>
<td>5.32</td>
<td>5.67</td>
<td>-2.451</td>
<td>**</td>
<td></td>
</tr>
<tr>
<td>SIDI</td>
<td>0.54</td>
<td>0.53</td>
<td>-1.311</td>
<td></td>
<td>0.53</td>
<td>0.50</td>
<td>1.668</td>
<td></td>
<td>0.54</td>
<td>0.51</td>
<td>2.335</td>
<td>*</td>
<td></td>
</tr>
<tr>
<td>PRPWRKOUT</td>
<td>0.62</td>
<td>0.73</td>
<td>6.002</td>
<td>***</td>
<td>0.73</td>
<td>0.75</td>
<td>-1.975</td>
<td></td>
<td>0.62</td>
<td>0.75</td>
<td>-6.067</td>
<td>***</td>
<td></td>
</tr>
<tr>
<td>R_RENT_OWN</td>
<td>2.67</td>
<td>0.62</td>
<td>-5.792</td>
<td>***</td>
<td>0.62</td>
<td>0.29</td>
<td>0.930</td>
<td></td>
<td>2.67</td>
<td>0.29</td>
<td>5.172</td>
<td>***</td>
<td></td>
</tr>
</tbody>
</table>

Notes: Z-score is the standardized difference between the means. Probability of equal means denoted by: * p < 0.10; ** p < 0.05; *** p < 0.01. P-values corrected for multiple testing using Bonferroni method (Dalgaard, 2002).
Fig. 2-1. Study area (Salt Lake Valley) showing geographic distribution of three neighborhood types. Neighborhoods are defined as the residential portion of a census block group (n = 542)
Fig. 2-2. Aerial photography showing neighborhoods representative of the three neighborhood types. The median year-built for residential homes in photo (a) is 1920, in photo (b) is 1960, and in photo (c) is 1996. Note all photos are depicted at the same map scale.
Fig. 2-3. Stripcharts for Density and Centrality metrics.
Fig. 2-4. Stripcharts for Accessibility and Neighborhood Mix metrics.
CHAPTER 3
DETERMINANTS OF URBAN TREE CANOPY IN RESIDENTIAL NEIGHBORHOODS: HOUSEHOLD CHARACTERISTICS, URBAN FORM, AND THE GEOPHYSICAL LANDSCAPE

Introduction

At the turn of the 21st century, more than half the world’s population lived in urban areas, and the number of people moving to urban areas is expected to increase over the next 25 years (UNU/IAS, 2003). Designing cities to be more livable is an important priority for this century, and urban vegetation is increasingly recognized as a key condition for human well-being in urbanized areas (Bolund and Hunhammar, 1999; Brown, 2008). Beyond aesthetic benefits, urban trees have been shown to reduce energy consumption (Huang et al., 1990) and control storm-water runoff (McPherson, 1992; Xiao et al., 1998). There is also evidence that urban trees improve air quality (Nowak et al., 2006) and aid in carbon sequestration (Nowak and Crane, 2002). It is not surprising that many cities in the U.S. and elsewhere are campaigning to cultivate more urban trees (Brown, 2008; McPherson et al., 2008). A better understanding of how human and environmental factors are related to urban forests will provide planners with information to improve residential neighborhood design, and will guide foresters in tree planting campaigns aimed at encouraging the cultivation of urban trees.

Factors Influencing Urban Vegetation

Urban ecosystems consist of multiple interlinked social, economic, institutional, ecological, and physical sub-systems (Pickett et al., 1997; Alberti et al., 2003; Grimm and
Redman, 2004; Alberti, 2008). Within the urban ecosystem, patterns of human activities influence patterns and processes of biotic systems (Grimm et al., 2000), and ecological processes and patterns are influenced by human activity (Whitney and Adams, 1980). Within the last decade, considerable attention has been given to the relationship between the socio-economic status of neighborhoods and urban vegetation (Iverson and Cook, 2000; Hope et al., 2003; Martin et al., 2004; Grove et al., 2006b; Troy et al., 2007). A common finding is that income and level of education are positively correlated with a greater abundance (Iverson and Cook, 2000; Grove et al., 2006a; Troy et al., 2007) and a greater diversity of urban vegetation (Hope et al., 2003; Martin et al., 2004). Several complementary social theories have been posited to explain why socio-economic status is highly correlated with urban vegetation patterns.

Grove et al. (2006b) suggest that homeowners are likely to maintain landscapes similar to their neighbors’ because of peer influences and the social status associated with lifestyle. In this case, income is an important component of lifestyle, but not the only driver of residential vegetation pattern (Grove et al., 2006b). Social stratification theory suggests that neighborhoods are able to influence public and private investments at a municipal level based on their position within the social hierarchy of a community. The implication is that social power influences patterns of urban vegetation. Yet another theory suggests that differentiation in urban vegetation patterns is primarily explained by the simple availability of financial resources (Grove, 1996). Hope et al. (2003) and Martin et al. (2004) call this the “luxury effect.” Neighborhoods with the wherewithal to maintain diverse or abundant vegetation do so primarily because they have the economic means.
Cultural factors are also believed to provide an explanation for heterogeneous patterns of urban vegetation. Choices about where to live or how residential vegetation is maintained may be linked to cultural values associated with race and/or ethnicity. Based on the observation that Oleander (*Nerium oleander*) is a popular landscape plant among Hispanics/Latinos, Martin et al. (2004) hypothesized that the abundance of Oleander might be correlated with the percentage of a neighborhood population that is Hispanic/Latino in Phoenix, AZ. They found however that a statistically significant relationship could not be established. Troy et al. (2007) investigated the relationship between the percent African-American families in a neighborhood and urban tree canopy cover, and found a positive relationship exists in Baltimore, MD. Rather than attribute this to a cultural affinity for denser tree canopy, they suggest the relationship is most likely explained by historical legacy (e.g. past tree plantings). Whether or not there is a relationship between cultural values and observed heterogeneity of urban vegetation is a matter that is not clearly resolved.

A factor that has received little attention, but may be related to urban vegetation heterogeneity is family life-stage. Zimmerman (1984) notes that the family life-cycle is a useful concept commonly used in social ecology to better understand demands and uses of natural resources by different social groups. She shows for example, that demand for air conditioning, and number of vehicles used per family fluctuates during the family life-cycle—mid stage families demand more air conditioning and have more vehicles than early or late-stage families (Zimmerman, 1984). The possible relationship between family life stage and urban vegetation heterogeneity has been noted, but not explicitly tested in previous studies (Grove et al., 2006b; Troy et al., 2007). One hypothesis may be that as
people pass through different stages of the family life cycle, the value placed on
amenities such as urban vegetation changes.

Much of the social science research to-date on urban vegetation has focused on
the relationship between urban vegetation (abundance and diversity) and the social,
economic, and demographic characteristics of households within neighborhoods (Martin
et al., 2004; Grove et al., 2006b; Troy et al., 2007). Little research has addressed the role
of the built environment—the spatial structure of urban areas and how it relates to urban
vegetation. Conway (2009) and Conway and Hackworth (2007) studied the relationship
between residential vegetation and urban design and found that New Urbanist design
principles (e.g. compact, pedestrian friendly neighborhoods, and mixed land uses
(Leccese et al., 2000)) did not necessarily support more vegetation than neighborhoods
designed under conventional planning patterns. Whitford et al. (2001) found that
indicators of ecological performance, such as surface temperature, carbon storage and
sequestration, and biodiversity were lower in compact cities than those with low-density
development (i.e. sprawl) due to the lack of green space, particularly urban trees. Both of
these studies raise questions about the merits of urban design principles focused primarily
on structural design without regard for its impact, or relationship, with urban vegetation.

While vegetation patterns in human-dominated ecosystems are highly influenced
by a variety of human factors, the physical environment is also important. Geomorphic
gradients, particularly in arid and semi-arid environments, influence plant diversity and
abundance by controlling water and nutrient supply (Parker and Bendix, 1996; Wondzell
et al., 1996). In their study of urban plant diversity in Phoenix, AZ, Hope et al. (2003)
found that elevation was a significant predictor of variation in plant diversity. However
they noted that elevation was also highly correlated with financial wealth, and conclude that diversity in urban vegetation has less to do with limiting natural resources than human factors that control the availability of those resources (Hope et al., 2003).

Previous research (Hope et al., 2003; Martin et al., 2004; Grove et al., 2006b; Troy et al., 2007) has shown that a key factor influencing the abundance and diversity of vegetation in residential urban and suburban neighborhoods is the amount of time since the land was developed. Grove et al. (2006) and Troy et al. (2007) found that in Baltimore, MD, the abundance of neighborhood tree canopy cover increases with median house age to a point at about 45 or 50 years, whereupon tree canopy decreases as neighborhoods get older. Hope et al. (2003) and Martin et al. (2004) found an opposite effect in Phoenix, AZ. They looked at urban vegetation in neighborhoods up to 50 years old, and found that both abundance and diversity decreased with neighborhood age—newer neighborhoods had higher plant abundance and diversity. Hope et al. (2003) noted that newer neighborhoods tended to be wealthier, suggesting that neighborhood wealth may be a covariate linked to neighborhood age that explains variations in urban vegetation. Another explanation for these observations in Baltimore and Phoenix could be the stark environmental differences between the two cities, and the types of landscaping common in each.

Study Area

Salt Lake County lies on the eastern rim of the Great Basin ecoregion of the western United States (41° N, 111° W) at an elevation of 1,280 meters (4,200 feet) above sea level. The region is considered a temperate desert, with a mean annual precipitation
of 380-760 mm (15-20 inches) and mean potential evaporation of 910-990 mm (36-39 inches) (Banner et al., 2009). At the center of the county is the Salt Lake Valley. When Mormon settlers first arrived in 1847 they encountered a valley dominated by sagebrush (*Artemisia tridentata* spp.) and a variety of bunchgrasses (Whitney, 1892). No trees were present, though gamble oak (*Quercus gambelii*) and two species of maple (*Acer glabrum* and *Acer gradidentatum*) could be found on the surrounding mountainsides (Whitney, 1892). Within the first year of their arrival the pioneers began developing canals to divert mountain stream water from the Wasatch Range (east side of the valley) and Jordan River (valley bottom) to irrigate crops (Arrington, 1958). An extensive irrigation system soon provided water for both agriculture and residential landscaping that is still in use today.

One million people live in Salt Lake County (U.S. Census Bureau, 2009) and the population is expected to rise to 1.6 million by 2050 (Salt Lake County, 2010). Nearly all of the county’s population resides in 15 cities within the valley, which has an areal extent of approximately 800 km² (308 mi²). The earliest residential structures were built at the northeast end of the valley near City Creek Canyon. As the valley population grew, residential development extended south primarily along the foothills of the Wasatch Range. The valley bottom remained in agricultural production through the 19th century until an era of suburban expansion started shortly after World War II, pushing residential development into agricultural lands in the central valley. Residential development continues today with growth to the south and western fringes of the valley. The spatial distribution of urban vegetation throughout the valley generally follows the course of residential expansion. In an arid environment that does not naturally support trees, human
dominance on the landscape appears to be the primary factor driving the growth and expansion of urban forests.

**Study Objectives**

This study aims to better understand how a broad range of factors relate to residential urban tree canopy density in Salt Lake County, Utah. Previous studies have focused primarily on how neighborhood socio-economic status relates to urban vegetation cover. Our goal is to broaden the scope of analysis to include an assessment of the relationship between the physical environment and residential tree canopy variability. We consider two aspects of the physical environment; namely the built environment, or aspects of urban design and form, and the physical geography of neighborhoods, which includes topographic gradients, precipitation, and soil characteristics. We consider neighborhood age (time since development) a confounding, or moderating, factor to be controlled. In other words, our question is if we hold neighborhood age constant, what is the relationship between human and physical dimensions of neighborhoods and residential tree canopy abundance? We frame this question within the context of three theories explaining variability in urban tree canopy abundance for residential neighborhoods.

1) *Variation in residential tree canopy is explained by social, economic and demographic characteristics of neighborhood households.* Building on the work of others (Grove et al., 2006b; Troy et al., 2007) we test the hypothesis that income and social status are positively related to the abundance of urban tree canopy. In addition we explore the hypotheses that racial/ethnic background of
households and family life-stage are related to the abundance of neighborhood tree canopy. A better understanding of how these household characteristics of neighborhoods relate to urban tree canopy will help urban foresters target certain social and/or demographic groups in tree planting and maintenance efforts.

2) *Variation in residential tree canopy is explained by the spatial structure (urban form) of residential neighborhoods.* Previous studies (Whitford et al., 2001; Conway and Hackworth, 2007; Conway, 2009) suggest that the spatial form of residential development is correlated with the abundance of urban vegetation. Using spatial metrics of urban form (e.g. median lot size, street connectivity, land use heterogeneity, etc.) we test the hypothesis that the spatial structure of residential neighborhoods is related to tree canopy abundance. Addressing this question is particularly relevant in light of current planning efforts that encourage specific types of urban form as a measure to curb urban sprawl.

3) *Variation in residential tree canopy is explained by gradients in physical geography.* Given the well-recognized influence of the physical environment on plant growth and production (Parker and Bendix, 1996; Wondzell et al., 1996) we hypothesize that urban vegetation, measured as tree canopy cover, will be correlated to these gradients in an urban setting.

As a final objective we evaluate the relative contribution of human and physical factors to observed heterogeneity in urban tree canopy cover. We address the question does any one theory have a stronger relationship with variation in tree canopy abundance than the others? Or can observed heterogeneity in tree canopy abundance be explained by a combination of all three theories?
Methods

Data

A common spatial unit of observation for neighborhood level studies in the U.S. is the census block group (Grove et al., 2006a; Grove et al., 2006b; Troy et al., 2007). Census block groups are considered fairly homogeneous areas with similar social, economic, and demographic characteristics, and with populations of typically around 1500 people (Peters and MacDonald, 2004). Block group boundaries however may include non-residential land uses. Given that our focus is on residential vegetation, we modified the census block group boundary to include only the residential portion of the block group. We did this by selecting residential parcels from the county parcels GIS dataset, dissolving them to create residential land use polygons, buffering them by 10 meters, and intersecting those polygons with the census block group dataset. This created a GIS dataset of 542 “neighborhoods” defined as the residential portion of census block groups.

We generated a GIS dataset of urban tree canopy from color-infra red digital orthophotography (NAIP, 2006) for Salt Lake County using an object-oriented image segmentation approach (Jensen, 2005). This dataset provided complete coverage of urban tree canopy as polygons for the study area, from imagery with approximately 1 meter spatial resolution. To assign a measure of tree canopy to each residential neighborhood, we intersected the tree canopy GIS dataset with the neighborhood GIS dataset and calculated the percent (i.e. proportion) of the neighborhood covered by canopy (from a
plan view). Fig. 3-1 shows the study area with 542 neighborhoods and percent tree canopy cover for each neighborhood depicted in graduated gray scale.

To investigate the relationship of various aspects of household characteristics, urban form, and the geophysical landscape to neighborhood tree canopy, we assembled a GIS database from a variety of sources (Table 3-1). The data were prepared within the GIS so that each observation unit (i.e. neighborhood) was populated with information for each of the database variables. Data on median household income and the percent of high school graduates have been used in previous studies (Grove et al., 2006b; Troy et al., 2007) as a measure socio-economic status. To measure race/ethnicity we calculated the percent of non-white persons using available census data. In Utah, Hispanics are the largest minority population group (U.S. Census Bureau, 2009). Average household size and median population age, are intended to measure family life-stage, with the assumption that neighborhoods with larger household sizes and mid to lower population age indicate neighborhoods predominated by middle-stage families in the family life cycle.

Measurements of urban form come largely from urban sprawl literature (Galster et al., 2001; Ewing et al., 2002; Song and Knapp, 2004). Median lot size and street density both measure the density of the built environment (Ewing et al., 2002; Song and Knapp, 2004). Median block perimeter may be thought of as a measure of urban density, but also measures the spatial structure of neighborhood design—suburban era neighborhoods typically have larger block perimeters because of a less-connected street network (Song and Knapp, 2004). A common method to measure street connectivity is the ratio of streets to intersections (Weston, 2002; Song and Knapp, 2004). Well-
connected neighborhoods have a high ratio of streets to intersections; a poorly connected design—with many cul-de-sacs—have a low ratio. Another common measure of urban form is land use mix, or land use heterogeneity (Galster et al., 2001). Sprawling urban form, typical of the suburban era, is commonly believed to be more homogeneous in both social/demographic characteristics, and land uses (Galster et al., 2001). For this metric we measure interspersion and juxtaposition (IJI) of commercial land uses within and around residential neighborhoods using a spatial index borrowed from landscape ecology (McGarigal et al., 2002).

A predominant limiting factor to plant growth in arid and semi-arid environments is water availability. Environmental gradients such as elevation, slope, and aspect, in addition to the water holding capacity of soils, are important factors influencing water as a resource to plants (Parker and Bendix, 1996; Wondzell et al., 1996). We selected five variables, readily available in spatial format, to measure these environmental gradients. We found slope and elevation to be highly correlated with mean annual precipitation, so chose only mean annual precipitation as a measure of water (non-irrigated) availability. Because both streams and canals tend to provide ground water to surrounding areas, we included a measure of the distance to these water bodies—measured as the neighborhood’s mean distance to the stream or canal. Available water storage capacity was obtained from a soils GIS dataset and is measured as the volume of water the soil is capable of storing that is available to plants (NRCS, 2009). Because measurements of aspect using conventional azimuths are problematic (north measured as both 0 and 360 degrees), we created two transformed metrics, one measuring westness and the other southness (Chang and Li, 2000).
A critical component of the GIS database is a variable measuring neighborhood age, or time since the land was developed. Neighborhood age was measured as the median age of residential homes in the neighborhood, based on the year the structure was built.

**Statistical Analysis**

For the sake of clarity we divide the discussion of our statistical analysis into three parts. Fundamentally our analytical approach utilizes linear regression to assess the relationship between urban tree canopy and variation in the household characteristics, urban form, and physical geography of residential neighborhoods. We begin with a brief discussion of regression diagnostics, followed by a description of our approach to address the question of neighborhood age as a moderating factor in the regression model. We conclude by describing our approach for determining the relative influence of the three theories (household characteristics, urban form, geophysical landscape) as explanations for vegetation abundance in residential neighborhoods.

**Regression Diagnostics**

Linear regression carries with it a number of assumptions that must be met in order to confidently interpret model results (Faraway, 2005). As a parametric model, linear regression assumes normally distributed data and often percentage data are not normally distributed. Deviation from normality, however, tends to be more pronounced at small and large percentages, approximating normality when data range between 30 and 70% (Zar, 1999). We tested for normality in the dependent variable (percent tree canopy) using Q-Q plots of model residuals, and residuals vs. fitted values plots and found
normality not to be a problem with our data (Appendix B, Fig. B-1). We tested for constant variance of residuals and non-linearity between the dependent and independent variables using residuals vs. fitted values plots (Appendix B, Fig. B-1) and partial residuals plots (Appendix B, Fig. B-2). We did not find enough of a problem with non-constant variance or non-linearity to warrant transformation of any of the independent variables. Using leverage plots (Appendix B, Fig. B-1) and added-variable plots (Appendix B, Fig. B-3) we checked for outliers and leverage points, finding 21 unusual or influential observations among the 542 samples. These were removed from further analysis. To test for collinearity among the independent variables we computed the variance inflation factor (VIF) for each independent variable, and examined a Pearson’s moment correlation matrix for large pairwise correlations. With the exception of one pair of variables (percent non-white population and percent high school graduates) correlation among independent variables was below 0.80 (Appendix B, Table B-1). We used the Moran’s I test for spatial autocorrelation and found high positive spatial dependence among model residuals (Appendix B, Table B-2). To account for spatial autocorrelation we adopted a simultaneous autoregressive model (SAR) for the regression analysis using spatial weights defined by the nearest neighborhood centroid (Bivand et al., 2008).

