Modeling Land Use Change using Spatial Statistics and Spatial Panel Data Regression in Dallas-Fort Worth Metropolitan Area (1990 - 2020)

by

ALI BEHSERESHT

Presented to the Faculty of the College of Architecture, Planning and Public Affairs

The University of Texas at Arlington in Partial Fulfillment

of the Requirements for the Degree of

DOCTOR OF PHILOSOPHY

THE UNIVERSITY OF TEXAS AT ARLINGTON

December 2023

To:

my parents

Hassangholi and ***Torkan***

They could not read or write but understood how important education was. They did not withhold any efforts to support their children in getting educated. I owe my success to their support, encouragement, and faith in me.

THE UNIVERSITY OF TEXAS
AT ARLINGTON

Acknowledgments

I sincerely thank my supervisor, Professor Ardeshir (Ard) Anjomani, for his guidance, support, and encouragement throughout my academic journey. Thanks to my dissertation committee members, Dr. Taner Ozdil and Dr. Joowon Im; their insightful comments were very helpful in identifying the strengths and weaknesses of my work and addressing the gaps and limitations in my analysis.

Also, I would like to thank my dear friend, Dr. Mohsen Ghiasi, who assisted me in collecting the relevant data for this study, and Dr. Emre Tepe for sharing his R scripts, which, as a road map, expedited my research modeling.

Finally, and more importantly, I would like to express my deepest gratitude to my family, my lovely wife, Zahra, and my little angel, Nicki, who have been my constant source of inspiration, motivation, and support throughout this research. Their love, patience, and encouragement have helped me overcome many challenges and reach the end of my Ph.D. journey.

The University of Texas at Arlington

DISSERTATION DEFENSE REPORT (DDR)

This report must be submitted to the Office of the Registrar, via email to gradteam@uta.edu, after the examination is administered regardless of the outcome of the defense. Students & advisors should consult the current Graduate Catalog for deadline dates applicable to the administration & report of the exam. The student must be enrolled in the term in which he/she takes the Dissertation Defense. An unconditional pass is required before degree can be conferred. In order to update a student from "passed with conditions" the Milestone Conditions Update form will need to be submitted.

Student: Behseresht (Last Name:) Ali (First Name:)

Date of Examination: 11/20/2023

UTA ID: 1000805800

Degree Name: PhD - Urban Planning and Public Policy

Example: History BA to PhD

College/School: CAPPA

Dept: CAPPA

Supervisor: Ard Anjomani

Title of Dissertation: Modeling Land Use Change Using Spatial Statistics and Spatial Panel Data Regression In Dallas-Fort Worth Metropolitan Area (1990 - 2020)

This is to report that the above-named student, under the direction of their committee, completed the Dissertation Defense with the following results

- [x] Passed Unconditionally, the dissertation document and defense fully satisfy requirements for the doctoral degree.
- [ ] Passed, with following conditions that must be satisfied
- [ ] Failed, with permission to retake with these stipulations
- [ ] Failed, dismissed from the program

Name (typed)	Signature	Date (mm/dd/yyyy)
Ard Anjomani Supervisor/Chair	Ard Anjomani Digitally signed by Ard Anjomani Date: 2023.11.20 21:03:57 -06'00'	11/20/2023
Taner R. Ozdil Committee Member	Taner R Ozdil Digitally signed by Taner R Ozdil Date: 2023.11.21 14:48:54 -06'00'	11/20/2023
Joowon Im Committee Member	Im, Joowon Digitally signed by Im, Joowon Date: 2023.11.20 17:48:10 -06'00'	11/20/2023
Select Title		
Select Title		
Select Title		
Ali Behseresht Student	Ali Behseresht Digitally signed by Ali Behseresht DN: OU=CAPPA, O=University of Texas at Arlington (UTA), CN=Ali Behseresht, E=ali.behseresht@mavs.uta.edu Reason: I am the author of this document Location: Date: 2023.11.22 08:09:58-06'00' Foxit PDF Editor Version: 12.1.3	11/20/2023

09/2021

Abstract

Modeling Land Use Change Using Spatial Statistics and Spatial Panel Data Regression

In the Dallas-Fort Worth Metropolitan Area (1990 - 2020)

Ali Behseresht, Ph.D.

The University of Texas at Arlington (UTA), December 2023

Supervising Professor: Dr. Ardeshir (Ard) Anjomani

In this dissertation research, we explored land use change for Activity and Residential land uses in the Dallas-Fort Worth Metropolitan Area (DFW) from 1990 to 2020 in association with spatial and socioeconomic factors. It aims to study three aspects of land use change: dynamics, drivers, and impacts. The research started with three hypotheses, all proven and accepted by the research findings.

We reviewed the relevant literature from five aspects of land use change: theories, models, methods, drivers, and impacts. Based on the reviewed literature, the land use change model we discussed is an economics-based spatiotemporal modeling approach in which the basic concept of land use change is a historical spatial development pattern.

For land use change dynamics, we used spatial statistics methods of Global Moran's I for spatial autocorrelation, Getis-Ord Gi* for hot spot analysis, and Anselin Local Moran's I for cluster/outlier analysis. For drivers of land use change, on the other hand, we used Spatial Panel Data Regression (SPDR) to model the factors that drive land use change. Finally, for the impacts of land use change, we used the results from the spatial statistics and the SPDR to discover the impacts of land use change on the urban structure.

The results of dynamics of land use change show that Activity land uses gradually leave central regions of the DFW Metropolitan Area in favor of the peripheral regions. On the other hand, Residential land uses, while expanding toward suburbs, tend to fill the vacancies in the CBDs created by Activity land uses. Also, the Z-score (as a measure of compactness) is decreasing while the P-value (the probability of randomness of the distribution pattern) is increasing for both Activity and Residential land uses over time, indicating the scatteredness of hot spots of these land uses. The new hot spots for Activity land uses are primarily located in the north, east, and south of the DFW Metropolitan Area. In contrast, the inner parts (including downtown Dallas, downtown Fort Worth, and the corridor between them) are losing Activity land uses. Therefore, while the null hypothesis of Complete Spatial Randomness (CSR) is rejected, economies of scale (agglomeration economies) and spatial dependency theories of land use distribution are weakening in explaining the distribution pattern of Activity and Residential land uses in the DFW Metropolitan Area.

We tested spatial and non-spatial regression models to model the drivers of land use change. Spatial regression models outperform non-spatial models; therefore, we used the Spatial Panel Data Regression (SPDR) model. Several theoretically and empirically essential variables were statistically significant, with the right signs explaining land use changes. Model results show that land use change is autoregressive, meaning the ratio of Activity and Residential land uses in a block group at a previous time (t-1) is the most significant driver of land use change. Also, vacant land is the second most significant driver of land use change, indicating ample land in the region is helping the sprawl. Finally, the second law of geography, the phenomenon external to a geographic area of interest that affects what goes on inside it, is applicable to land use change.

By combining the results of dynamics and drivers of land use change, we analyzed how land use change may have impacted the urban structure of the DFW Metropolitan Area. To do so, we cross-examined the Z-score and distribution of AADT using the QQ Plot. The results of the analysis show that land use change resulted in urban sprawl and the spread of traffic congestion in the region. Therefore, the DFW Metropolitan Area has evolved into a sprawled multi-centric Metropolitan Area; it is a region