Interpreting Interaction Terms

Evidence (Martin et al., 2004; Grove et al., 2006b; Troy et al., 2007) suggests that the age of a neighborhood is an important covariate explaining tree canopy abundance in urban residential neighborhoods. The more time that has passed since a neighborhood was developed, the more time there has been for trees to grow. If we wish to test the
hypothesis that household income, for example, is positively related to tree canopy abundance, we must recognize the confounding effect of neighborhood age. One way to do this is to sample neighborhoods varying in household income, but which were built at the same time. For practical reasons this is not possible, as the number of samples for any given time would be unreasonably small. It is possible, however, to account for the moderating effects of a covariate in multiple linear regression using interaction terms (Aiken and West, 1991; Jaccard and Turrisi, 2003). For example, in the equation (following Aiken and West (1991) we place the intercept \( b_0 \) term in the last position):

\[
\hat{Y} = b_1X + b_2Z + b_3XZ + b_0
\]

The \( XZ \) interaction term signifies the regression of \( Y \) on \( X \) is dependent on specific values of \( Z \). So if \( Y \) is percent tree canopy, \( X \) is household income, and \( Z \) is neighborhood age, the regression equation above estimates the slope of canopy cover (\( Y \)) on household income (\( X \)) along a continuous range of neighborhood ages (\( Z \)). To determine the slope of \( Y \) on \( X \) for any value \( Z \), we restructure the regression equation through simple algebra,

\[
\hat{Y} = (b_1 + b_3Z)X + (b_2Z + b_0)
\]

and substitute the desired value \( Z \) in the equation. Aiken and West (1991) refer to \( (b_1 + b_3Z) \) as the simple slope of the regression of \( Y \) on \( X \) at the single value of \( Z \). To illustrate how the simple slope is computed it is useful to present a numerical example. Given a regression model where the following parameters are estimated:
\[ b_0 = 13.98 \]
\[ b_1 = 2.60 \]
\[ b_2 = 0.89 \]
\[ b_3 = 0.29 \]

The regression equation,

\[ \hat{Y} = 2.60X + 0.89Z + 0.29XZ + 13.98 \]

is restructured:

\[ \hat{Y} = (2.60 + 0.29Z)X + (0.89Z + 13.98) \]

To find the slope and intercept at neighborhood age 25 we substitute 25 for \( Z \):

\[ \hat{Y} = [2.60 + 0.29(25)]X + [0.89(25) + 13.98] \]

\[ \hat{Y} = 9.85X + 36.23 \]

Using this approach we can graph simple slopes for different values of \( Z \) to visualize the relationship of \( Y \) to \( X \) conditional on values of \( Z \). It is also possible to test whether the slopes are statistically different from zero, and to determine whether the
simple slopes are statistically different from one another (an interaction effect). In terms of our analysis this allows us to determine whether there is a statistically significant relationship between percent tree canopy and household income, for example, at any given neighborhood age, and to determine whether that relationship differs between neighborhood ages.

*Model Selection and Multi-Model Inference*

Our second objective is to explore the question of which theory best explains variation in the abundance of tree canopy in residential neighborhoods. We follow the information-theoretic approach of Burnham and Anderson (2002) to evaluate and compare a set of candidate models. This approach is both philosophical and procedural in nature. Philosophically it asserts that the goal of scientific data analysis is to make inferences from models by separating out the information in the data from the noise (Burnham and Anderson, 2002). Consistent with the philosophy of “multiple working hypotheses” (Chamberlin, 1965) the approach begins with thoughtful selection of a set of plausible candidate models based on sound theoretical reasoning. All candidate models are considered reasonable approximations of the issue at interest. To evaluate the relative “correctness” of the models Burnham and Anderson (2002) adopt the Akaike information criterion (AIC). As an information-theoretic criterion AIC operates under the principle of parsimony—it seeks to measure how well the model fits the data while penalizing unnecessary complexity. Alone the AIC score is simply a measure of model fit taking into consideration model complexity. As a relative measure, it can be used to compare
alternative models with different parameters as long as the set of observations remains
the same (Burnham and Anderson, 2002)

To assess the relative contribution of the three theories (household characteristics,
urban form, and geophysical landscape) as explanations of variation in residential
vegetation we used AIC to compare eight models: a model for each of the three theories
with interaction terms, and without interaction terms; and two full models (all three
theories combined) with and without interaction terms.

Results

Fig. 3-2 presents plots of calculated simple slopes for the 15 explanatory variables
tested. Each plot represents a regression model of percent tree canopy ($Y$) on one
explanatory variable ($X$) with the inclusion of the covariates neighborhood age ($Z$) and
the two-way interaction of neighborhood age with the $X$ variable ($XZ$). For each
explanatory variable we calculated simple slopes for three neighborhood ages: 15 years,
55 years, and 95 years old. Table 3-2 provides a numerical account of the plots in Fig. 3-2
and is useful to illustrate which simple slopes are statistically different from zero.

Simple slopes that are zero are interpreted to mean there is no relationship
between percent tree canopy and a given explanatory variable for that neighborhood age.
For example, in the median income model we find that the simple slope for
neighborhoods that are 15 years old is not significantly different from zero, suggesting
that there is no relationship between household income and percent tree canopy in newer
neighborhoods. For older neighborhoods a significant positive relationship between
income and percent tree canopy emerges, and the relationship becomes more pronounced in 95 year old neighborhoods.

The plots of simple slopes in Fig. 3-2 also inform us about the interaction between neighborhood age and the explanatory variable. Simple slope lines for different neighborhood ages that run parallel to one another indicate there is no interaction effect between neighborhood age and that variable. Thus the mean annual precipitation model suggests there is a positive relationship between annual precipitation and tree canopy, and that the relationship remains the same regardless of neighborhood age. The median income model suggests a strong positive interaction (i.e. strengthening) effect with neighborhood age. The effects of income combined with neighborhood age increase the likelihood of greater urban tree canopy. An example of a negative interaction (i.e. weakening) effect is presented in the land use mix model, which suggests that the combined effect of greater land use heterogeneity and neighborhood age is correlated with lower tree canopy.

Table 3-3 presents the results of an AIC information-theoretic approach to model evaluation for the three theories explaining urban tree canopy variation across different types of residential neighborhoods. The eight candidate models are ranked from the best model to the poorest model according to AIC score. To evaluate the set of candidate models, AIC is scaled so that the best model has a value of 0 and all other models are ranked relative to the best model ($\Delta_i AIC$). Burnham and Anderson (2002) suggest that relative differences between 1 and 2 AIC scores of the best model should be considered as good as the best model. Models between 4 and 7 AIC scores larger than the best model are moderately good relative to the best model, and models with AIC scores larger than
10 AIC scores are considerably less good in terms of fit and complexity than the best overall model. The weighted AIC ($w_i$) score gives the probability that any candidate model is as good as the best model (Burnham and Anderson, 2002).

The results indicate that taking into consideration model fit and accounting for model complexity, the best model is the full model with neighborhood age interactions. The next best model (full model without interactions) is 23.57 AIC scores higher, and by Burnham and Anderson’s recommendation not a suitable alternative to the best model. According to this analysis all other candidate models do considerably poorer at explaining the relationship between urban tree canopy and their respective model variables. Because the probability that the alternative models are as good as the best model is zero, averaging the models to make inferences on the entire set of models is not advantageous (Burnham and Anderson, 2002). Instead, our analysis suggests that the level to which each model differs from other candidate models is quite pronounced—none of the models rank equally with each another.

Following Burnham and Anderson’s (2002) AIC multi-model inference approach it is possible to assess the relative importance of individual predictor variables. This is done by fitting separate models for all possible combinations of the model variables and summing the AIC weights ($w_i$) for all models in which the variable occurs. We did this for the 16 explanatory variables, producing $2^{16} - 1 = 65,536$ models. The results presented in Table 3-4 identify the probability that each variable is the most important variable among the set of variables analyzed. The results suggest that neighborhood age, and aspect (westness) are the most important model variables, and that available water storage (soils) and median household income are essentially just as important. The most
important urban form variable is *street connectivity* with *land use mix* coming in a close second. The variable least likely to be important in understanding the determinants of urban tree canopy variation is *mean annual precipitation*.

**Discussion**

This analysis makes at least two important contributions to the growing body of theory aimed at understanding the social and physical determinants of urban tree canopy in residential neighborhoods. First, we have attempted to disentangle the role of neighborhood age from factors (i.e. household characteristics, urban form, geophysical landscape) determining urban forest structure to better understand how they relate to tree canopy variation in residential neighborhoods. Second, we have illuminated the combined influence of multiple determinants of urban forest canopy and have attempted to quantify their relative importance in a semi-arid urban environment.

The importance of *time since development* is revealed through both our analysis of the interaction effects of neighborhood age with other explanatory variables, and through the AIC multi-model analysis. We demonstrate empirically what is often logically assumed by social theory—that while social, economic, and demographic factors are related to urban vegetation abundance, the relationship is moderated by the amount of time the vegetation has had to grow. The “luxury effect” (Hope et al., 2003; Martin et al., 2004) or the positive relationship between household income and vegetation abundance is influenced by the age of the neighborhood, and in fact, we see that in relatively new neighborhoods there is no relationship. However, as time progresses the luxury effect becomes more pronounced. Time strengthens (multiplies) the effect of
income on neighborhood canopy abundance because wealthy homeowners have the financial resources to invest in growing vegetation.

Troy et al. (2007) observed a positive relationship between variables measuring family life stage (marriage rates and household size) and tree canopy in Baltimore, and hypothesized that families with more children either plant and maintain more trees or self-select by moving to neighborhoods with more trees. In Salt Lake County we hypothesized a negative relationship exists between mid-stage families with children and tree canopy, not because families in Salt Lake County are adverse to trees, but because urban trees are a rare commodity in this environment, and the opportunity cost of either planting and maintaining trees or moving to an established neighborhood with mature trees is high. We find that our analysis supports this hypothesis only in new neighborhoods. At the 15-year-old level, tree canopy decreases with household size, and increases with median population age. In Utah it is common for mid-stage families to build a new home where they invest most of their wealth in the house and only gradually invest in landscaping. This is less likely to occur in new neighborhoods with older or younger populations where landscaping expenses might be covered by a homeowners association or by the developer. In our study there appears to be no relationship between family life-stage and tree canopy in older neighborhoods. We interpret this to mean that mid-stage families with children are not self-selecting for neighborhoods with either more or less tree canopy.

When we look at the relationship of race/ethnicity to urban tree canopy our analysis indicates that taking neighborhood age into consideration makes a difference. Troy et al. (2007) note a positive relationship between percent African-American
population and urban vegetation in Baltimore. Our analysis suggests that in new neighborhoods there is no relationship between race/ethnicity and tree canopy abundance, but in older neighborhoods with higher percentages of non-white populations, tree canopy abundance is likely to be lower. The immediate explanation might be that older neighborhoods with higher percentages of minorities are less likely to have the financial resources (luxury effect) or social capital (social stratification) required to replace trees that die after their natural life span is complete. Whether income, social capital, transiency, racial or ethnic values, or some other factor explains this relationship is not clear, and suggests that additional studies of race/ethnicity and urban vegetation are needed.

Our study of how household characteristics are related to urban tree canopy provides useful information to urban foresters interested in cultivating and maintaining urban forests. Many municipalities around the nation are sponsoring campaigns to increase urban tree canopy for the benefits it provides (Brown, 2008; McPherson et al., 2008). Knowing more about how urban tree canopy is related to different social groups provides practitioners with information about how to distribute scarce financial resources. For example, new neighborhoods dominated by mid-stage families will not have much tree canopy to start with, but will eventually increase the amount of tree canopy over the course of time. On the other hand, older neighborhoods with high percentages of non-white populations will likely not increase tree canopy without some kind of assistance. Given these two situations, a tree planting campaign would benefit most by directing its resources to the older neighborhood.
Treating neighborhood age as a moderating factor in our analysis of urban form and its relationship to tree canopy is informative and sheds light on some important questions currently debated in urban planning. The built structure of urban areas is known to change over time (Song and Knapp, 2004) so controlling for neighborhood age is particularly relevant to a better understanding of urban form as a determinant of urban tree canopy. In our analysis, the 95-year-old neighborhood level corresponds roughly to the period around 1910 and is representative of the pre-suburban era prior to the end of World War II. The 55-year-old level represents the early suburban era around 1950, and the 15-year-old level represents the late suburban era around 1990. It must be remembered that our analysis presents models based on patterns in the data, so the relationships can be thought of as empirical “trends” rather than actual observations (there are no pre-suburban neighborhoods with a connectivity index of 0.05, in fact, the mean connectivity index for the pre-suburban era is 2.4 (see chapter 2)). What this means is that by modeling an urban form metric, such as street connectivity conditional on a given time since development (neighborhood age), we are predicting what tree canopy would be like in neighborhoods with that level of connectivity after a given number of years have passed. Because many of the questions in urban planning theory and practice involve speculating about how a particular planning strategy or design will improve urban function in the future, this approach, which could be viewed as a “place for time” substitution, is particularly useful as a tool to better understand the potential outcomes of spatial planning ideas and efforts.

Over the last two decades American urban planning has been greatly influenced by the ideals espoused by the philosophy of New Urbanism. However, despite its allure,
New Urbanism, as any new planning philosophy or approach, is an experiment underway. New Urbanism has both advocates (Bressi, 2002) and critics (Grant, 2006). Conway (2009) and Conway and Hackworth (2007) are among the few who have begun to critically examine New Urbanism and the implications of this new planning approach to the ecological function of cities. Our study contributes to this discussion by specifically looking at urban form metrics commonly associated with New Urbanism ideals, and how they relate to the biotic environment in cities, namely urban tree canopy.

An important tenet of New Urbanism is the idea that residential neighborhoods should be walkable, which means residential blocks should be small, and street networks dense and well-connected. Critics argue, however, that making neighborhoods more connected for humans makes them less connected for ecological processes (wildlife movement, seed dispersal, etc.) and that a cul-de-sac design may be better for natural ecosystem connectivity (Grodon and Tamminga, 2002). If we consider residential tree canopy as one possible measure of natural ecosystem connectivity, an examination of the relationship between neighborhood design and tree canopy sheds some light on the question about how design influences natural urban ecosystems. What our analysis shows is that in new neighborhoods there is a positive relationship (statistically significant at $\alpha = 0.05$) between both street connectivity and density, and residential tree canopy. Neighborhoods with street designs that are dense and well-connected start off with greater tree canopy than less dense and less well-connected streets, but after about 50 years street design no longer makes a difference. So the question of street design for ecosystem connectivity may not be as important in the long term as it is in the first few decades the urban forest is being established.
The other urban form metrics we tested (land use mix, block perimeter, and median lot size) suggest that the amount of time that has passed influences the "performance" of the design principle measured by the metric. Although the relationship is not statistically significant at $\alpha = 0.05$, the trend in the slopes for the median lot size metric provides useful information. When we look at the 15-year time level the relationship is negative—larger lots have less tree canopy. However over the course of 50 years, the model suggests that tree canopy density is about the same for both small and large lots. After 95 years, it appears that large lots are capable of supporting more tree canopy and the relationship between lot size and tree canopy slopes in the positive direction. This highlights the importance of considering time in any assessment of how the design of residential neighborhoods influences urban vegetation.

The importance of geophysical landscape determinants becomes apparent through both our analysis of the simple slope models and through AIC multi-model analysis. Two factors that have the same relationship with urban tree canopy regardless of neighborhood age are mean annual precipitation and distance to streams/canals. As hypothesized the further a neighborhood is from a stream or canal the lower the tree canopy, which suggests groundwater flow from these streams/canals is influencing urban vegetation. Mean annual precipitation is positively related to tree canopy abundance regardless of neighborhood age, however the importance of the variable is highly questionable based on the AIC multi-model analysis. That precipitation is of little importance is not surprising, however, in an environment that receives little precipitation during the summer, and where nearly all homeowners rely on irrigation water for landscape maintenance. When we look at soil water storage capacity we see that an interaction
effect exists between neighborhood age and soils in determining tree canopy abundance. There is a strengthening effect between how old the neighborhood is and the type of soils in the neighborhood—older neighborhoods with good soils have higher tree canopy cover. This suggests that lands with the highest water storage capacity were the earliest to be developed.

Our analysis of topographical aspect presents an interesting case where variables chosen may not always measure what one initially expects, and demonstrates the uniqueness of the urban ecosystem relative to more natural ecosystems. Initially our choice of aspect as a determinant of urban tree canopy was to capture the effects of insolation and water availability. Upon examining the results we find that in new and moderately old neighborhoods there is a positive relationship between west-facing azimuths and tree canopy abundance, which runs contrary to our hypothesis that west-facing neighborhoods would have lower tree canopy cover. We have determined that aspect in this particular urban setting explains historic settlement patterns and homeowners preferences for view sheds, rather than limiting water resources due to insolation. We conclude that the geophysical landscape is indeed an important determinant of urban vegetation, but that ecological processes in the urban ecosystem are greatly influenced by the dominant role of humans.

**Conclusion**

Over 25 years ago Sanders (1984) wrote the paper “Some Determinants of Urban Forest Structure” in the journal *Urban Ecology*. One of Sanders’ objectives was to identify and examine factors that influence urban vegetation patterns in American cities.
His work was largely speculative, based primarily on observation and available knowledge. He recognized “there are presently no rapid, accurate, and systematic methods for collecting inventory data for an entire urban forest complex, a fact that precludes an entirely empirical approach to the subject of urban forest science” (Sanders, 1984). One of our objectives in this study has been to show that with advances in data collection and management technologies, inventories and the analysis of complete urban forest ecosystems are possible. Sanders (1984) suggested urban forest structure is determined by three broad factors: urban morphology, natural factors, and human management systems. He argued that urban morphology creates the spaces available for vegetation, natural factors such as precipitation and soils influence the types of biomass that are likely to thrive, and human management systems account for intra-urban heterogeneity based on variations in human choice.

Research addressing urban tree canopy in the last decade has certainly taken advantage of the data management and analysis technologies currently available (Grove et al., 2006a; Grove et al., 2006b; Conway and Hackworth, 2007; Troy et al., 2007; Conway, 2009) but little has been done to integrate all the determinants of urban tree canopy in a comprehensive analysis. We suggest that our study merely scratches the surface of what can be done given currently available data and technology, and serves more to illuminate this need than to satisfy it.

References


and embedded small parks of Phoenix, AZ. Landscape and Urban Planning 69, 355-368.


Nowak, D. J., Crane, D. E., 2002. Carbon storage and sequestration by urban trees in the USA. Environmental Pollution 116 (381-389), 381.


Whitney, O. F., 1892. History of Utah. Cannon & Sons, Salt Lake City, UT.


### Table 3-1. Determinants of urban tree canopy. Description of explanatory variables and data sources.

<table>
<thead>
<tr>
<th>Variable name</th>
<th>Abbrev.</th>
<th>Description &amp; Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>Neighborhood Age</td>
<td>NHAGE</td>
<td>Median age of single family residences. Source: SLCo Parcel GIS data.¹</td>
</tr>
<tr>
<td>Street Connectivity</td>
<td>STCON</td>
<td>Street connectivity index; ratio of links (streets) to nodes (intersections). Source: SLCo Street Centerline GIS data.</td>
</tr>
<tr>
<td>Land Use Mix</td>
<td>LUMIX</td>
<td>Land use mix, or contiguity. Measured by Interspersion &amp; Juxtaposition (IJ) index. Source: SLCo Parcel GIS data.</td>
</tr>
<tr>
<td>Median Lot Size</td>
<td>MDLOT</td>
<td>Median single family lot size (acres). Source: SLCo Parcel GIS data.</td>
</tr>
<tr>
<td>Res. Street Density</td>
<td>RSDEN</td>
<td>Residential street density. Line density function in ArcGIS. Source: SLCo Parcel GIS data.</td>
</tr>
<tr>
<td>Median Block Perimeter</td>
<td>MBLKP</td>
<td>Median length of neighborhood block perimeter (m.). Source SLCo Parcel GIS data.</td>
</tr>
<tr>
<td>Mean Annual Precip.</td>
<td>MPRCP</td>
<td>Mean annual precipitation (mm) average for neighborhood. Source: PRISM GIS data.³</td>
</tr>
<tr>
<td>Avail. Water Storage</td>
<td>AVWTS</td>
<td>Available water storage capacity (cm) of soils to the depth of 150 cm. Source: STATS.GO GIS data.⁴</td>
</tr>
<tr>
<td>Dist. to Streams/Canals</td>
<td>DSTSC</td>
<td>Average Euclidean distance to streams/canals for each neighborhood. Source: SLCo stream/canal GIS data.</td>
</tr>
<tr>
<td>Aspect (Westness)</td>
<td>ASWST</td>
<td>Transformed azimuth measuring degree to which topographic location is facing west (west = 180). Mean azimuth for neighborhood. Source: USGS 10 m DEM.⁵</td>
</tr>
<tr>
<td>Aspect (Southness)</td>
<td>ASSTH</td>
<td>Transformed azimuth measuring degree to which topographic location is facing south (south = 180). Mean azimuth for neighborhood. Source: USGS 10 m DEM.⁵</td>
</tr>
</tbody>
</table>

Table 3-2. Estimated simple slopes for 15 explanatory variables. Slope of $Y$ (percent tree canopy) on explanatory variable ($X$) at neighborhood age ($Z$).

<table>
<thead>
<tr>
<th>Independent Variable</th>
<th>15 years old</th>
<th>55 years old</th>
<th>95 years old</th>
<th>Interaction Effect $^\S$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Median Income</td>
<td>0.25</td>
<td>1.49</td>
<td>2.74</td>
<td>6.07 **</td>
</tr>
<tr>
<td>Ave. Household Size</td>
<td>-1.71</td>
<td>-0.17</td>
<td>1.37</td>
<td>1.00</td>
</tr>
<tr>
<td>Med. Population Age</td>
<td>0.32</td>
<td>0.17</td>
<td>0.02</td>
<td>0.13</td>
</tr>
<tr>
<td>Non-White Population</td>
<td>0.04</td>
<td>-0.07</td>
<td>-0.18</td>
<td>-4.74 **</td>
</tr>
<tr>
<td>High School Graduates</td>
<td>0.09</td>
<td>0.15</td>
<td>0.22</td>
<td>4.04 **</td>
</tr>
<tr>
<td>Street Connectivity</td>
<td>3.23</td>
<td>4.09</td>
<td>4.96</td>
<td>2.83 **</td>
</tr>
<tr>
<td>Land Use Mix</td>
<td>0.01</td>
<td>-0.07</td>
<td>-0.15</td>
<td>-4.22 **</td>
</tr>
<tr>
<td>Median Lot Size</td>
<td>-8.56</td>
<td>2.97</td>
<td>14.49</td>
<td>1.49</td>
</tr>
<tr>
<td>Res. Street Density</td>
<td>0.69</td>
<td>0.57</td>
<td>-0.37</td>
<td>-0.68</td>
</tr>
<tr>
<td>Median Block Perimeter</td>
<td>0.00</td>
<td>0.92</td>
<td>0.01</td>
<td>1.55</td>
</tr>
<tr>
<td>Mean Annual Precip.</td>
<td>0.04</td>
<td>0.04</td>
<td>0.04</td>
<td>1.80</td>
</tr>
<tr>
<td>Avail. Water Storage</td>
<td>-0.16</td>
<td>0.32</td>
<td>0.81</td>
<td>3.20 **</td>
</tr>
<tr>
<td>Dist. to Streams/Canals</td>
<td>-1.59</td>
<td>-2.28</td>
<td>-3.16</td>
<td>-1.44</td>
</tr>
<tr>
<td>Aspect (Westness)</td>
<td>0.09</td>
<td>0.05</td>
<td>0.01</td>
<td>0.33</td>
</tr>
<tr>
<td>Aspect (Southness)</td>
<td>-0.02</td>
<td>1.50</td>
<td>0.06</td>
<td>2.45 *</td>
</tr>
</tbody>
</table>

Notes: Probability that slope is not equal to zero denoted by * $p < 0.05$; ** $p < 0.01$. $^\S$ Interaction significant $p < 0.05$. 


Table 3-3. AIC information-theoretic multi-model analysis of eight candidate models from three theories explaining tree canopy abundance in residential neighborhoods.

<table>
<thead>
<tr>
<th>Model Name</th>
<th>Log-Likelihood</th>
<th>K</th>
<th>AIC</th>
<th>ΔAIC</th>
<th>AIC weights (w_i)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Full Model w/ Age Interactions</td>
<td>-1619.08</td>
<td>25</td>
<td>3288.17</td>
<td>0.00</td>
<td>1.00</td>
</tr>
<tr>
<td>Full Model</td>
<td>-1636.87</td>
<td>19</td>
<td>3311.73</td>
<td>23.57</td>
<td>0.00</td>
</tr>
<tr>
<td>Geophysical Landscape w/ Age Interactions</td>
<td>-1649.52</td>
<td>12</td>
<td>3323.05</td>
<td>34.88</td>
<td>0.00</td>
</tr>
<tr>
<td>Household Characteristics w/ Age Interactions</td>
<td>-1669.59</td>
<td>11</td>
<td>3361.18</td>
<td>73.01</td>
<td>0.00</td>
</tr>
<tr>
<td>Urban Form w/ Age Interactions</td>
<td>-1687.46</td>
<td>10</td>
<td>3394.93</td>
<td>106.76</td>
<td>0.00</td>
</tr>
<tr>
<td>Geophysical Landscape</td>
<td>-1736.63</td>
<td>8</td>
<td>3489.25</td>
<td>201.09</td>
<td>0.00</td>
</tr>
<tr>
<td>Urban Form</td>
<td>-1778.37</td>
<td>8</td>
<td>3572.73</td>
<td>284.57</td>
<td>0.00</td>
</tr>
<tr>
<td>Household Characteristics</td>
<td>-1788.93</td>
<td>8</td>
<td>3593.85</td>
<td>305.69</td>
<td>0.00</td>
</tr>
</tbody>
</table>

Notes: AIC = Akaike Information Criterion, K = number parameters estimated.
Table 3-4. Computed variable importance for 16 determinants of urban tree canopy in residential neighborhoods using the sum of AIC weights ($w_i$) from analysis of $2^{16} - 1 = 65,535$ models in which the explanatory variable appears.

<table>
<thead>
<tr>
<th>Variable</th>
<th>$\Sigma$AIC weights ($w_i$)</th>
<th>Theory</th>
</tr>
</thead>
<tbody>
<tr>
<td>Neighborhood Age</td>
<td>1.0000</td>
<td>N/A*</td>
</tr>
<tr>
<td>Aspect (westness)</td>
<td>1.0000</td>
<td>Geophysical Landscape</td>
</tr>
<tr>
<td>Available Water Storage (Soils)</td>
<td>0.9977</td>
<td>Geophysical Landscape</td>
</tr>
<tr>
<td>Median Household Income</td>
<td>0.9973</td>
<td>Household Characteristics</td>
</tr>
<tr>
<td>Street Connectivity</td>
<td>0.9497</td>
<td>Urban Form</td>
</tr>
<tr>
<td>Land Use Mix</td>
<td>0.9050</td>
<td>Urban Form</td>
</tr>
<tr>
<td>Percent Non-White Population</td>
<td>0.8998</td>
<td>Household Characteristics</td>
</tr>
<tr>
<td>Percent High School Graduates</td>
<td>0.8089</td>
<td>Household Characteristics</td>
</tr>
<tr>
<td>Distance to Streams/Canals</td>
<td>0.7606</td>
<td>Geophysical Landscape</td>
</tr>
<tr>
<td>Median Lot Size</td>
<td>0.6587</td>
<td>Urban Form</td>
</tr>
<tr>
<td>Aspect (southness)</td>
<td>0.6510</td>
<td>Geophysical Landscape</td>
</tr>
<tr>
<td>Residential Street Density</td>
<td>0.5657</td>
<td>Urban Form</td>
</tr>
<tr>
<td>Median Block Perimeter</td>
<td>0.4499</td>
<td>Urban Form</td>
</tr>
<tr>
<td>Average Household Size</td>
<td>0.3583</td>
<td>Household Characteristics</td>
</tr>
<tr>
<td>Median Population Age</td>
<td>0.3467</td>
<td>Household Characteristics</td>
</tr>
<tr>
<td>Mean Annual Precipitation</td>
<td>0.2856</td>
<td>Geophysical Landscape</td>
</tr>
</tbody>
</table>

Notes: * Neighborhood Age is not considered pertaining to any single theory
Fig. 3-1. Relative spatial distribution of urban tree canopy by neighborhood in Salt Lake County, Utah in 2006. Number of neighborhoods = 542.
Fig. 3-2. Plots of simple slopes for 15 explanatory variables. Slope of $Y$ (percent tree canopy) on explanatory variable ($X$) at neighborhood age ($Z$).
CHAPTER 4
PREDICTING URBAN FOREST GROWTH AND ITS IMPACT ON RESIDENTIAL LANDSCAPE WATER DEMAND IN A SEMIARID ENVIRONMENT

Introduction

The importance of trees in urban areas has been recognized for millennia (Miller, 1997). However in a rapidly urbanizing world potential environmental and economic benefits offered by urban forests are increasingly important factors influencing urban planning and policy (Heynen and Lindsey, 2003; Conway and Urbani, 2007). Potential benefits include reduced energy consumption (Huang et al., 1990; McPherson, 1990) stormwater control (McPherson, 1992; Xiao et al., 1998) habitat for increased biodiversity (Johnson, 1988) improved air quality (Nowak et al., 2006) and carbon storage and sequestration (Nowak and Crane, 2002). Based on growing evidence that urban trees provide numerous benefits, cities in the U.S. and around the world are mounting campaigns to cultivate more urban trees (Brown, 2008; McPherson et al., 2008). Few have questioned the benefits of the urban forest, and in benefit-cost analyses the costs associated with urban forests have focused primarily on maintenance, planting, removal, and damage caused by roots and limbs (Nowak and Dwyer, 2007). An important challenge however, associated with urbanization in the 21st century is an increasing demand for water (Gleick, 2000). Population growth and the specter of a changing climate make the question of urban water demand particularly critical (Vorosmarty et al., 2000). Urban landscaping in cities in the western U.S. can consume up to 50% of the municipal water budget (Vickers, 1991; Bishop and Hughes, 1993).
With campaigns promoting the cultivation of trees for the benefits they offer, a better understanding of how growing urban forests impact the demand on available irrigation water is critical to the design and management of sustainable cities.

**Urban Forest Growth**

A common measure of urban forest growth is the amount of leaf area provided by urban trees, and is typically measured as the percentage of an area that is covered (from a plan view) by tree canopy (Sanders, 1984). Within the last decade considerable research has focused on how urban forests grow across the urban landscape, and specifically how certain factors account for spatial differentiation in urban tree canopy cover. Household income for example has been shown to be positively correlated with higher tree canopy (Iverson and Cook, 2000; Grove et al., 2006b; Troy et al., 2007) and there is evidence that education (Heynen and Lindsey, 2003; Troy et al., 2007) and family life-stage (see Chapter 3) are also correlated to tree canopy abundance. While there is some evidence that topographic gradients such as aspect and elevation are correlated with urban tree canopy abundance, it is most likely that limitations of resources such as water and nutrients in an urban setting are controlled more by landscape management than by natural processes (Hope et al., 2003). Clearly the role of humans in urban areas accounts for much of the variation observed in urban tree canopy.

An additional factor influencing forest growth in urban areas that receives less attention—perhaps because of its obvious role—is the amount of time that has passed since urban land was developed. Chapter 3 demonstrated that of several social and physical landscape factors, the age of a residential neighborhood was the most influential
factor explaining tree canopy abundance. The importance of neighborhood age has been noted by others. In Phoenix AZ, Hope et al. (2003) and Martin et al. (2004) found that the abundance and diversity of urban vegetation decreases with neighborhood age, while in Baltimore MD, Grove et al. (2006) and Troy et al. (2007) found the abundance of neighborhood tree canopy increases with housing stock age to a point at about 45 or 50 years, whereupon it decreases as neighborhoods get older. Troy et al. (2007) suggest the parabolic nature (inverted U) of the urban tree canopy-to-neighborhood age relationship may reflect changes in past planting efforts. Maco and McPherson (2002) suggested that the inverted U relationship between tree canopy and time reflects a natural climax of a first generation urban forest stand. As mature trees die off and are not replaced, tree canopy abundance of the first generation stand reaches a maximum cover at around 50%, then decreases, stabilizing at around 30% cover. Maco and McPherson (2002) were primarily concerned with canopy of urban street trees, but it is reasonable to extend their explanation of urban forest cover in relation to time to residential forest stands.

Urban Water Demand

Forecasting urban water demand is a topic that has been extensively studied in civil and environmental engineering, econometrics, and other related fields (Baumann et al., 1998). The simplest models estimate current demand as a function of per capita water-use multiplied by total urban population. Future demand is forecast by extrapolating current demand to projected populations based on an accepted population growth rate (Baumann et al., 1998). More sophisticated models incorporating additional explanatory variables such as population density, land use, per capita income, housing
size, water costs, and climatic factors have also been developed (Jain et al., 2001; Arbues et al., 2003). Methodological approaches include empirically based models using a variety of regression techniques (Jain et al., 2001; Arbues et al., 2003) mechanistic models (Dziegielewski and Boland, 1989) and artificial neural networks (ANN) (Jain et al., 2001; Msiza et al., 2008).

Urban water requirements are typically divided into municipal and industrial water uses, with municipal water including both residential and commercial allotments (Baumann et al., 1998). Residential water consumption can be divided into indoor and outdoor uses but estimates of residential demand typically do not distinguish the two (Arbues et al., 2003) probably because water supply companies rarely account for these two uses separately (UDNR-DWR, 2009). Outdoor water uses vary dramatically depending on income and geography. Landscaping and swimming pools, for example, account for a large portion of residential outdoor water use in Phoenix, AZ (Wentz and Gober, 2007). Residential landscaping is one of the most promising segments of the urban water budget for conservation (Farag, 2003), though there have been few assessments of water demand for residential landscape uses at the city or municipal level (Farag, 2003; Endter-Wada et al., 2008).

**Water Needs for Residential Landscapes**

To assess water requirements for residential and commercial landscapes, plant and horticultural scientists have adopted methods originally developed in agronomy to quantify the amount of water lost through evapotranspiration by different plant types (Kjelgren et al., 2000). Reference evapotranspiration (ETo) is a hypothetical water-loss
rate adopted by the United Nations Food and Agriculture Organization (UNFAO) as a standard measure by which other crops can be compared (Allen et al., 1998), and is calculated using meteorological data, making it possible to determine water requirements for given plants in different climatic regions. Using this approach, the evapotranspiration rate of a plant is expressed as a proportion of ETo. The proportion of ETo required by a plant to thrive productively is referred to as the water-loss crop coefficient (Kc). Thus, the Kc for a given crop may be 0.60 meaning it requires 60% of ETo for maximum growth and production. Applying the concept of ETo and crop coefficients to landscape plants is relatively new, and with the exception of turf grasses (Kneebone et al., 1992) few coefficients have been derived for specific landscape plants. It is nevertheless widely used by plant scientists and landscape management professionals to assess water needs of urban/suburban vegetated landscapes (Pettenger and Shaw, 2007). Because coefficients for agricultural crops are derived to maximize crop growth and production, and the goal for landscapes is to identify the minimum amount of water needed to maintain acceptable appearance and function, a distinction is made between coefficients for crops and coefficients for landscape plants. In the case of landscape plants, the adjustment factor is often referred to as a “plant factor” (PF) rather than a crop coefficient (Pettenger and Shaw, 2007).

Study Area

Salt Lake County, Utah (41° N, 111° W) is located on the eastern rim of the Great Basin ecoregion of the western United States. The region is considered a temperate desert with a mean annual precipitation of 380-760 mm (15-20 inches) and mean annual
reference evapotranspiration of 910-990 mm (36-39 inches) (Banner et al., 2009).

Precipitation falls predominantly as snow during the winter months, with very little occurring during the growing season from May to September. Capturing spring runoff in reservoirs and aquifers is therefore critical to the regions’ water supply. The Salt Lake County boundary lies coincident with the Jordan River hydrological basin, which receives water from the Wasatch Range to the east, the Oquirrh Range to the west, and the headwaters of the Jordan River in the Utah Lake hydrological basin to the south. Drainage from the Jordan River Basin flows north into the Great Salt Lake. Water supply comes by way of surface, ground, and imported sources (UDNR-DWR, 2009).

Municipal and industrial water use and supply is managed by the Jordan Valley Water Conservancy District, which estimated a total of 288,653 acre-feet of reliable potable water supplies for public community systems in 2005 (UDNR-DWR, 2009). Agricultural uses are managed separately.

Salt Lake County is the most populous county in the state with an estimated population of nearly a million people in 2005 (U.S. Census Bureau, 2005) and a projected population of 1.6 million by 2050 (Salt Lake County, 2010). Within the last 40 years the county has witnessed substantial residential expansion, primarily in the south and west portions of the valley. Nearly all residential expansion has occurred at the expense of agricultural land. Thus, water requirements in the county have shifted markedly from primarily agricultural uses during most of the 20th century to municipal and industrial uses at present and into the foreseeable future (UDNR-DWR, 1997, 2009).
Study Objectives

The goal of this study is to better understand how a growing urban forest influences residential landscape water demand in a semiarid metropolitan environment. To this end we have two primary objectives. First, to estimate residential landscape water demands based on the spatial distribution of tree/shrub canopy, turf grass cover, and under-canopy turf grass cover for all residential landscapes in the Salt Lake County metropolitan area. Second, to predict the spatial distribution of future urban forest canopy across the same area and to estimate future residential landscape water demands based on the growing urban forest. An important question to be addressed by this study is whether a growing urban forest increases overall residential irrigation demand, decreases demand, or has no apparent effect.

Methods

Data

We generated an urban vegetation GIS dataset (Fig. 4-1) from 1-meter color-infra digital orthophotography flown for the entire county in 2006. Overall map accuracy for five landscape classes (bare ground, irrigated grass, tree/shrub, impervious surfaces, and water bodies) was 75% using 1,197 reference sites. Residential neighborhoods were defined as the residential portion of U.S. Census block groups, and identified by residential parcels from a county parcel GIS database. This database was also used to determine the median age of housing stock for each residential neighborhood (Figs. 4-1 and 4-2). To assign a measure of tree/shrub canopy to each residential neighborhood we intersected the urban vegetation dataset with the neighborhood dataset and calculated the
proportion (later converted to percent) of the neighborhood covered by tree/shrub canopy (Fig. 4-1). We also did this for irrigated grass; producing a GIS dataset with 542 neighborhoods, each neighborhood containing an attribute for neighborhood age, percent tree/shrub canopy, and percent irrigated grass cover (not under-canopy).

Relative evapotranspiration (ETo) and precipitation data were obtained from the Salt Lake International Airport (4,225 ft (1287 m)) and Cottonwood Weir (4,960 ft (1,511 m)) weather stations (Utah Climate Center, 2010). Located in the northwest portion of the valley, the Salt Lake International Airport station is representative of the valley bottom and the drier western portion of the study area, while the Cottonwood Weir station, near the mouth of Big Cottonwood Canyon, is representative of the foothills on the east side of the valley. Since we considered these stations representative of the extreme conditions of study area as a whole, we averaged ETo and precipitation from the two stations for our analysis. Table 4-1 presents climate data used in our study.

**Estimating Water Demand**

To estimate water demand for residential landscapes we used a modification of the approach used by (Endter-Wada et al., 2008) and pioneered by the Irvine Ranch Water District in Southern California to establish water rate structures (Wong, 1999). Equation 1 estimates the amount of irrigation water required to replace water used by the vegetated landscape:

\[
I_m = \sum_{i=1}^{y} \left( \left( P_{F_i} \cdot ETo_m \right) - R_m \right) \cdot \left( A_L / D_U \right) \quad [1]
\]
where:

\begin{align*}
I_m &= \text{total irrigation demand for month } m \\
PF_L &= \text{plant factor or “crop coefficient” for landscape type } L \\
ETO_m &= \text{total reference evapotranspiration for month } m \\
R_m &= \text{total rainfall (precipitation) for month } m \\
A_L &= \text{area of landscape type } L \\
DU &= \text{distribution uniformity (irrigation inefficiency) factor}
\end{align*}

Measurements for landscape area ($A_L$) came from the urban vegetation GIS dataset, and $ETO_m$ and precipitation ($R_m$) from the weather station data. The choice of parameters for $PF_L$ and $DU$ was guided by the work by others. A reasonable $PF$ for warm season turf grass ranges between 0.80 and 0.90 (Kneebone et al., 1992). Few studies have reported adjustment factors for trees and woody plants partly because of the wide variety of species found in urban landscapes (Kjelgren et al., 2000). Reasonable factors have been reported from 0.20 to 0.80 (Costello and Jones, 1994). In the absence of clear standards for Plant Factors for urban vegetation, we followed the guidance of Farag (2003) and Enter-Wada et al. (2008) who successfully estimated residential landscape water demand in Layton, Utah using GIS data. For exposed irrigated grass we used a $PF$ of 0.80, for tree/shrub canopy we used 0.50, and following Farag (2003) we used a $PF$ of 0.40 for irrigated grass under tree canopy. Distribution Uniformity ($DU$) is an inefficiency factor that takes into consideration the non-uniformity with which most urban irrigation systems apply water to urban landscapes. Kjelgren et al. (2000) suggest
that $DUs$ ranging from 0.60 are common 0.70, and Farag (2003) used a $DU$ of 0.85 in Layton. We used a $DU$ factor of 0.75, which assumes that the irrigation systems for the residential neighborhoods in our study are on average 75% efficient.

Using Equation 1 we estimated residential landscape water demand for each neighborhood. For irrigated grass and tree/shrub canopy we used the measured area provided by the GIS dataset, for grass under tree canopy we used 60% of the area that was tree/shrub canopy. While the date of the imagery from which the urban vegetation GIS dataset was generated was 2006, we estimated water demand for 2005 (using 2005 ETo and precipitation) because a reliable report on municipal water use was available for 2005 and was not available for 2006. This allowed us to compare our estimates to reported water use data and a means to validate our modeling approach. We considered it reasonable to assume that vegetation cover did not change significantly between 2005 and 2006. It should be noted that this approach used the measured data from the urban vegetation GIS dataset. Since we have measured landscape data for only one year, our next step was to develop models to predict future quantities of the three landscape types used by the water demand model (i.e. Equation 1).

Relationship of Tree Canopy and Grass Cover to Neighborhood Age

The plots in Fig. 4-3 were created from the GIS dataset of 542 neighborhoods containing attributes for neighborhood age, percent tree canopy, and grass cover. As observed by others (Maco and McPherson, 2002; Grove et al., 2006b; Troy et al., 2007) the relationship between neighborhood age and percent canopy cover is curvilinear and appears to reach an asymptote of 50% tree canopy at about 100 years. The grass cover-to-
neighborhood age relationship is nearly the reciprocal of the tree canopy cover relationship, with low grass cover in older neighborhoods and higher grass cover in newer neighborhoods.

After running diagnostics required for linear regression analysis (Faraway, 2005) we fit regression models to both the tree canopy and grass cover data. Because we were using spatial data and used all the neighborhoods in the study area as observations, we tested model residuals for spatial dependence. Finding high positive spatial autocorrelation in our data, we adopted a simultaneous autoregressive model (SAR) for the regression analysis using spatial weights defined by the nearest neighborhood centroid (Bivand et al., 2008). We found that a quadratic relationship fit the data well for both models, and used the following model for percent tree canopy,

\[
\text{PercentTC} \sim 19.94 + \text{Age} \ast 0.41774 + \text{Age}^2 \ast -0.00134 \quad [2]
\]

and

\[
\text{PercentGC} \sim 36.17 + \text{Age} \ast -0.24965 + \text{Age}^2 \ast -0.00308 \quad [3]
\]

to model percent grass cover.

*Model Validation (Back-casting)*

Using equations 2 and 3 we predicted past tree canopy and grass cover by subtracting years from the Age parameter in the regression model and applying the model
to all neighborhoods in the study area. To obtain the area of a given landscape type (e.g. tree/shrub canopy) in a neighborhood we converted the predicted percent to a proportion and multiplied the proportion by the total area of the neighborhood. With this approach we tested our combined water demand/landscape prediction model by back-casting tree canopy and grass cover using the regression models, and calculating grass under tree canopy as 60% of tree canopy, and then applying Equation 1 using the predicted landscape areas. We did this for three dates for which we could obtain water use data to use as a benchmark for validation. Water use data for previous years did not report residential outdoor use separately as did the report for 2005 so we estimated residential outdoor use for past years based on the proportion of municipal water use reported as residential outdoor use in 2005. In 2005 88,672 acre-feet of a total 224,019 acre-feet, or 39% of municipal water, was used for residential outdoor purposes. Using this fraction we calculated the amount of water that was likely used for residential outdoor purposes based on total municipal water use reported in 1975 (SLCo, 1977), 1985 (USGS, 2010) and 1995 (USGS, 2010).

Results

Fig. 4-4 presents the results of our back-cast validation using the tree/shrub and grass regression models to predict the amount of these landscape types for past decades. Under-canopy grass was calculated at 60% of predicted tree/shrub canopy. Reported residential outdoor water use for all of 2005 was 88,672 acre-feet (76,495 potable + 12,177 secondary) based on estimates reported by the Utah Division of Water Resources (UDNR-DWR, 2009). Our estimate for residential landscape water demand based on
measured GIS data was 89,616 acre-feet for the summer months of 2005. Our estimate based on modeled data (vegetation from regression models) for 2005 was 85,532 acre-feet. The model performs less well for the other three benchmark dates (Fig. 4-4). It should be noted that benchmark residential outdoor use for 1975, 1985, and 1995 are estimates based on the proportion of total municipal use used for residential outdoor purposes in 2005. Total municipal water use data for 1985 and 1995 come from national reports produced by the USGS (Solley et al. 1988, 1998) and while figures in these reports are based on local data, there may be uncertainty in these estimates. Our estimate for 1975 comes within 10,000 acre-feet of the benchmark estimate based on data obtained from Salt Lake County for that year (SLCo, 1977).

Errors in our modeling approach can be attributed to one or both of the major model components—the landscape area prediction using regression analysis (and the choice of 60% of canopy for under-canopy grass cover), or the choice of parameters (Plant Factors and Distribution Uniformity factor) for the water demand equation. Variables influencing model performance are the amount of landscape vegetation (urban vegetation GIS dataset) and climatic conditions for a given summer season (weather station data) both of which are subject to measurement error.

Apparent Influence of Climatic Conditions

The influence of climatic conditions on the model is demonstrated by a comparison of the prediction for 2005 and the prediction for 1995 (Fig. 4-4). New residential land developed in the 1995-2005 decade was approximately 11,000 acres compared to approximately 18,000 acres in the decade from 1985-1995 and 16,000 acres
in the decade from 1975-1985. Based on new residential land area alone, we would expect the water demand slope between 1995 and 2005 to lessen or at least remain constant given a slowing of residential growth for that time period. Instead the model suggests there is a 32% increase in water demand between 1995 and 2005. To understand why the model predicts an increase in water demand for 2005, we looked at the climate data (Table 4-1) and found the summer of 2005 to be hotter and drier than the summer of 1995 (the difference between ETo and precipitation was greater in 2005 for four of the five months). To determine whether climatic factors influenced our water demand prediction, we ran the model for again for 2005 using 1995 climate data and found the increased demand in 2005 to be only 11%—demand that would be accounted for by the relatively moderate increase in residential land development. This suggests that the model is sensitive to climatic conditions as well as increases in residential development.

*Effects of a Maturing Urban Forest on Water Demand*

To explore the effects of a maturing urban forest on water demand we forecast urban forest growth under the hypothetical scenario of no residential expansion (i.e. no new neighborhoods). Future abundances of tree/shrub canopy and grass cover were estimated with regression models, and under-canopy grass was estimated at 60% of predicted tree/shrub canopy. Considering the reciprocal relationship between tree/shrub canopy and grass cover (Fig. 4-3) we expected the model to predict increased tree/shrub cover and decreased grass cover over time, which is what it did (Fig. 4-5). Using predicted amounts of tree/grass canopy, grass cover, and grass under-canopy and the
water demand equation, we estimated water demand for 10 year increments from 2010 to 2050 (Fig. 4-6). We used 30-year averages (1979-2009) for ETo and precipitation (Table 4-1). Fig. 4-6 demonstrates that as tree/shrub canopy increases and exposed grass decreases, overall water demand decreases. This is explained by the relative differences in which our model accounts for water lost through evapotranspiration by grass, tree/shrub canopy, and grass under-canopy (0.80 of ETo for grass, 0.50 for tree canopy, and 0.40 for under-canopy grass) and changes in the overall composition of these landscape types as the urban forest matures. It suggests that a maturing forest in a metropolitan area such as Salt Lake County has the effect of slightly decreasing overall residential water demand over time.

*Projecting Residential Water Demand Based on Landscape Needs*

Modeling urban forest growth under a no residential expansion scenario sheds light on how urban forest canopy dynamics influence irrigation water demand, but is not realistic in Salt Lake County where there is still room for residential expansion, and little to no likelihood of a moratorium on residential expansion in the near future (Envision Utah, 2009). For a more realistic projection of residential water demand we implemented the same urban vegetation growth models previously described, but with the inclusion of simulated residential expansion. Future residential expansion is likely to occur in the southwest section of the valley and in a few islands of farmland currently surrounded by residential development. This is apparent by examining the availability of developable land in these areas and suburban growth patterns over the last 40 years.
To simulate future growth we identified areas (polygons) that are likely to be developed as residential neighborhoods based on their zoning designation within the county parcels GIS dataset. During the decade from 1995-2005 approximately 11,000 acres of new residential land was developed in the study area. Based on this figure we simulated residential growth by adding an additional 10,000 acres of residential development for each decade from 2015 to 2045. The exact location of the polygons is not critical for the water demand model since water demand depends only on the area of landscape type. However, the spatial configuration of the simulated neighborhoods is important in order to justify the application of the simultaneous autoregressive regression (SAR) model for prediction. For this reason we simulated new neighborhoods with similar spatial structure and size as the actual neighborhoods.

Figure 4-7 presents maps of predicted urban vegetation change and water demand for 2010 and 2040. Projections of water demand by landscape type are presented in Fig. 4-8. Unlike Fig. 4-6 we see a steady increase in water demand for all three landscape types due to the increasing area of land in residential development. Fig. 4-9 presents the model’s prediction of residential landscape water demand from 1975 to 2050 with simulated residential expansion. We used recorded ETo and precipitation and modeled urban vegetation for prediction from 1975-2005. For predictions from 2010-2050 we used 30-year (1979-2009) averages for ETo and precipitation, and modeled urban vegetation with linear growth of about 10,000 acres of new residential development per decade.

While uncertainty exists in several components of the model, it is only possible to measure uncertainty associated with the regression components of the vegetation
prediction model. Figure 4-9 shows the upper and lower bounds of uncertainty attributed to the regression predictions for tree/shrub canopy and grass cover. These bounds were computed by determining pseudo prediction intervals at 95% confidence for the regression models (Fig. 4-3) and calculating water demand at decennial intervals based on tree/shrub canopy and grass cover predictions within that range of uncertainty.

Discussion

An important question addressed by this study is how a growing urban forest affects residential landscape water demand. In order to properly address this question we must first consider how a growing urban forest affects the overall structure of residential landscaped areas. Data on tree canopy and neighborhood age show that over time urban tree canopy increases, and similar data on exposed turf grass suggests that the amount of exposed turf grass decreases with time (Fig. 4-3). We make the assumption that most of the decrease in exposed turf grass is not grass converted to some other landscape type, but grass covered by tree canopy. This is an assumption, but it is a logical one. A key point brought out by this research is that a growing urban forest canopy has the often unacknowledged effect of increasing the amount of covered turf grass. An important consequence of this is a reduction in water lost through evapotranspiration by residential turf grass landscaping.

Our analysis of the effects of urban forest growth under the hypothetical scenario of no residential expansion demonstrates that increasing tree canopy results in a slight reduction in overall water demand from the landscape. This is due to increasing under-canopy turf grass and decreasing exposed turf grass. It is important to recognize that
these results depend on the chosen water-loss coefficients for the landscape types. For example, our choice of water-loss coefficient (i.e. Plant Factor) for under-canopy grass is half that of exposed turf grass (0.40 versus 0.80). This is a logical assumption which produced reasonable results, and has been followed by others (Farag, 2003; Endter-Wada et al., 2008) but is nevertheless an assumption. Having more reliable water-loss coefficients, particularly for under-canopy turf grass, would improve confidence in our ability to assess these effects. Nevertheless, based on the outcome of our model, it is reasonable to conclude that a growing urban forest does not increase landscape irrigation demand, and may in fact result in a decreased demand for irrigation water. It should be noted however that our analysis focuses on irrigation demand, and trees with deep or extensive root systems are likely benefiting from groundwater resources. A greater abundance of tree canopy therefore may not reduce overall water consumption at the watershed level.

Another goal of this study was to determine whether we could estimate residential landscape water demand, and whether we could reasonably project future water requirements in the study area. Our estimate of 89,616 acre-feet using GIS data for 2005 is very close to the 88,672 acre-feet estimated by the Utah Division of Water Resources, which suggests the model includes the most important variables determining residential landscape water demand, and that those variables are reasonably measured and parameterized. However it is commonly recognized that more water than is required is often applied to residential landscapes (Kjelgren et al., 2000; Farag, 2003; Endter-Wada et al., 2008). In addition to inefficiency due to non-uniform application of irrigation water by irrigation systems (accounted for by the DU factor in our model) inefficiencies
occur when automatic sprinkler systems apply water when there has been adequate 
precipitation or even while it is raining (Kjelgren et al., 2000). In a parcel-based study in 
Layton, Utah, Farag (2003) found for example that excesses of 15% water use were 
common and that some homeowners applied up to 70% excess irrigation water. Given 
that more water is typically applied than is actually needed, we suspect that while our 
estimate for 2005 is in the range of being reasonable, it is probably high if the benchmark 
of 88,672 acre-feet from the Utah Division of Water Resources is correct.

Predicting future water needs is a major concern for all urbanized areas, but is 
especially critical in semi-arid environments such as Salt Lake County. A variety of 
methods can be used to predict future water needs, most involving some form of 
population projection (Baumann et al., 1998). To our knowledge no one has attempted to 
predict residential landscape water demand as a separate component of municipal water 
needs. However, residential (and commercial) landscape watering offers one of the most 
promising segments of the urban water budget for conservation (Kjelgren et al., 2000; 
Endter-Wada et al., 2008). Evidence for this comes from the success of Utah’s “Slow the 
Flow: Save H2O” campaign to encourage homeowners to be more conscientious about 
how they irrigate their landscapes (Endter-Wada et al., 2008). Projecting future demand 
based on current landscape management practices and comparing it to more water-
conserving practices would allow water suppliers the ability to assess potential 
conservation gains. Operationally this could be implemented in our model simply by 
raising the DU factor to simulate more efficient water application. Our analysis of 
projected future water demands are based on linear growth in residential development, 
and 30-year averages for ETo and precipitation. Both are rough approximations of actual
growth and potential future climatic conditions. Given the sensitivity of the model to these two components, as evidenced by fluctuation during the 1995-2005 period, there would be merit in running the landscape water demand prediction using more reliable estimates of land use change and predicted future climatic conditions.

**Conclusion**

This study presents an innovative approach to a better understanding of how residential water demand is influenced by a growing urban forest. As cities throughout the world encourage the cultivation of urban trees, it is important to more fully understand the consequences of these actions. Our study suggests that an increase in urban forest abundance does not result in an increase in irrigation water demand, and that it may in fact decrease demand. Relative differences in evapotranspiration rates of tree canopy, exposed turf grass, and under-canopy turf grass account for this observation. An often unacknowledged outcome of a growing forest canopy is the effect of an increased area of under-canopy turf grass. Our vegetation-based water demand model also suggests that climatic factors play an important role in determining the overall irrigation requirements for a large metropolitan area such as Salt Lake County. Future climatic conditions are likely to influence residential water requirement, particularly in semi-arid environments where evapotranspiration exceeds precipitation. Water conservation policies and practices in urban environments where water is scarce are important now, and likely to be increasingly important in the future.
References


interface using remote sensing and geographic information systems. In: *Irrigation
Engineering*, Utah State University, Logan, Utah, pp. 310.


International 25 (1), 127-138.

Grove, J. M., Troy, A. R., O'Neil-Dunne, J., Burch, W. R. J., Cadenasso, M. L., Pickett,
S. T., 2006. Characterization of households and its implications for vegetation of
urban ecosystems. Ecosystems 9, 578-597.

Management & Policy 8 (1), 33-47.

PNAS 100 (15), 8788-8792.

Huang, Y. J., Akbari, H., Taha, H., 1990. The wind-sheilding and shading effects of trees
on residential heating and cooling. ASHRAE 96, 1403-1411.

Iverson, L. R., Cook, E. A., 2000. Urban forest cover of the Chicago region and its
relation to household density and income. Urban Ecosystems 4 (105-124).

modelling at IIT Kanpur using artificial neural networks. Water Resources


HortScience 35 (6), 1037-1040.

irrigation. Water requirements and irrigation, ASA-CSSA-SSSA, Madison, WI,
pp. 441-444.

Maco, S. E., McPherson, E. G., 2002. Assessing canopy cover over streets and sidewalks

a useful predictor of perennial landscape vegetation in residential neighborhoods
and embedded small parks of Phoenix, AZ. Landscape and Urban Planning 69, 355-368.


Nowak, D. J., Crane, D. E., 2002. Carbon storage and sequestration by urban trees in the USA. Environmental Pollution 116 (381-389), 381.


Table 4-1. Relative evapotranspiration and precipitation data used to estimate residential water demand.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Cottonwood Weir</td>
<td>4.25</td>
<td>3.57</td>
<td>5.37</td>
<td>1.22</td>
<td>6.52</td>
<td>0.17</td>
<td>6.32</td>
<td>0.24</td>
<td>3.94</td>
<td>0.31</td>
<td>26.40</td>
<td>5.50</td>
</tr>
<tr>
<td>SL Int'l Airport</td>
<td>4.92</td>
<td>2.61</td>
<td>6.38</td>
<td>1.83</td>
<td>8.05</td>
<td>0.30</td>
<td>6.89</td>
<td>0.14</td>
<td>4.91</td>
<td>0.09</td>
<td>31.15</td>
<td>4.96</td>
</tr>
<tr>
<td>Average</td>
<td>4.58</td>
<td>3.09</td>
<td>5.87</td>
<td>1.53</td>
<td>7.29</td>
<td>0.23</td>
<td>6.61</td>
<td>0.19</td>
<td>4.42</td>
<td>0.20</td>
<td>28.77</td>
<td>5.23</td>
</tr>
<tr>
<td>Cottonwood Weir</td>
<td>5.80</td>
<td>4.71</td>
<td>7.10</td>
<td>2.53</td>
<td>7.73</td>
<td>1.26</td>
<td>6.71</td>
<td>0.00</td>
<td>4.18</td>
<td>3.13</td>
<td>31.53</td>
<td>11.63</td>
</tr>
<tr>
<td>SL Int'l Airport</td>
<td>5.88</td>
<td>3.00</td>
<td>7.23</td>
<td>1.31</td>
<td>8.03</td>
<td>0.88</td>
<td>6.93</td>
<td>0.04</td>
<td>4.13</td>
<td>2.00</td>
<td>32.20</td>
<td>7.21</td>
</tr>
<tr>
<td>Average</td>
<td>5.84</td>
<td>3.85</td>
<td>7.17</td>
<td>1.92</td>
<td>7.88</td>
<td>1.07</td>
<td>6.82</td>
<td>0.02</td>
<td>4.15</td>
<td>2.56</td>
<td>31.86</td>
<td>9.42</td>
</tr>
<tr>
<td>Cottonwood Weir</td>
<td>4.73</td>
<td>6.91</td>
<td>5.93</td>
<td>2.86</td>
<td>7.13</td>
<td>0.76</td>
<td>6.52</td>
<td>0.11</td>
<td>4.66</td>
<td>1.47</td>
<td>28.97</td>
<td>12.11</td>
</tr>
<tr>
<td>SL Int'l Airport</td>
<td>4.78</td>
<td>3.71</td>
<td>6.23</td>
<td>1.51</td>
<td>7.90</td>
<td>0.34</td>
<td>7.21</td>
<td>0.24</td>
<td>4.39</td>
<td>1.35</td>
<td>30.51</td>
<td>7.14</td>
</tr>
<tr>
<td>Average</td>
<td>4.76</td>
<td>5.31</td>
<td>6.08</td>
<td>2.19</td>
<td>7.51</td>
<td>0.55</td>
<td>6.87</td>
<td>0.17</td>
<td>4.52</td>
<td>1.41</td>
<td>29.74</td>
<td>9.62</td>
</tr>
<tr>
<td>Cottonwood Weir</td>
<td>5.27</td>
<td>5.03</td>
<td>6.04</td>
<td>2.52</td>
<td>7.38</td>
<td>0.00</td>
<td>6.01</td>
<td>0.26</td>
<td>4.28</td>
<td>1.52</td>
<td>28.98</td>
<td>9.33</td>
</tr>
<tr>
<td>SL Int'l Airport</td>
<td>5.35</td>
<td>2.91</td>
<td>6.38</td>
<td>1.68</td>
<td>8.34</td>
<td>0.02</td>
<td>6.74</td>
<td>0.74</td>
<td>4.58</td>
<td>0.41</td>
<td>31.39</td>
<td>5.76</td>
</tr>
<tr>
<td>Average</td>
<td>5.31</td>
<td>3.97</td>
<td>6.21</td>
<td>2.10</td>
<td>7.86</td>
<td>0.01</td>
<td>6.38</td>
<td>0.50</td>
<td>4.43</td>
<td>0.97</td>
<td>30.19</td>
<td>7.54</td>
</tr>
<tr>
<td>Cottonwood Weir</td>
<td>5.79</td>
<td>1.97</td>
<td>7.18</td>
<td>0.97</td>
<td>8.26</td>
<td>0.67</td>
<td>7.10</td>
<td>0.66</td>
<td>4.75</td>
<td>1.23</td>
<td>33.08</td>
<td>5.50</td>
</tr>
<tr>
<td>Average</td>
<td>5.71</td>
<td>2.52</td>
<td>7.02</td>
<td>1.24</td>
<td>7.97</td>
<td>0.86</td>
<td>6.89</td>
<td>0.88</td>
<td>4.62</td>
<td>1.58</td>
<td>32.20</td>
<td>7.08</td>
</tr>
</tbody>
</table>
Fig. 4-1. Urban vegetation classification of 1-meter digital orthophotography used to attribute neighborhood polygons.
Fig. 4-2. Map of study area showing median age of housing stock by neighborhood.
Fig. 4-3. Plots showing the relationship of tree canopy to neighborhood age, and exposed turf grass cover to neighborhood age.
Fig. 4-4. Plot of predicted residential irrigation water demand compared to benchmark reported water use for 1975-2005.

* Note: For 2005, 85,532 acre-feet were predicted using regression modeled tree canopy and grass cover. Using the measured GIS data, 89,616 acre-feet were predicted.
Fig. 4-5. Maps of predictions of tree canopy, turf grass cover, and water demand for 2010 and 2040 under hypothetical scenario of no residential expansion.
**Fig. 4-6.** Estimated irrigation water demand from 2010-2050 under hypothetical scenario of no residential expansion.
Fig. 4-7. Maps of predictions of tree canopy, turf grass cover, and water demand for 2010 and 2040 with simulated residential growth.
Fig. 4-8. Estimated water demand from 2010-2050 with simulated residential expansion.
Fig. 4-9. Plot of predictions of irrigation water demand for residential landscapes.
Urban areas are complex systems comprised of a variety of biotic and abiotic components that interact and are often interdependent upon one another. *Systems analysis* is a methodological framework that provides terms and concepts that help us understand and describe relationships within complex systems (Aber and Melillo, 2001). A systems approach also allows us to focus on specific components or relationships while recognizing that they comprise only portions of a larger system. The focus of this research has been on just a few components of the Salt Lake County urban ecosystem.

An understanding of the structure of an ecosystem is an important first step before we can fully understand how it functions (Aber and Melillo, 2001; Alberti, 2005). Chapter 2 focused on measuring spatial patterns of human activities on the urban landscape (Fig. 5-1). Using a variety of metrics common to the literature on urban sprawl we mapped the demographic and spatial structure of three types of neighborhood defined by the era during which it was built. We found that the urban landscape is indeed heterogeneous (Band et al., 2005), and that much of the spatial form of an urban landscape is related to the era during which it was developed. While not specifically studied by our research, we hypothesize that the drivers for the structure we measured can be attributed to differences in cultural values, available technology, and economic conditions of the era during which a neighborhood was developed (Fig. 5-1).

Chapter 3 focused on the relationship between urban tree canopy abundance and human and physical characteristics of the urban landscape (Fig. 5-1). In particular we
found that the amount of urban tree canopy is related to household income and family life-stage, but that this relationship changes with the amount of time since the neighborhood was developed. We tested the hypothesis that variations in urban form might be related to tree canopy abundance, but found the relationship weak to non-existent, depending on the age of the neighborhood.

Perhaps the most interesting results from this research are in Chapter 4, where we begin to understand how the urban ecosystem functions, and in particular how urban vegetation patterns influence residential irrigation water demands. Using terms common to systems analysis we may say there is a negative feedback effect between a growing forest and the amount of exposed turf grass. As the forest canopy grows, more turf grass is covered by canopy, which has a stabilizing effect on the amount of irrigation water required by the system. It is true that irrigation water demand increases with continued residential expansion, but based on what our analysis suggests a growing urban forest has at least some attenuating effect on overall water demand.

Resiliency of a system refers to the degree to which a system responds to a disturbance and returns to a stable state (Aber and Melillo, 2001). Although we did not explicitly address the effects of changes in climatic conditions (Fig. 5-1) on irrigation water demand, our analysis seems to indicate that water demand for any given year is dependent on reference evapotranspiration and precipitation conditions for that year. An analysis of the sensitivity of our model to changes in climatic conditions and the resiliency of the residential landscape water demand system would be an important area for future research.
Finally, this study has examined the Salt Lake County urban ecosystem within a spatial analytical framework. The spatial nature of the datasets used in the statistical analyses required special attention to spatial processes within the data. In all three parts of the study (Chapters 2, 3, and 4) a simultaneous autoregressive (SAR) regression model was used instead of an ordinary least squares (OLS) regression model because we found spatial autocorrelation in the OLS model residuals. From an operational perspective this was a justifiable approach because the SAR model improves estimation by taking advantage of spatial autocorrelation in the data (Bailey and Gatrell, 1995). Trying to understand the theoretical implications of spatial autocorrelation in the data is a little less straightforward. At the most basic level it tells us that residential neighborhoods that are near one another are similar to one another (Tobler’s (1970) First Law of Geography).

What this means in terms of the questions addressed by this research is a prime topic for future work.

References


Fig. 5-1. Urban ecosystem conceptual model for Salt Lake Valley residential landscapes.
APPENDICES
Appendix A
Supplemental Information for Chapter 2
### Table A-1. Pearson’s test for correlation for 18 urban form metrics

<table>
<thead>
<tr>
<th>Density Metrics</th>
<th>MEDLOTACRE</th>
<th>HUTPERSQKM</th>
<th>MEDNURMS</th>
<th>POPPERSQKM</th>
<th>POPPERHU</th>
</tr>
</thead>
<tbody>
<tr>
<td>MEDLOTACRE</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HUTPERSQKM</td>
<td>-0.38</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MEDNURMS</td>
<td>0.46</td>
<td>-0.60</td>
<td>1.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>POPPERSQKM</td>
<td>-0.42</td>
<td>0.91</td>
<td>-0.62</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>POPPERHU</td>
<td>0.33</td>
<td>-0.48</td>
<td>0.54</td>
<td>-0.28</td>
<td>1.00</td>
</tr>
<tr>
<td>Medlotacre</td>
<td>Hutsperkm</td>
<td>Mednurms</td>
<td>Popperskm</td>
<td>Popperskm</td>
<td>Popperhu</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Centrality Metrics</th>
<th>COMMEAN</th>
<th>PRKMEAN</th>
<th>SCHMEAN</th>
<th>BUSMEAN</th>
</tr>
</thead>
<tbody>
<tr>
<td>COMMEAN</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PRKMEAN</td>
<td></td>
<td>1.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SCHMEAN</td>
<td>0.23</td>
<td>0.07</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>BUSMEAN</td>
<td>0.50</td>
<td>0.04</td>
<td>0.37</td>
<td>1.00</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Accessibility Metrics</th>
<th>R_LINK_NOD</th>
<th>MED_BLK_PE</th>
<th>R_LINK_CUL</th>
<th>MED_CUL_LE</th>
</tr>
</thead>
<tbody>
<tr>
<td>R_LINK_NOD</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MED_BLK_PE</td>
<td>-0.20</td>
<td>1.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>R_LINK_CUL</td>
<td>0.06</td>
<td>-0.11</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>MED_CUL_LE</td>
<td>-0.04</td>
<td>-0.11</td>
<td>-0.03</td>
<td>1.00</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Neighborhood Mix Metrics</th>
<th>IJI</th>
<th>PR</th>
<th>SIDI</th>
<th>PRPWRKOUT</th>
<th>R_RENT_OWN</th>
</tr>
</thead>
<tbody>
<tr>
<td>IJI</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PR</td>
<td>0.32</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SIDI</td>
<td>0.28</td>
<td>0.07</td>
<td>1.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PRPWRKOUT</td>
<td>-0.07</td>
<td>0.02</td>
<td>-0.01</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>R_RENT_OWN</td>
<td>0.34</td>
<td>0.08</td>
<td>0.04</td>
<td>-0.21</td>
<td>1.00</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>IJI</th>
<th>PR</th>
<th>SIDI</th>
<th>PRPWRKOUT</th>
<th>R_RENT_OWN</th>
</tr>
</thead>
<tbody>
<tr>
<td>IJI</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PR</td>
<td>0.32</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SIDI</td>
<td>0.28</td>
<td>0.07</td>
<td>1.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PRPWRKOUT</td>
<td>-0.07</td>
<td>0.02</td>
<td>-0.01</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>R_RENT_OWN</td>
<td>0.34</td>
<td>0.08</td>
<td>0.04</td>
<td>-0.21</td>
<td>1.00</td>
</tr>
</tbody>
</table>
### Table A-2. Results for Moran’s I test for spatial dependence.

<table>
<thead>
<tr>
<th>Metric Name</th>
<th>Moran’s I (LM)</th>
<th>Moran’s I (SAR)</th>
<th>Z-Score (SAR)</th>
<th>P-Value (SAR)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Density Metrics</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Median SF Res. Lot Size</td>
<td>MEDLOTACRE</td>
<td>0.4537</td>
<td>0.0442</td>
<td>1.1424</td>
</tr>
<tr>
<td>Housing Density</td>
<td>HUTPERSQKM</td>
<td>0.3096</td>
<td>-0.0176</td>
<td>-0.2924</td>
</tr>
<tr>
<td>Median Number of Rooms</td>
<td>MEDNURMS</td>
<td>0.5302</td>
<td>-0.0344</td>
<td>-0.6037</td>
</tr>
<tr>
<td>Population Density</td>
<td>POPPERSQKM</td>
<td>0.3472</td>
<td>-0.0204</td>
<td>-0.3440</td>
</tr>
<tr>
<td>Population /Housing Unit</td>
<td>POPPERHU</td>
<td>0.4974</td>
<td>0.0033</td>
<td>0.0962</td>
</tr>
<tr>
<td><strong>Centrality Metrics</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Distance to Commercial Zone</td>
<td>COMMEAN</td>
<td>0.8911</td>
<td>0.1758</td>
<td>3.2918</td>
</tr>
<tr>
<td>Distance to Neighborhood and City Parks</td>
<td>PRKMEAN</td>
<td>0.7552</td>
<td>0.0262</td>
<td>0.5201</td>
</tr>
<tr>
<td>Distance to K-12 Schools</td>
<td>SCHMEAN</td>
<td>0.4930</td>
<td>0.0418</td>
<td>0.8085</td>
</tr>
<tr>
<td>Distance to Bus Stops (public transportation)</td>
<td>BUSMEAN</td>
<td>0.5373</td>
<td>0.0492</td>
<td>1.2654</td>
</tr>
<tr>
<td><strong>Accessibility Metrics</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ratio Streets to Intersections</td>
<td>R_LINK_NOD</td>
<td>0.4550</td>
<td>-0.0138</td>
<td>-0.2214</td>
</tr>
<tr>
<td>Median Perimeter of Residential Blocks</td>
<td>MED_BLK_PE</td>
<td>0.2308</td>
<td>-0.0254</td>
<td>-0.4367</td>
</tr>
<tr>
<td>Ratio of Streets to Cul-de-sacs</td>
<td>R_LINK_CUL</td>
<td>0.2314</td>
<td>-0.0162</td>
<td>-0.2665</td>
</tr>
<tr>
<td>Median Length of Cul-de-sacs</td>
<td>MED_CUL_LE</td>
<td>0.1671</td>
<td>-0.0127</td>
<td>-0.2011</td>
</tr>
<tr>
<td><strong>Neighborhood Mix Metrics</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Land use Contiguity</td>
<td>IJI</td>
<td>0.2800</td>
<td>-0.0071</td>
<td>-0.0971</td>
</tr>
<tr>
<td>Land use Richness</td>
<td>PR</td>
<td>0.2017</td>
<td>-0.0098</td>
<td>-0.1465</td>
</tr>
<tr>
<td>Land use Diversity</td>
<td>SIDI</td>
<td>0.1150</td>
<td>-0.0002</td>
<td>0.0308</td>
</tr>
<tr>
<td>Proportion of People Working Outside City of Residence</td>
<td>PRPWRKOUT</td>
<td>0.6819</td>
<td>-0.0875</td>
<td>-1.5880</td>
</tr>
<tr>
<td>Ratio Renters to Owners</td>
<td>R_RENT_OWN</td>
<td>0.2614</td>
<td>0.0531</td>
<td>1.0180</td>
</tr>
</tbody>
</table>

Notes: Spatial neighbors were defined as the first nearest centroid for all urban form metrics, except MEDLOTACRE, and BUSMEAN. These models required a proximity matrix based on 2 nearest neighbors to bring the z-score for the Moran’s I statistic within the 95% confidence interval (-1.96 < z < 1.96). It was not possible to achieve a z-score within the 95% confidence interval for the COMMEAN model. All p-values are for a two-tailed test.
Fig. A-1. Maps of urban form metrics 1-6.
Fig. A-2. Maps of urban form metrics 7-12.
Fig. A-3. Maps of urban form metrics 13-18.
Appendix B
Supplemental Information for Chapter 3
Residuals vs Fitted: Check for non-constant variance.
Normal Q-Q: Check for normality
Scale-Location: Check for non-constant variance.
Residuals vs Leverage: Check for leverage points.

![Diagnostic plots for non-constant variance, normality, and leverage points.](image)

**Fig. B-1.** Diagnostic plots for non-constant variance, normality, and leverage points.
Fig. B-2. Partial residual plots. Check for non-constant variance and non-linearity.
Fig. B-3. Added-variable plots. Check for influential or unusual observations.
Table B-1. Pearson’s moment correlations. Check for multi-collinearity among explanatory variables.

<table>
<thead>
<tr>
<th></th>
<th>NHAGE</th>
<th>MDINC</th>
<th>AVHSZ</th>
<th>POAGE</th>
<th>PNWT</th>
<th>PGTHS</th>
<th>STCON</th>
<th>LUMIX</th>
<th>MDLOT</th>
<th>RSDENS</th>
<th>MBLKP</th>
<th>MPRCP</th>
<th>AVWTS</th>
<th>DSTSC</th>
<th>ASWST</th>
<th>ASSTH</th>
</tr>
</thead>
<tbody>
<tr>
<td>Neighborhood Age*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.4</td>
<td>0.1</td>
<td>-0.6</td>
<td>0.5</td>
<td>-0.4</td>
<td>0.2</td>
<td>0.1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Median Income</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-0.3</td>
<td>0.6</td>
<td>-0.4</td>
<td>0.3</td>
<td>0.5</td>
<td>-0.3</td>
<td>-0.1</td>
</tr>
<tr>
<td>Ave. Household Size</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-0.4</td>
<td>-0.3</td>
<td>0.4</td>
<td>-0.3</td>
<td>0.4</td>
<td>-0.2</td>
<td>0.1</td>
<td></td>
<td>-0.2</td>
<td>0.1</td>
</tr>
<tr>
<td>Med. Population Age</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.2</td>
<td>-0.1</td>
<td>0.0</td>
<td>0.1</td>
<td>-0.1</td>
<td>0.4</td>
<td>0.0</td>
<td>0.4</td>
<td></td>
<td>-0.1</td>
</tr>
<tr>
<td>Non-White Population</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.2</td>
<td>0.2</td>
<td>-0.5</td>
<td>0.2</td>
<td>-0.2</td>
<td>-0.5</td>
<td>0.0</td>
<td>0.0</td>
<td></td>
<td>-0.1</td>
</tr>
<tr>
<td>High School Graduates</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-0.1</td>
<td>-0.2</td>
<td>0.4</td>
<td>-0.2</td>
<td>0.2</td>
<td>0.5</td>
<td>-0.1</td>
<td>-0.1</td>
<td>0.2</td>
<td>0.3</td>
</tr>
<tr>
<td>Street Connectivity</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.3</td>
<td>-0.3</td>
<td>0.0</td>
<td>-0.3</td>
<td>0.2</td>
<td></td>
<td>-0.1</td>
</tr>
<tr>
<td>Land Use Mix</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-0.1</td>
<td>0.1</td>
<td>-0.2</td>
<td>0.1</td>
<td>0.0</td>
</tr>
<tr>
<td>Median Lot Size</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.1</td>
<td>-0.1</td>
<td>-0.1</td>
<td>0.1</td>
<td>0.2</td>
<td>-0.1</td>
</tr>
<tr>
<td>Res. Street Density</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.3</td>
<td>0.3</td>
<td>0.2</td>
<td>0.0</td>
</tr>
<tr>
<td>Median Block Perim.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.2</td>
<td>-0.1</td>
<td>0.2</td>
</tr>
<tr>
<td>Mean Annual Precip.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-0.5</td>
<td>0.0</td>
</tr>
<tr>
<td>Avail. Water Storage</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-0.1</td>
</tr>
<tr>
<td>Dist. to Streams/Canals</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Aspect (Westness)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Aspect (Southness)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Variance Inflation Factors: Check for multi-collinearity among explanatory variables

| NHAGE | MDINC | AVHSZ | POAGE | PNWT | PGTHS | STCON | LUMIX | MDLOT | RSDENS | MBLKP | MPRCP | AVWTS | DSTSC | ASWST | ASSTH |
|-------|-------|-------|-------|------|-------|-------|-------|-------|--------|-------|-------|-------|-------|-------|-------|-------|
| 4.10  | 4.83  | 4.05  | 1.97  | 1.97 | 4.52  | 4.59  | 1.59  | 1.66  | 2.49   | 1.92  | 1.47  | 3.94  | 1.77  | 1.25  | 2.64  | 1.51  |
Table B-2. Moran’s I test for spatial autocorrelation of model residuals.

**Ordinary Least Squares model:**

Moran's I test under normality

data: resid(lm.model)
weights: listW.pts

Moran I statistic standard deviate = 8.2545, p-value < 2.2e-16
alternative hypothesis: two.sided
sample estimates:
Moran I statistic       Expectation          Variance
0.452980115      -0.001923077       0.003037097

**Simultaneous Autoregressive model:**

Moran's I test under normality

data: resid(sar.model)
weights: listW.pts

Moran I statistic standard deviate = -0.2626, p-value = 0.7929
alternative hypothesis: two.sided
sample estimates:
Moran I statistic       Expectation          Variance
-0.016392358      -0.001923077       0.003037097
Fig. B-4. Maps of determinants of urban tree canopy 1-6.
Fig. B-5. Maps of determinants of urban tree canopy 7-12.
Fig. B-6. Maps of determinants of urban tree canopy 13-15.
Appendix C
R-Code for Statistical Analyses
# Research Part I
# Purpose: Runs a sensitivity analysis of K nn for a sar model, outputs *.csv with morans statistics
# Requires the spdep library to be set
# Runs a SAR model with the following user defined parameters:
# spatob: spatial dataframe object, created with readShapePoint function
# x: dependent variable from the dataframe (e.g. ERA)
# y: independent variable from the dataframe (e.g. MEDLOTSIZE)
# k: number of k nearest neighbors
#
# create data frame from spatob
df.pts = as.data.frame(spatob)

# create empty output matrix
out.mat = matrix(nrow = k, ncol = 3)

for (i in 1:k) {
    # create proximity matrix based on k nn
    knn.pts = knearneigh(spatob, i)

    # create neighbors list from knn object
    list.nb = knn2nb(knn.pts)

    # create proximities object with weight style W
    listW.pts = nb2listw(list.nb, style = "W")

    # run the SAR model
    sar.Model = spautolm(paste(y, '~', x), df.pts, listW.pts, family="SAR")

    # Morans I test for autocorrelation of residuals of SAR model
    m.test = moran.test (resid(sar.Model), randomisation=FALSE, listW.pts)

    # send k, morans i, and standard to output matrix
    out.mat[i,1] = i
    out.mat[i,2] = m.test$estimate[[1]]
    out.mat[i,3] = m.test$statistic
}

###
## write out.mat to .csv file and plot graph
###
# write out.mat to text file
colnames(out.mat) = c("kval", "mstat", "sdev")
fileName = paste("C:\PhD_Work\DATA\ANALYSIS\PART_I\STATS\OUT\",y,".csv")
write.table(out.mat, file = fileName, col.names = TRUE, sep="", row.names=FALSE)

# plot graph
plot(out.mat[,2], type="o", pch=16, main=paste("SPAUTOLM (",y,"~ ERA)"), cex.main=1.0,
xlab="K number of nearest neighbors", ylab= "Moran's I statistic")

abline(h=0, col="grey")
}
# Research Part I
# Name: Run.SAR.R
# Purpose: Runs SAR.function for 18 models.
# Used to determine best K nn for each model.

# Clean up
# list all vars, then remove all variables, then show all are gone
ls()
rm(list=ls(all.names=TRUE))
ls()

# My Data

# Set working directory
setwd("C:/PhD_Work/DATA/ANALYSIS/PART_I/STATS/SAR")

# Set libraries
library(spdep)

# read in shapefile to a Spatial Data frame Object
pts.shpObj = readShapePoints("DataPts.shp", proj4string=CRS("+proj=utm +zone=12 +datum=NAD83"))

# Set working directory
setwd("C:/PhD_Work/DATA/ANALYSIS/PART_I/STATS/")
source("SAR.function.R")

### Setup map area 3x3 maps
par(mfrow=c(3,3), mex=0.6)

# Density metrics
# run SAR.function.R (spatob, df, x, y, k)
SAR.function.R (pts.shpObj, 'ERA', 'MEDLOTSIZE', 20)
SAR.function.R (pts.shpObj, 'ERA', 'HUTPERSQKM', 20)
SAR.function.R (pts.shpObj, 'ERA', 'EHUMNR', 20)
SAR.function.R (pts.shpObj, 'ERA', 'POPPERSQKM', 10)
SAR.function.R (pts.shpObj, 'ERA', 'POPPERHU', 35)

# Centrality metrics
# run SAR.function.R (spatob, df, x, y, k)
SAR.function.R (pts.shpObj, 'ERA', 'COMMEAN', 50)
SAR.function.R (pts.shpObj, 'ERA', 'PRKMEAN', 200)
SAR.function.R (pts.shpObj, 'ERA', 'SCHMEAN', 50)
SAR.function.R (pts.shpObj, 'ERA', 'BUSMEAN', 50)

# new page, set up number of plots per page
x11(); par(mfrow=c(3,3), mex=0.6)

# Accessibility metrics
# run SAR.function.R (spatob, df, x, y, k)
SAR.function.R (pts.shpObj, 'ERA', 'R_LINK_NOD', 20)
SAR.function.R (pts.shpObj, 'ERA', 'MED_BLK_PE', 20)
SAR.function.R (pts.shpObj, 'ERA', 'R_CUL_LNKS', 20)
SAR.function.R (pts.shpObj, 'ERA', 'MED_CUL_LE', 20)

# Neighborhood Mix metrics
# run SAR.function.R (spatob, df, x, y, k)
SAR.function.R (pts.shpObj, 'ERA', 'IJI', 20)
SAR.function.R (pts.shpObj, 'ERA', 'PR', 20)
SAR.function.R (pts.shpObj, 'ERA', 'SIDI', 20)
SAR.function.R (pts.shpObj, 'ERA', 'PRPWRKOUT', 20)
SAR.function.R (pts.shpObj, 'ERA', 'R_RENT_OWN', 20)
KNN.MORAN.function.R <- function(spatob, x, y, k){
  # Purpose: determines morans stats for a lm and sar model
  # (contiguous neighbors) sends to moran.out.csv
  # Requires the spdep library to be set
  #
  # Runs a LM then moran.test, then runs SAR and moran.test
  # Sends output from the two moran.test to .csv file.
  #
  # spatob: spatial dataframe object, created with readShapePoint function
  # x: dependent variable from the dataframe (e.g. ERA)
  # y: independent variable from the dataframe (e.g. MEDLOTSIZE)
  # k: k number of nearest neighbors
  #
  # create data frame from spatob
df.pts = as.data.frame(spatob)
  # create empty output matrix
out.mat = matrix(nrow = 2, ncol = 6)
  # create proximity matrix based on k nearest neighbors
knn.pts = knearneigh(spatob, k)
  # create neighbors list form knn object
list.nb = knn2nb(knn.pts)
  # Create proximities object with weight style W
listW.pts = nb2listw(list.nb, style = "W")
  # create the linear model
lm.Model = (lm(paste(y,'~',x),df.pts))
  # run Morans I on linear model
lm.m.test = moran.test(resid(lm.Model), randomisation=FALSE, alternative="two.sided",listW.pts)
  # send output of Morans I to out.mat
out.mat[1,1] = paste("lm(",y,"~ERA)")
out.mat[1,2] = lm.m.test$estimate[[1]]
out.mat[1,3] = lm.m.test$estimate[[2]]
out.mat[1,4] = lm.m.test$estimate[[3]]
out.mat[1,5] = lm.m.test$statistic
out.mat[1,6] = lm.m.test$p.value
  # create the SAR model
sar.Model = spautolm(paste(y,'~',x),df.pts, listW.pts, family="SAR")
  # run Morans I on SAR model
sar.m.test = moran.test(resid(sar.Model), randomisation=FALSE, alternative="two.sided", listW.pts)
  # send output of Morans I to out.mat
out.mat[2,1] = paste("sar(",y,"~ERA)")
out.mat[2,2] = sar.m.test$estimate[[1]]
out.mat[2,3] = sar.m.test$estimate[[2]]
out.mat[2,4] = sar.m.test$estimate[[3]]
out.mat[2,5] = sar.m.test$statistic
out.mat[2,6] = sar.m.test$p.value
  # write out.mat to text file
fName = "C:\PhD_Work\DATA\ANALYSIS\PART_I\STATS\OUT\knn.moran.out.csv"
write.table(out.mat, file = fName, append=TRUE, col.names = TRUE, sep="",
)}
# Research Part I
# Name: Run.MORAN.R
# Purpose: Runs KNN.MORAN.function for 18 models.
# Used to determine Moran's I for lm and sar

# Clean up
# list all vars, then remove all variables, then show all are gone
ls()
rm(list=ls(all.names=TRUE))
ls()

# My Data

# Set working directory
setwd("C:/PhD_Work/DATA/ANALYSIS/PART_I/STATS/SAR")

# Set libraries
library(spdep)

# read in shapefile to a Spatial Data frame Object
ptsObj = readShapePoints("DataPtsXYnew.shp", proj4string=CRS("+proj=utm +zone=12 +datum=NAD83"))

# Set working directory
setwd("C:/PhD_Work/DATA/ANALYSIS/PART_I/STATS/")
source("KNN.MORAN.function.R")

# Density metrics
# run KNN.MORAN.function.R (spatob, x, y, k)
KNN.MORAN.function.R (ptsObj, 'ERA', 'MEDLOTACRE', 2)
KNN.MORAN.function.R (ptsObj, 'ERA', 'HUTPERSQKM', 1)
KNN.MORAN.function.R (ptsObj, 'ERA', 'MEDNURMS', 1)
KNN.MORAN.function.R (ptsObj, 'ERA', 'POPPERSQKM', 1)
KNN.MORAN.function.R (ptsObj, 'ERA', 'POPPERHU', 1)

# Centrality metrics
# run KNN.MORAN.function.R (spatob, x, y, k)
KNN.MORAN.function.R (ptsObj, 'ERA', 'COMMEAN', 1)
KNN.MORAN.function.R (ptsObj, 'ERA', 'PRKMEAN', 1)
KNN.MORAN.function.R (ptsObj, 'ERA', 'SCHMEAN', 1)
KNN.MORAN.function.R (ptsObj, 'ERA', 'BUSMEAN', 2)

# Accessiblity metrics
# run KNN.MORAN.function.R (spatob, x, y, k)
KNN.MORAN.function.R (ptsObj, 'ERA', 'R_LINK_NOD', 1)
KNN.MORAN.function.R (ptsObj, 'ERA', 'MED_BLK_PES', 1)
KNN.MORAN.function.R (ptsObj, 'ERA', 'MED_CULK', 1)
KNN.MORAN.function.R (ptsObj, 'ERA', 'MED_CUL_LE', 1)

# Neighborhood Mix metrics
# run KNN.MORAN.function.R (spatob, x, y, k)
KNN.MORAN.function.R (ptsObj, 'ERA', 'IJI', 1)
KNN.MORAN.function.R (ptsObj, 'ERA', 'PR', 1)
KNN.MORAN.function.R (ptsObj, 'ERA', 'SIDI', 1)
KNN.MORAN.function.R (ptsObj, 'ERA', 'R_RENT_OWN', 1)
KNN.MORAN.function.R (ptsObj, 'ERA', 'PRPWRKOUT', 1)
P.VALUES.function.R <- function(spatob, x, y, k){
# Research Part I
# Purpose: Determines mean, zscore, and pvalues for metrics
# and outputs to a .csv file.
# Requires the spdep library to be set
# Runs a sar model, then uses relevel to change the order of the factors
# presub, sub, postsub to get the p-values of each needed combination.
# Requires:
# spatob:  spatial dataframe object, created with readShapePoint function
# x:  dependent variable (e.g. ERA)
# y:  independent variable (e.g. MEDLOTSIZE)
# k:  number of nn to use for sar model
#
# create data frame from spatob
df.pts = as.data.frame(spatob)

# create neighbors list from spatob !!using knearneigh!!
knn.pts = knearneigh(spatob, k)

# create neighbors list from knn object
list.nb = knn2nb(knn.pts)

# Create proximities object with weight style W
listW.pts = nb2listw(list.nb, style = "W")

# Create p-value table

# run the model
sar.Model= spautolm(paste(y,'~',x),df.pts, listW.pts, family="SAR")
print(summary(sar.Model))

# Get the means
mu.post = sar.Model$fit$coefficients[[1]]; print(mu.post)
mu.pre = (sar.Model$fit$coefficients[[1]] + sar.Model$fit$coefficients[[2]]); print(mu.pre)
mu.sub = (sar.Model$fit$coefficients[[1]] + sar.Model$fit$coefficients[[3]]); print(mu.sub)

# create summary object & get the z scores and p values
sum.sar = summary(sar.Model)
z.presub.post = sum.sar$coef[2,3]; z.presub.post
p.presub.post = sum.sar$coef[2,4]; p.presub.post
z.sub.post = sum.sar$coef[3,3]; z.sub.post
p.sub.post = sum.sar$coef[3,4]; p.sub.post

## Use relevel to re-arrange the ERA factors, then run the model again
df.pts$ERA = relevel(df.pts$ERA, ref = "presub") # move presub up to intercept position
attach(df.pts)
sar.Model= spautolm(paste(y,'~',ERA),df.pts, listW.pts, family="SAR")
print(summary(sar.Model))
detach(df.pts)

# create summary object & get the z scores and p values
sum.sar = summary(sar.Model)
z.sub.presub = sum.sar$coef[3,3]; z.sub.presub
p.sub.presub = sum.sar$coef[3,4]; p.sub.presub

# create and populate the matrix
# multiply by 3 for Bonferroni correction
mat = matrix(0,1,13)
mat[1,1] = y
mat[1,2] = mu.pre
mat[1,3] = mu.sub
mat[1,4] = z.sub.presub
mat[1,5] = p.sub.presub * 3
mat[1,6] = mu.sub
mat[1,7] = mu.post
mat[1,8] = z.sub.post
mat[1,9] = p.sub.post * 3
mat[1,10] = mu.pre
mat[1,11] = mu.post
mat[1,12] = z.presub.post
mat[1,13] = p.presub.post * 3
# print(mat)

# write mat to text file
fName = "C:\PhD_Work\DATA\ANALYSIS\PART_I\STATS\OUT\PVALUES\PVALUES_KNN1.new.csv"
write.table(mat, file = fName, append = TRUE, col.names = TRUE, sep="",""
)}}

############################################################
# Research Part I
# Name: Run.PVALUES_1.R
# Purpose: Runs P.VALUES.function for 18 models.
# Used to determine mean, zscore, pvalues (for table 2-2 in ms)
############################################################
# Clean up
# list all vars, then remove all varaibles, then show all are gone
ls()
rm(list=ls(all.names=TRUE))
ls()

############################################################
# My Data
############################################################
# Set working directory
setwd("C:/PhD_Work/DATA/ANALYSIS/PART_I/STATS/SAR")

# Set libraries
library(spdep)

# read in shapefile to a Spatial Data frame Object
pts.shpObj = readShapePoints("DataPtsXYnew.shp", proj4string=CRS("+proj=utm +zone=12 +datum=NAD83"))

# Set working directory
setwd("C:/PhD_Work/DATA/ANALYSIS/PART_I/STATS/")
source("P.VALUES.function.R")

# Density metrics
# run P.VALUES.function.R (spatob, x, y, k)
P.VALUES.function.R (pts.shpObj, 'ERA', 'MEDLOTACRE', 2)
P.VALUES.function.R (pts.shpObj, 'ERA', 'HUTPERSQKM', 1)
P.VALUES.function.R (pts.shpObj, 'ERA', 'MEDNURMS', 1)
P.VALUES.function.R (pts.shpObj, 'ERA', 'POPPERHU', 1)
P.VALUES.function.R (pts.shpObj, 'ERA', 'POPPERSQKM', 1)
P.VALUES.function.R (pts.shpObj, 'ERA', 'MED_BLK_PE', 1)
P.VALUES.function.R (pts.shpObj, 'ERA', 'MED_CUL_LE', 1)

# Centrality metrics
# run P.VALUES.function.R (spatob, x, y, k)
P.VALUES.function.R (pts.shpObj, 'ERA', 'CUMMEAN', 1)
P.VALUES.function.R (pts.shpObj, 'ERA', 'FRKMEAN', 1)
P.VALUES.function.R (pts.shpObj, 'ERA', 'SCHMEAN', 1)
P.VALUES.function.R (pts.shpObj, 'ERA', 'BUSMEAN', 2)

# Accessibility metrics
# run P.VALUES.function.R (spatob, x, y, k)
P.VALUES.function.R (pts.shpObj, 'ERA', 'R_LINK_NOD', 1)
P.VALUES.function.R (pts.shpObj, 'ERA', 'R_LINK_CUL', 1)
P.VALUES.function.R (pts.shpObj, 'ERA', 'MED_CUL_LE', 1)

# Neighborhood Mix metrics
# run P.VALUES.function.R (spatob, x, y, k)
P.VALUES.function.R (pts.shpObj, 'ERA', 'IJI', 1)
P.VALUES.function.R (pts.shpObj, 'ERA', 'PR', 1)
P.VALUES.function.R (pts.shpObj, 'ERA', 'SIDI', 1)
P.VALUES.function.R (pts.shpObj, 'ERA', 'PRPWRKOUT', 1)
P.VALUES.function.R (pts.shpObj, 'ERA', 'R_RENT_OWN', 1)

STRIP.function.R <- function(spatob,x, y, k){
  # Research Part I
  # Purpose: Runs a knn SAR model and plots a stripchart.
  # to be run by the script Run.STRIPCHART_1nn.R
  # Requires the spdep library to be set
  #
  # Runs a SAR model then creates stripchart
  #
  # spatob: spatial dataframe object, created with readShapePoint function
  # x: dependent variable from the dataframe (e.g. ERA)
  # y: independent variable from the dataframe (e.g. MEDLOTSIZE)
  # k: number of nn to use for sar model
  #
  # create data frame from spatob
  df.pts = as.data.frame(spatob)
  # create neighbors list from spatob !!using knearneigh!!
  knn.pts = knearneigh(spatob, k)
  # create neighbors list from knn object
  list.nb = knn2nb(knn.pts)
  # Create proximities object with weight style W
  listW.pts = nb2listw(list.nb, style = "W")
  # create the SAR model
  sar.Model= spautolm(paste(y,'~',x,'+X_DIS+Y_DIS'),df.pts, listW.pts, family="SAR")
  print(sar.Model)
  print(summary(sar.Model))
  print(moran.test(resid(sar.Model), randomisation=FALSE, alternative="two.sided", listW.pts))
  # get the betas from the sar model
  beta.hat = sar.Model$fit$coefficients
  beta.hat
  # get the sigma hats from the sar model
  Sigma.hat = summary(sar.Model)$resvar
  Sigma.hat
  # create the contrast matrix
  vect = c(1,1,1,1,0,0,0,1,0)
  cont.mat = matrix(vect,3,3)
  cont.mat
  # create empty mu matrix for simulations
  mu = matrix(0,1000,3)
  # Simulation loop
  for (i in 1:1000){
    beta = as.vector(rmvnorm(1, beta.hat, Sigma.hat))
    mu[i,] = cont.mat%*%beta
  }
  # write mu to text file
  fName = "C:\PhD_Work\DATA\ANALYSIS\PART_I\STATS\OUT\STRIP\mu.csv"
  write.table(mu, file = fName, col.names = TRUE, sep="","")
}
SCRIPTS USED FOR PART II OF THE RESEARCH (CHAPTER 3)

Diag.function.R <- function(data, n){
  # Research Part I
  # Purpose: produces diagnostic graphs
  # Requires the spdep library to be set
  #
  # Creates LM, runs diagnostics
  # Parameters:
  # data: a data frame from which to create the LM
  # n: number of X variables in the model
  #
  # OLS Regression
  data = as.data.frame(data)
  ols.model = lm(data)
  sum = summary(ols.model)
  print(sum)

  ### Diagnostics (basic)
  ### Checking for constant variance (residuals X fitted values)
  ### Checking for normality (Q-Q plot)
  ### Checking for constant variance (standardized residuals X fitted values)
  ### Checking for influential observations (Cooks distance plot)
  print ("Running basic diagnostics...")
  x11()
  par(mfrow=c(2,2), mex=0.6)
  plot(ols.model)

  ### Diagnostics (additional)
  ### Checking for influential observations
  ### DFFITS --> how an obs impacts the whole model (R^2)
  ### DFBETAS --> how an obs impacts the Beta coefficients
  print ("New page. Running additional diagnostics...")
  x11()
  par(mfrow=c(2,2), mex=0.6)
  plot(rstandard(ols.model),main = "Obs. vs Standarized Residuals", xlab = "Index")
  abline (h=0, col="grey")
  plot(rstudent(ols.model),main = "Obs. vs Studentized Residuals", xlab = "Index.")
  abline (h=0, col="grey")
  plot(dffits(ols.model), type ="l",main = "Obs. vs DFFITS", xlab = "Index")
  abline (h=0, col="grey")
  matplot(dfbetas(ols.model), type="l", col="black",main = "Obs. vs DFFITS", xlab = "Index")
  lines(sqrt(cooks.distance(ols.model)), lwd=2)
  abline (h=0, col="grey")
  par(mfrow=c(1,1), mex=1)

  print ("New page. Partial regression Plots...")
  x11()

  ### PARTIAL REGRESSION PLOTS: X RESIDUALS AGAINST MODEL (Y) RESIDUALS
  ### Provides intuition of meaning of Beta coefficients
  ### Useful to detect outliers, as well as check relationship of Xs on Y
  library(car) # load the car package to av.plot
  par(mfrow=c(4,3), mex=0.6)
  avplots(ols.model)

  print ("New page. Partial Residual Plots...")
  x11()

  ### PARTIAL RESIDUAL PLOTS: X RESIDUALS AGAINST MODEL RESIDUALS
  ### Check for non-linearity between Xs on Y
  library(faraway) # load the faraway package to use prplot
  par(mfrow=c(4,3), mex=0.6)
  for (i in 1:n) {
    prplot(ols.model, i)
    title(main = "partial residuals plot")
  } # end faraway function

  print ("New page. Model residuals vs. X variables")
par(mfrow=c(4,3), mar=0.6)
n = n+1
for (i in 2:n) {
plot(data[,i], residuals(ols.model), xlab="X var", main="Resid. v. X var")
abline(h=0, col="grey")
} # end residuals v. X vars

# Research Part II
# Name: Run.Diag.R
# Purpose: Runs Diag.function for 15 models.
# Used to run regression diagnostics

# Clean up
# list all vars, then remove all variables, then show all are gone
ls()
rm(list=ls(all.names=TRUE))
ls()

# My Data
# Set working directory
setwd("C:/PhD_Work/DATA/ANALYSIS/PART_II/DATA/COMBINE")
# Set libraries
library(spdep)

# read in shapefile to a Spatial Data frame Object
pts.shpObj = readShapePoints("DataPts7.shp", proj4string=CRS("+proj=utm +zone=12 +datum=NAD83"))

# create dataframe from spatial object
df.pts = as.data.frame(pts.shpObj)

# Set working directory
setwd("C:/PhD_Work/DATA/ANALYSIS/PART_II/STATS/")
source("Diag.function.R")
source("Moran.function.R")

# Socio Economic demographic
# Not transformed (except AGE)
AGE = df.pts[,4]^0.5
AVHS = df.pts[,28]
MAGE = df.pts[,27]
MINC = df.pts[,20]
PNWT = (df.pts[,37]+df.pts[,38]+df.pts[,39]+df.pts[,40]+df.pts[,41])
PGTHS = (df.pts[,23]+df.pts[,24]+df.pts[,25])
Xdat = cbind(AVHS,MAGE,MINC,PNWT,PGTHS)
Ydat = df.pts[,3]
data = as.data.frame(cbind(Ydat,Xdat,AGE))
datSED = Xdat
lm.model = lm(Ydat~AGE+AVHS+MINC+MAGE+PNWT+PGTHS, data=data)
summary(lm.model)

# urban structure
# Not transformed (except AGE)
AGE = (df.pts[,4]^0.5)
MDLOT = df.pts[,8]
HSSIZ = df.pts[,55]
MBLKP = df.pts[,14]
LNKND = df.pts[,15]
IJI = df.pts[,16]

Ydat = df.pts[,3]
Xdat = cbind(MDLOT, HSSIZ, MBLKP, LNKND, IJI)
data = as.data.frame(cbind(Ydat, Xdat, AGE))
datURB = Xdat
lm.model = lm(Ydat~MDLOT+HSSIZ+MBLKP+LNKND+IJI+AGE, data=data)
summary(lm.model)
## run the function
Diag.function.R(data, 6)
Moran.function.R('URB', data, pts.shpObj)

############################################################
## physical landscape
AGE = (df.pts[,4]^.5)
PRCP = df.pts[,44]
MAWS = df.pts[,53]
MAWT = df.pts[,6]
MSTH = df.pts[,46]
MWST = df.pts[,47]

Ydat = df.pts[,3]
Xdat = cbind(PRCP, MAWS, MAWT, MSTH, MWST)
data = as.data.frame(cbind(Ydat, Xdat, AGE))
datPHY = Xdat
lm.model = lm(Ydat~PRCP+MAWS+MAWT+MSTH+MWST+AGE, data=data)
summary(lm.model)
## run the function
Diag.function.R(data, 6)
Moran.function.R('PHY', data, pts.shpObj)

############################################################
## full model
Ydat = df.pts[,3]
data = as.data.frame(cbind(Ydat, datSED, datURB, datPHY, AGE))

lm.model =
lm(Ydat~AVHS+MINC+MAGE+PNWT+PGTHS+PRCP+MDLOT+HSSIZ+MBLKP+LNKND+IJI+MAWS+MAWT+MSTH+MWST+AG E, data=data)
s

summary(lm.model)
## run the function
Diag.function.R(data, 17)

SimpleSlope.function.R <- function(data, y, Xvar, prox, vName){
  # Research Part II
  # Purpose: Creates simple slope table
  # Parameters:
  # data: data frame
  # y: dependent variable (PercentTCC)
  # Xvar: x-variable string
  # prox: proximith matrix (listW.pts)
  # vName: variable name as string
  # create output matrix
  out.mat = matrix(0,1,13)
  # send data to out matrix
  out.mat[,1] = vName
  # 15 years old
  Zcv = data$AGE - 15;data2 = cbind(data, Zcv)
g = spautolm(paste(y, '~', Xvar), data2, prox, tol.solve=1e-20)
sum.g = summary(g)
out.mat[,2] = sum.g$Coef[2,1]
out.mat[,3] = sum.g$Coef[2,1]
out.mat[,4] = sum.g$Coef[2,4]
print(summary(g))}
# 55 years old
Zcv = data$AGE - 55;data2 = cbind(data,Zcv)
g = spautolm(paste(y,'~',Xvar),data2,prox,tol.solve=1e-20)
sum.g = summary(g)
out.mat[1,5] = sum.g$Coef[2,1]
out.mat[1,6] = sum.g$Coef[2,3]
out.mat[1,7] = sum.g$Coef[2,4]
print(summary(g))

# 95 years old
Zcv = data$AGE - 95;data2 = cbind(data,Zcv)
g = spautolm(paste(y,'~',Xvar),data2,prox,tol.solve=1e-20)
sum.g = summary(g)
out.mat[1,8] = sum.g$Coef[2,1]
out.mat[1,9] = sum.g$Coef[2,3]
out.mat[1,10] = sum.g$Coef[2,4]
print(summary(g))

# Interaction estimate, z-score & p-value (same for all levels)
out.mat[1,11] = sum.g$Coef[4,1]
out.mat[1,12] = sum.g$Coef[4,3]
out.mat[1,13] = sum.g$Coef[4,4]

# write mat to text file
fName = "C:/PhD_Work/DATA/ANALYSIS/PART_II/STATS/OUT/simpleslope2X.csv"
write.table(out.mat, file = fName, append=TRUE, col.names = FALSE, sep="", )

########################################
# Research Part II
# Name: Run.SimpleSlope.R
# Purpose: Determines slopes and p-values for 15 models (table 3-2)
########################################
# Clean up
# list all vars, then remove all varaibles, then show all are gone
ls()
rm(list=ls(all.names=TRUE))
ls()

########################################
# SET UP DATA
########################################
# Set working directory
setwd("C:/PhD_Work/DATA/ANALYSIS/PART_II/DATA/COMBINE")

# Set libraries
library(spdep)

# read in shapefile to a Spatial Data frame Object
pts.shpObj = readShapePoints("DataPts9.shp",
proj4string=CRS("+proj=utm +zone=12 +datum=NAD83"))

# create dataframe from spatial object
df.pts = as.data.frame(pts.shpObj)

# Set up spatial neighbors
# create data frame from spatob
df.pts = as.data.frame(pts.shpObj)

# create proximity matrix based on 1 nn
knn.pts = knearneigh(pts.shpObj, 1)
list.nb = knn2nb(knn.pts)
listW.pts = nb2listw(list.nb, style = "W")

### SOCIAL ECONOMIC DEMOGRAPHIC VARIABLES

AGE = df.pts[,4]
AVHS = df.pts[,28]
MAGE = df.pts[,27]
MINC = df.pts[,20]/10000
PNWT = (df.pts[,37]+df.pts[,38]+df.pts[,39]+df.pts[,40]+df.pts[,41])
PGTHS = (df.pts[,23]+df.pts[,24]+df.pts[,25])

# Dependent Variable & Data Matrix
PercentTCC = df.pts[,3]; data = as.data.frame(cbind(PercentTCC, AGE, AVHS, MAGE, MINC, PNWT, PGTHS))

## Simple slope analysis per pages 18 & 19 of Aiken and West (1991)
# Set working directory
setwd("C:/PhD_Work/DATA/ANALYSIS/PART_II/STATS/")
source("SimpleSlope.function.R")

# Run SimpleSlope.function.R
SimpleSlope.function.R(data,'PercentTCC','MINC+Zcv+MINC*Zcv',listW.pts, 'Median Income')
SimpleSlope.function.R(data,'PercentTCC','AVHS+Zcv+AVHS*Zcv',listW.pts, 'Ave. Household Size')
SimpleSlope.function.R(data,'PercentTCC','PNWT+Zcv+PNWT*Zcv',listW.pts, 'Non-Whie Pop.')
SimpleSlope.function.R(data,'PercentTCC','PGTHS+Zcv+PGTHS*Zcv',listW.pts, 'High School Grads."

### URBAN STRUCTURE VARIABLES ###################################
AGE = df.pts[,4]
LNKND = df.pts[,56]
IJI = df.pts[,16]
MDLOT = df.pts[,8]
RSDEN = df.pts[,57]
MBLKP = df.pts[,14]

# Dependent Variable and create data matrix
PercentTCC = df.pts[,3]; data = as.data.frame(cbind(PercentTCC, AGE, MDLOT, RSDEN, MBLKP, LNKND, IJI))

# Run SimpleSlope.function.R
SimpleSlope.function.R(data,'PercentTCC','LNKND+Zcv+LNKND*Zcv',listW.pts, 'Street Connectivity')
SimpleSlope.function.R(data,'PercentTCC','IJI+Zcv+IJI*Zcv',listW.pts, 'Land Use Mix')
SimpleSlope.function.R(data,'PercentTCC','MDLOT+Zcv+MDLOT*Zcv',listW.pts, 'Median Lot Size')
SimpleSlope.function.R(data,'PercentTCC','RSDEN+Zcv+RSDEN*Zcv',listW.pts, 'Residential Street Density')

### PHYSICAL LANDSCAPE VARIABLES ###################################
AGE = df.pts[,4]
PRCP = df.pts[,44]
MAWS = df.pts[,53]
MAWT = df.pts[,6]/1000
MSTH = df.pts[,46]
MWST = df.pts[,47]

# Dependent Variable and create data matrix
PercentTCC = df.pts[,3]; data = as.data.frame(cbind(PercentTCC, AGE, PRCP, MAWS, MAWT, MSTH, MWST))

# Run SimpleSlope.function.R
SimpleSlope.function.R(data,'PercentTCC','PRCP+Zcv+PRCP*Zcv',listW.pts, 'Mean Annual Precip.')
SimpleSlope.function.R(data,'PercentTCC','MAWS+Zcv+MAWS*Zcv',listW.pts, 'Avail. Water Storage')
SimpleSlope.function.R(data,'PercentTCC','MAWT+Zcv+MAWT*Zcv',listW.pts, 'Distance to Streams/Canals')
SimpleSlope.function.R(data,'PercentTCC','MWST+Zcv+MWST*Zcv',listW.pts, 'Aspect (Westness)')
SimpleSlope.function.R(data,'PercentTCC','MSTH+Zcv+MSTH*Zcv',listW.pts, 'Aspect (Southness)')
SimplePlot.function.R <- function(sarmod, prox, var, var2, xlab, title) {
  # Research Part II
  # Purpose: Creates simple plots
  # Parameters:
  # sarmod: a sar model from errorsarlm function
  # prox: neighbors list (e.g. listW.pts)
  # var: variable as object to create the plot from
  # var2: variable name as text
  # title: title of graph
  # SET UP MATRIX FOR PREDICTED VALUES
  z1910 = (2005-1910) # ~ 95 yrs
  z1950 = (2005-1950) # ~ 55 yrs
  z1990 = (2005-1990) # ~ 15 yrs
  v = var
  mat = matrix(nrow = 6, ncol = 2)
  mat[1,1] = z1990; mat[3,1] = z1990
  mat[3,1] = z1950; mat[4,1] = z1950
  mat[5,1] = z1910; mat[6,1] = z1910
  mat[1,2] = min(v); mat[2,2] = max(v)
  mat[3,2] = min(v); mat[4,2] = max(v)
  mat[5,2] = min(v); mat[6,2] = max(v)
  # ADD COLUMN NAMES AND MAKE DATAFRAME
  colnames(mat) = c("AGE", paste(var2))
  nDat = as.data.frame(mat)
  # DETERMINE PREDICTED VALUES FOR PERCENTTCC BASED ON THE SARMODEL
  PercentTCC = as.vector(predict.sarlm(sarmod, nDat, prox))
  # CREATE DATA MATRIX FOR PLOT
  xDat = cbind(nDat, PercentTCC)
  # MAKE THE PLOT
  plot(PercentTCC~xDat[,2], xDat, pch=19, ylab="Percent Tree Canopy", xlab=(paste(xlab)),
       ylim=(c(20,80)))
  title(main=paste(title), cex.main=1.1)
  x2 = xDat[1:2,]; lines(x2[,2], x2$PercentTCC, lty=1, lwd=1)
  x4 = xDat[3:4,]; lines(x4[,2], x4$PercentTCC, lty=15, lwd=1)
  x6 = xDat[5:6,]; lines(x6[,2], x6$PercentTCC, lty=14, lwd=1)
}

########################################
# Research Part II
# Name: Run.SimpleSlope.R
# Purpose: Creates simple slope plots for 15 models (figure 3-2)
########################################
# Clean up
# list all vars, then remove all varaibles, then show all are gone
ls()
rm(list=ls(all.names=TRUE))
ls()
# SET UP DATA
# Set working directory
setwd("C:/PhD_Work/DATA/ANALYSIS/PART_II/DATA/COMBINE")
# Set libraries
library(spdep)
# read in shapefile to a Spatial Data frame Object
pts.shpObj = readShapePoints("DataPts9.shp",
                            proj4string=CRS("+proj=utm +zone=12 +datum=NAD83"))
# create dataframe from spatial object
df.pts = as.data.frame(pts.shpObj)

# Set up spatial neighbors
# create data frame from spatob
df.pts = as.data.frame(pts.shpObj)

# create proximity matrix based on 1 nn
knn.pts = knearneigh(pts.shpObj, 1)

# create neighbors list from knn object
list.nb = knn2nb(knn.pts)

# create proximities object with weight style W
listW.pts = nb2listw(list.nb, style = "W")

# Set working directory
setwd("C:/PhD_Work/DATA/ANALYSIS/PART_II/STATS/")
source("SimplePlot.function.R")
source("SimplePlot.function1.R")

### SET UP THE WINDOWS CANVAS ###################################
windows(10.0,12.0)
par(mfrow=c(5,3), mex=0.6)

### SOCIAL ECONOMIC DEMOGRAPHIC VARIABLES ###########################
AGE = df.pts[,4]
AVHS = df.pts[,28]
MAGE = df.pts[,27]
MINC = df.pts[,20]/10000
PNWT = (df.pts[,37]+df.pts[,38]+df.pts[,39]+df.pts[,40]+df.pts[,41])
PGTHS = (df.pts[,23]+df.pts[,24]+df.pts[,25])

# Dependent Variable & Data Matrix
PercentTCC = df.pts[,3]; data = as.data.frame(cbind(PercentTCC,AGE,AVHS,MAGE,MINC,PNWT,PGTHS))

#### Run SimplePlot.function.R ###############################
g = errorsarlm(PercentTCC~AGE+MINC+AGE*MINC,data,listW.pts,tol.solve=1e-20);summary(g)
SimplePlot.function1.R(g,listW.pts,MINC,'MINC','Income ($10k)','Median Income')

g = errorsarlm(PercentTCC~AGE+AVHS+AGE*AVHS,data,listW.pts, tol.solve=1e-20);summary(g)
SimplePlot.function.R(g,listW.pts,AVHS,'AVHS','Houshold Size', 'Ave. Houshold Size')

# Dependent Variable and create data matrix
PercentTCC = df.pts[,3]; data = as.data.frame(cbind(PercentTCC,AGE,MAGE,MINC,PNWT,PGTHS))

### URBAN STRUCTURE VARIABLES ####################################
AGE = df.pts[,4]
LNKND = df.pts[,15]
#LNKCL = df.pts[,56]
#IJI = df.pts[,16]
MDLOT = df.pts[,8]
#HSSIZ = df.pts[,55]
RSDEN = df.pts[,57]
MBLKP = df.pts[,14]
#HPRSQ = df.pts[,7]

# Dependent Variable and create data matrix
PercentTCC = df.pts[,3]; data = as.data.frame(cbind(PercentTCC,AGE,MDLOT,MBLKP,LNKND,LNKCL,IJI))
```
### Run SimplePlot.function.R (sarmod,prox,var)############################
g = errorsarlm(PercentTCC~AGE+LNKND+AGE*LNKND,data,listW.pts,tol.solve=1e-20);summary(g)
SimplePlot.function.R(g,listW.pts,LNKND,'LNKND','Link:Node Ratio (more-to-less)', 'Street Connectivity')

# g = errorsarlm(PercentTCC~AGE+LNKCL+AGE*LNKCL,data,listW.pts,tol.solve=1e-20);summary(g)
# SimplePlot.function.R(g,listW.pts,LNKCL,'LNKCL','Streets:Cul-de-sac (more-to-less)', 'Street Connectivity')

g = errorsarlm(PercentTCC~AGE+IJI+AGE*IJI,data,listW.pts,tol.solve=1e-20);summary(g)
SimplePlot.function.R(g,listW.pts,IJI,'IJI','IJI Index (less-to-more)', 'Land Use Mix')

g = errorsarlm(PercentTCC~AGE+MDLOT+AGE*MDLOT,data,listW.pts,tol.solve=1e-20);summary(g)
SimplePlot.function.R(g,listW.pts,MDLOT,'MDLOT','Size (acres)', 'Median Lot Size')

# g = errorsarlm(PercentTCC~AGE+HSSIZ+AGE*HSSIZ,data,listW.pts,tol.solve=1e-20);summary(g)
# SimplePlot.function.R(g,listW.pts,HSSIZ,'HSSIZ','Size (sq ft)','Median House Size')

g = errorsarlm(PercentTCC~AGE+RSDEN+AGE*RSDEN,data,listW.pts,tol.solve=1e-20);summary(g)
SimplePlot.function.R(g,listW.pts,RSDEN,'RSDEN','Streets/km^2','Res. Street Density')

g = errorsarlm(PercentTCC~AGE+MBLKP+AGE*MBLKP,data,listW.pts,tol.solve=1e-20);summary(g)
SimplePlot.function.R(g,listW.pts,MBLKP,'MBLKP','Length (ft)','Med. Blk. Perimeter')

### PHYSICAL LANDSCAPE VARIABLES ##########################################
AGE = df.pts[,4]
PRCP = df.pts[,44]
MAWS = df.pts[,53]
MAWT = df.pts[,6]/1000
MSTH = df.pts[,46]
MWST = df.pts[,47]

# Dependent Variable and create data matrix
PercentTCC = df.pts[,3]; data = as.data.frame(cbind(PercentTCC,AGE,PRCP,MAWS,MAWT,MSTH,MWST))

### Run SimplePlot.function.R (sarmod,prox,var)############################
g = errorsarlm(PercentTCC~AGE+PRCP+AGE*PRCP,data,listW.pts,tol.solve=1e-20);summary(g)
SimplePlot.function.R(g,listW.pts,PRCP,'PRCP','Precip. (mm)','Mean Annual Precip.')

g = errorsarlm(PercentTCC~AGE+MAWS+AGE*MAWS,data,listW.pts,tol.solve=1e-20);summary(g)
SimplePlot.function.R(g,listW.pts,MAWS,'MAWS','Volume (cm)','Avail. Water Storage(Soil)')

g = errorsarlm(PercentTCC~AGE+MAWT+AGE*MAWT,data,listW.pts,tol.solve=1e-20);summary(g)
SimplePlot.function.R(g,listW.pts,MAWT,'MAWT','Distance (Km)','Dist. to Streams/Canals')

g = errorsarlm(PercentTCC~AGE+MWST+AGE*MWST,data,listW.pts,tol.solve=1e-20);summary(g)
SimplePlot.function.R(g,listW.pts,MWST,'MWST','Degrees (180=west azimuth)','Aspect (Westness)')

g = errorsarlm(PercentTCC~AGE+MSTH+AGE*MSTH,data,listW.pts,tol.solve=1e-20);summary(g)
SimplePlot.function.R(g,listW.pts,MSTH,'MSTH','Degrees (180=south azimuth)','Aspect (Southness)')
```

SAR.function.R <- function(data, prox, formula){
# Research Part II
# Purpose: output SAR information to do MMI
# Requires the spdep library to be set
#
# Runs a SAR model with the following user defined parameters:
#
# data: dataframe object, created from spatial object
# prox: proximities object
# formula: model formula as string

# run the SAR model
sar.Model = spautolm(paste(formula), data, prox, family="SAR")
print(summary(sar.Model))
AIC =(-2*sar.Model$LL) + (2*sar.Model$parameters)
pseudoR = cor(sar.Model$fit$fitted.values, data[,1])^2
# print(pseudoR)
# print(sar.Model$lambda)
# print(sar.Model$LL)
# print(sar.Model$parameters)
print (length(as.vector(sar.Model$fit$coefficients)) - 1)
# print(AIC)

# create matrix
out.mat = matrix(0,1,6)
# send values to matrix
out.mat[1,1] = formula
out.mat[1,2] = sar.Model$lambda
out.mat[1,3] = sar.Model$LL
out.mat[1,4] = sar.Model$parameters
#out.mat[1,4] = length(as.vector(sar.Model$fit$coefficients)) - 1
out.mat[1,5] = AIC
out.mat[1,6] = pseudoR

# write out.mat to text file
fName = "C:\PhD_Work\DATA\ANALYSIS\PART_II\STATS\OUT\MMItable.csv"
write.table(out.mat, file = fName, append=TRUE, col.names = FALSE, sep="",")
}

******************************************************************************
# Research Part II
# Name: Run.SAR.MMI.Full.R
# Purpose: Runs SAR.function.R
# Used to produced data for Table 3-3
******************************************************************************
# Clean up
# list all vars, then remove all varaibles, then show all are gone
ls()
rm(list=ls(all.names=TRUE))
ls()
******************************************************************************
# My Data
******************************************************************************
# Set working directory
setwd("C:\PhD_Work\DATA\ANALYSIS\PART_II\DATA\COMBINE")
# Set libraries
library(spdep)

# read in shapefile to a Spatial Data frame Object
pts.shpObj = readShapePoints("DataPts9.shp", proj4string=CRS("*proj=utm +zone=12 +datum=NAD83*"))

# Set up spatial neighbors
# create data frame from spatob
df.pts = as.data.frame(pts.shpObj)
# create proximity matrix based on 1 nn
knn.pts = knearneigh(pts.shpObj, 1)

# create neighbors list from knn object
list.nb = knn2nb(knn.pts)

# create proximities object with weight style W
listW.pts = nb2listw(list.nb, style = "W")

### SOCIAL ECONOMIC DEMOGRAPHIC VARIABLES ###########################
NHAGE = df.pts[,4]
MDINC = df.pts[,20]/10000
AVHSZ = df.pts[,28]
POAGE = df.pts[,27]
PNWHT = (df.pts[,37]+df.pts[,38]+df.pts[,39]+df.pts[,40]+df.pts[,41])
PGTHS = (df.pts[,23]+df.pts[,24]+df.pts[,25])

### SOCIAL ECONOMIC DEMOGRAPHIC VARIABLES ###########################
x1 = NHAGE
x2 = MDINC
x3 = AVHSZ
x4 = POAGE
x5 = PNWHT
x6 = PGTHS

### URBAN STRUCTURE VARIABLES ###################################
STCON = df.pts[,56]
LUMIX = df.pts[,16]
MDLOT = df.pts[,8]
RSDEN = df.pts[,57]
MBLKP = df.pts[,14]

x7 = STCON
x8 = LUMIX
x9 = MDLOT
x10 = RSDEN
x11 = MBLKP

### PHYSICAL LANDSCAPE VARIABLES ##################################
MPRCP = df.pts[,44]
AVWTS = df.pts[,53]
DSTSC = df.pts[,6]/1000
ASWST = df.pts[,47]
ASSTH = df.pts[,46]

x12 = MPRCP
x13 = AVWTS
x14 = DSTSC
x15 = ASWST
x16 = ASSTH

Xdat = cbind(x1,x2,x3,x4,x5,x6,x7,x8,x9,x10,x11,x12,x13,x14,x15,x16)
Ydat = df.pts[,3]
data = as.data.frame(cbind(Ydat,Xdat))

# Set working directory
setwd("C:/PhD_Work/DATA/ANALYSIS/PART_II/STATS/")
source("SAR.function.R")

### Run SAR.function.R (data,prox,formula)##########################
# SED separate
#SAR.function.R(data,listW.pts,'Ydat~x1')
#SAR.function.R(data,listW.pts,'Ydat~x2')
#SAR.function.R(data,listW.pts,'Ydat~x1+x2*x1')
#SAR.function.R(data,listW.pts,'Ydat~x3')
#SAR.function.R(data,listW.pts,'Ydat~x4')
#SAR.function.R(data,listW.pts,'Ydat~x5')
#SAR.function.R(data,listW.pts,'Ydat~x1+x5*x1')
#SAR.function.R(data,listW.pts,'Ydat~x6')
# URB separate
#SAR.function.R(data,listW.pts,'Ydat~x7')
#SAR.function.R(data,listW.pts,'Ydat~x8')
#SAR.function.R(data,listW.pts,'Ydat~x1+x8*x1')
#SAR.function.R(data,listW.pts,'Ydat~x9')
#SAR.function.R(data,listW.pts,'Ydat~x10')
#SAR.function.R(data,listW.pts,'Ydat~x11')
# PHY separate
#SAR.function.R(data,listW.pts,'Ydat~x12')
#SAR.function.R(data,listW.pts,'Ydat~x13')
#SAR.function.R(data,listW.pts,'Ydat~x1+x13*x1')
#SAR.function.R(data,listW.pts,'Ydat~x14')
#SAR.function.R(data,listW.pts,'Ydat~x15')
#SAR.function.R(data,listW.pts,'Ydat~x16')
#SAR.function.R(data,listW.pts,'Ydat~x1+x16*x1')
# SED w/out age
SAR.function.R(data,listW.pts,'Ydat~x2+x3+x4+x5+x6')
# SED w age
#SAR.function.R(data,listW.pts,'Ydat~x1+x2+x3+x4+x5+x6')
# SED w age interactions
SAR.function.R(data,listW.pts,'Ydat~x1+x2+x3+x4+x5+x6+x2*x1+x5*x1')
# URB w/out age
SAR.function.R(data,listW.pts,'Ydat~x7+x8+x9+x10+x11')
# URB w age
#SAR.function.R(data,listW.pts,'Ydat~x1+x7+x8+x9+x10+x11')
# URB w age interactions
SAR.function.R(data,listW.pts,'Ydat~x1+x7+x8+x9+x10+x11+x8*x1')
# PHY w/out age
SAR.function.R(data,listW.pts,'Ydat~x12+x13+x14+x15+x16')
# PHY w age
#SAR.function.R(data,listW.pts,'Ydat~x1+x12+x13+x14+x15+x16')
# PHY w age interactions
SAR.function.R(data,listW.pts,'Ydat~x1+x12+x13+x14+x15+x16+x13*x1+x15*x1+x16*x1')
# FULL w/out age interactions
SAR.function.R(data,listW.pts,'Ydat~x1+x2+x3+x4+x5+x6+x7+x8+x9+x10+x11+x12+x13+x14+x15+x16')
# FULL w age interactions
SAR.function.R(data,listW.pts,'Ydat~x1+x2+x3+x4+x5+x6+x7+x8+x9+x10+x11+x12+x13+x14+x15+x16')
# FULL w age interactions
SAR.function.R(data,listW.pts,'Ydat~x1+x2+x3+x4+x5+x6+x7+x8+x9+x10+x11+x12+x13+x14+x15+x16+x2*x1+x5*x1+x8*x1+x13+x15*x1+x16*x1')
# Research Part III
# Name: Tree_Lawn_Graph.R
# Purpose: Creates plots for tree canopy and grass cover (figure 3)
# Also produces regression coefficients for OLS and SAR models

# Clean up
# list all vars, then remove all variables, then show all are gone
ls()
rm(list=ls(all.names=TRUE))
ls()

# Set up Data
# Set working directory
setwd("C:/PhD_Work/DATA/ANALYSIS/PART_III/DATA/PREDICT/")

# Set libraries
library(spdep)

# read in shapefile to a Spatial Data frame Object
pts.shpObj = readShapePoints("DataPts9Prt3.shp", proj4string=CRS("+proj=utm +zone=12 +datum=NAD83"))

# create dataframe from spatial object
df.pts = as.data.frame(pts.shpObj)

# Set up spatial neighbors
# create proximity matrix based on 1 nn
knn.pts = knearneigh(pts.shpObj, 1)

# create neighbors list from knn object
list.nb = knn2nb(knn.pts)

# create proximities object with weight style W
listW.pts = nb2listw(list.nb, style = "W")

data = cbind(df.pts$AGE, df.pts$PerTrsh)
colnames(data) = c("AGE","PercentTCC")
data = as.data.frame(data)

### GRAPHING THE POLYNOMIAL

# Set up variables
MEANTCC = data$PercentTCC
AGE = data$AGE
YRBLT = abs(data$AGE - 2006)

### LM
lm.Age <- lm(MEANTCC~AGE+I(AGE^2))
lm.Age <- lm(MEANTCC~AGE+I(AGE^2)-1) # force through intercept!
summary(lm.Age)

### SAR
sar.Model = spautolm(MEANTCC~AGE+I(AGE^2)-1, data, listW.pts, family="SAR")
sar.Model = spautolm(MEANTCC~AGE+I(AGE^2), data, listW.pts, family="SAR")
summary(sar.Model)
pseudoR = cor(sar.Model$fit$fitted.values, data[,2])^2
pseudoR

### SET up for Graph/figure
windows(8.5, 4.25)
windows(3.5, 7.5)
par(mfrow=c(2,1), mex=0.75)

## CREATE the GRAPH
pred.frame = data.frame(AGE=seq(0,110,1))
#predict(lm.Age, interval="pred", newdata=pred.frame)

pp = predict(lm.Age, newdata=pred.frame, interval="pred")
pc = predict(lm.Age, newdata=pred.frame, interval="conf")
plot(AGE, MEANTCC, ylim=c(0,70), xlim=rev(c(0,110)), cex = .6, cex.lab = .7, cex.axis = .7, ylab="Percent Tree Canopy", xlab="Neighborhood Age (yrs)", font=1, pch=20)
title(main=list("Tree Canopy ~ Neighborhood Age",font=2, cex = .85))
matlines(pred.frame$AGE, pp, lty=c(1,2,2), col = "black")
matlines(pred.frame$AGE, pc, lty=c(1,3,3), col = "black")

# Legend
linetypes = c("Fitted Line","Prediction Bands","Confidence Bands")
legend("bottomleft", legend=linetypes, lty=1:3, cex=.7)
text(25,66,"PercentTC-Age+Age^2", font=1, cex=-.65)

text("(2005)", side = 1, at=c(0), padj=3, col=1,font=2,cex=.65)
text("(1985)", side = 1, at=c(20),padj=3,col=1,font=2,cex=.65)
text("(1965)", side = 1, at=c(40),padj=3,col=1,font=2,cex=.65)
text("(1945)", side = 1, at=c(60),padj=3,col=1,font=2,cex=.65)
text("(1925)", side = 1, at=c(80),padj=3,col=1,font=2,cex=.65)
text("(1905)", side = 1, at=c(100),padj=3,col=1,font=2,cex=.65)

# Find mean lwr confidence band and upr conf. band
pc = data.frame(pc)
meanLwr= mean(pc$fit-pc$lwr)
meanUpr= mean(pc$fit-pc$upr)

# Lawn Graph

# GRAPHING THE POLYNOMIAL

# Set up variables
MEANTCC = data$PercentLawn
AGE = data$AGE
YRBLT = abs(data$AGE - 2006)

### LM
lm.Age <-lm(MEANTCC~AGE+I(AGE^2))
summary(lm.Age)

### SAR
sar.Model = spautolm(MEANTCC~AGE+I(AGE^2),data, listW.pts, family="SAR")
summary(sar.Model)
pseudoR = cor(sar.Model$fit$fitted.values, data[,2])^2
pseudoR

# CREATE the GRAPH
pred.frame = data.frame(AGE=seq(0,110,1))
#predict(lm.Age, interval="pred", newdata=pred.frame)

pp = predict(lm.Age, newdata=pred.frame, interval="pred")
pc = predict(lm.Age, newdata=pred.frame, interval="conf")
plot(AGE, MEANTCC, ylim=c(0,70), xlim=rev(c(0,110)), cex = .6, cex.lab = .7, cex.axis = .7, ylab="Percent Grass Cover", xlab="Neighborhood Age (yrs)", font=1, pch=20)
title(main=list("Grass Cover ~ Neighborhood Age",font=2, cex = .85))
matlines(pred.frame$AGE, pp, lty=c(1,2,2), col = "black")
matlines(pred.frame$AGE, pc, lty=c(1,3,3), col = "black")

# Legend
linetypes = c("Fitted Line","Prediction Bands","Confidence Bands")
legend("topleft", legend=linetypes, lty=1:3, cex=.7)
text(80,50,"PercentGC~Age+Age^2", font=1,cex=.65)
text(80,45,"Psuedo R-sqr 0.82", font=1, cex =.65)
mtext("(2005)", side = 1, at=c(0), padj=3, col=1,font=2,cex =.65)
mtext("(1985)", side = 1, at=c(20),padj=3,col=1,font=2,cex =.65)
mtext("(1965)", side = 1, at=c(40),padj=3,col=1,font=2,cex =.65)
mtext("(1945)", side = 1, at=c(60),padj=3,col=1,font=2,cex =.65)
mtext("(1925)", side = 1, at=c(80),padj=3,col=1,font=2,cex =.65)
mtext("(1905)", side = 1, at=c(100),padj=3,col=1,font=2,cex =.65)

# Find mean lwr confidence band and upr conf. band
pc = data.frame(pc)
meanLwr = mean(pc$fit - pc$lwr)
meanLwr
meanUpr = mean(pc$fit - pc$upr)
meanUpr

# Research Part III
# Name: PredictIntervalR
# Purpose: Determines pseudo prediction intervals for SAR model

# Clean up
# list all vars, then remove all variables, then show all are gone
ls()
rm(list=ls(all.names=TRUE))
ls()

# Set up Data
# Set working directory
setwd("C:/PhD_Work/DATA/ANALYSIS/PART_III/DATA/PREDICT/")

# Set libraries
library(spdep); library(mvtnorm); library(geoR)

# read in shapefile to a Spatial Data frame Object
pts.shpObj = readShapePoints("DataPts9Prt3.shp",
proj4string=CRS("+proj=utm +zone=12 +datum=NAD83"))

data = cbind(df.pts$AGE, df.pts$PerTrsh)
colnames(data) = c("AGE","PercentTCC")
data = as.data.frame(data)

### Run SAR model
# Set up variables
MEANTCC = data$PercentTCC
AGE = data$AGE

## STEP 1, SAR
# sar.Model = spautolm(MEANTCC~AGE+I(AGE^2),data, listW.pts, family="SAR")
sar.Model = errorsarlm(PercentTCC~AGE+I(AGE^2),data, listW.pts,tol.solve=1e-20)

# get the betas from the sar model
beta.hat = sar.Model$fit$coefficients # for spautolm
beta.hat = sar.Model$coefficients # for errorsarlm
beta.hat

# get proximity matrix from knearneigh object
W = as.matrix(as_dgRMatrix_listw(listW.pts))

# get estimated variance component (s2)
# s2 = sar.Model$fit$s2 # for spautolm
s2 = sar.Model$s2 # for errorsarlm
s2

# construct identity matrix
n = length(sar.Model$fitted.values)
I = diag(n)

# get rho
rho = sar.Model$lambda

# create residual covariance matrix
C = s2*solve(t(I-rho*W)%*%(I-rho*W))
Sigma.hat = C

n=length(MEANTCC)
X.new=matrix(1,n,3)
X.new[,2]=AGE
X.new[,3]=(AGE)^2

n.sim=1000
y.pred.mat=t(rmvnorm(n.sim, X.new%*%beta.hat, Sigma.hat))

# create empty matrix for prediction bounds for all n observations
pred.Bounds = matrix(0,n,2)

# populate matrix
for (i in 1:n){
  lwr = quantile(y.pred.mat[,i], 0.025)
pred.Bounds[i,1] = lwr
  upr = quantile(y.pred.mat[,i], 0.975)
pred.Bounds[i,2] = upr
}

X.new = data.frame(X.new)

# get predicted values for data on which the model was fitted
pred.Values = as.vector(predict.sarlm(sar.Model,newdata=X.new, listW.pts))

# combine predicted values with lwr and upr bounds
pred.Table = cbind(pred.Values, pred.Bounds)
colnames(pred.Table) = c("pred","lwr","upr")
head(pred.Table)

# Find out mean lwr band & upr band
pred.Table = data.frame(pred.Table)
meanLwr = mean(pred.Table$pred - pred.Table$lwr)
meanUpr = mean(pred.Table$pred - pred.Table$upr)
meanLwr
meanUpr
# Python script used to generate a text file of $2^{16} - 1 = 65,536$ combinations of 16 variables for Multiple Model Inference Analysis (MMI) in chapter 3.
# Code courtesy of Chris Garrard

# define the function MMIList with two arguments, "list" and "prefix"
def MMIList(list, prefix):
    for i in xrange(len(list)):
        new_prefix = prefix + '+' + list[i]
        line1 = 'SAR.function.R(data, listW.pts, Ydat'+ new_prefix +')'
        line2 = line1.replace("Ydat+","Ydat~")
        line3 = line2.replace(");",")")
        outFile = open('C:/PhD_Work/DATA/ANALYSIS/PART_II/STATS/MMIList.txt', "a")
        outFile.write(line3+"
")
        outFile.close()
        MMIList(list[i+1:], new_prefix)

list = ['x1','x2','x3','x4','x5','x6','x7','x8','x9','x10','x11','x12','x13','x14','x15','x16']

# run the function MMIList on the list called "List"
MMIList(list, '')
CURRICULUM VITAE

JOHN H. LOWRY JR.
Remote Sensing/GIS Lab (UMC 5275)
College of Natural Resources
Utah State University, Logan, Utah 84322

Phone: 435-797-0653
E-Mail: john.lowry@usu.edu
http://www.gis.usu.edu/~jlowry

June, 2010

RESEARCH and TEACHING INTERESTS

Spatial urban ecology; Human-environment interactions in urban and regional settings; Landscape ecology; Geographic information science (GIScience); Spatial modeling; Bioregional planning, Remote sensing-based mapping; GIS project management in natural resources.

EDUCATION


PROFESSIONAL EXPERIENCE

Associate Director (2005-present) RS/GIS Lab, Utah State University, Logan, Utah
Assistant Director (1999-2005) RS/GIS Lab, Utah State University, Logan, Utah
Graduate Research Assistant (1992-1994) Dept. of Geography, Univ. of Utah, SLC.
English Teacher (1991) Cambridge English Institute, Chungli, Taiwan, ROC.

TEACHING at UTAH STATE UNIVERSITY (full responsibility)

• Introduction to Geographic Information Science (WATS/GEOG 2930, F 2009)
• Geographic Information Systems (WATS 4930/6920, F 2008)
• GIS Programming with Python (WILD 6900, Sp 2008; Sp 2009)
• Geographic Information Systems (WATS 4930/6920, Distance Ed., F 2007, F 2008)
• Geographic Information Analysis (WATS 5930/6930, Distance Ed., Sp 2008; Sp 2009)
• Introduction to GIS fundamentals (GEOG 593/693, Winter 1997)
• Geographic Information Analysis (GEOG 594/694, Spring 1997)

ACCADEMIC ADVISING (adjunct advisor)


Vaugh, Dusty (present) GIS-based assessment of wilderness areas within the state of Colorado. Plan A Masters of Science in Recreation Resource Management.

PEER-REVIEWED PUBLICATIONS


PUBLICATIONS in PREPARATION

Lowry, J. H., Determinants of urban tree canopy in residential neighborhoods: Household characteristics, urban form, and the geophysical landscape.

Lowry, J. H., Predicting urban forest growth and its impact on residential landscape water demand in a semi-arid environment.
Lowry, J. H., Measuring urban/suburban form: A spatial and demographic characterization of residential neighborhoods.

PROCEEDINGS and BULLETINS


TECHNICAL REPORTS and DOCUMENTS


PROFESSIONAL and ACADEMIC HONORS

Selected Participant, National Science Foundation and Social Science Research Council, Snowbird Charrette for Environmental Research Design, Cliff Lodge, Snowbird, Utah, August 24-28, 2006.


**GRANTS, CONTRACTS and PROJECT MANAGEMENT (As Principle Investigator)**

- Feature Extraction Wildland-Urban Interface (2009; $10,000; State of Utah AGRC).
- Feature Extraction of Urban Structures (2008; $25,000; State of Utah AGRC).
- Utah State University Campus Planning GIS and Internet Map Server (2007; $10,000; Utah State University).
- Utah Test and Training Range, Hill Air Force Base, SPOT Imagery Vegetation Classification (2000; $17,500; Utah Division of Wildlife Resources).
- World Wide Web-Page Design and Implementation (2000; $17,000; Utah Division of Wildlife Resources).
- GIS Services of Photo Interpretation and Digitizing Riparian Cover (2000; $3500; US Bureau of Reclamation).

**GRANTS, CONTRACTS and PROJECT MANAGEMENT (As Co-Principle Investigator)**

- Development and Coordination of a Region-wide Gap Analysis Project for Utah, Nevada, Colorado, Arizona and New Mexico (1999-2005; $1,120,000; USGS BRD GAP).

**GRANTS, CONTRACTS and PROJECT MANAGEMENT (As Project Manager)**

- Spatial Analysis Project (SAP) for Utah Division of Forestry, Fire and State Lands (2006-2007; $11,000; USU Extension).
• Mule Deer Mapping Project (2003-2004; $40,000; Western Association of Fisheries and Wildlife Agencies).

CONSULTING ACTIVITIES


INVITED WORKSHOPS, SHORT COURSES and ORGANIZED FIELD TRIPS


Southwest Regional GAP Project Data Product Debut, National Gap Analysis Conference and Interagency Symposium, Reno, NV, December 5, 2005.

GIS Workshop for SEDESOL (Secretaria de Desarollo Social), Mexico City, Mexico, February 22-26, 1993.

ORAL PRESENTATIONS (As Presenter and Primary Author)


An Overview of the Southwest Regional Gap Land cover Dataset, National Gap Analysis Conference and Interagency Symposium, Reno, NV, Dec. 5-8, 2005.


Land Cover Mapping Background: Training Data and Classification Methods The 19th Annual Symposium International Association for Landscape Ecology United States Regional Association (US-IALE), March 30-April 2, 2004, Las Vegas, NV.

Land Cover Mapping for the Southwest Regional Gap Analysis Project Tenth Biennial USDA Forest Service Remote Sensing Applications Conference, April 5-9, 2004, Salt Lake City, UT.


POSTER PRESENTATIONS (As Primary Author)


Herramientas para la aplicación espacial de Árboles de Clasificación en el mapeo de la cobertura de la Tierra. VI Congreso Nacional de Áreas Protegidas de Mexico. Monterrey, Mexico. November 3-6, 2004.


Provo River Instream Vegetation and Habitat Mapping, Southwest Arc/Info Users Group Conference, Midway, Utah, October 15-17, 1995.

PROFESSIONAL AFFILIATIONS

US International Association for Landscape Ecology (US-IALE) 2003-present
Society for Conservation GIS (SGIS) 2003-present

JOURNAL REVIEWER

Photogrammetric Engineering & Remote Sensing
Ecological Modeling

SERVICE

Vice President, Intermountain Region ASPRS. (2007- present)
USU Ecology Center, Graduate Student committee member (2007-2008)
USU College of Natural Resources, Graduate Student Council (2004-2005)

LANGUAGES

English (native), Spanish (fluent), Mandarin Chinese (survival)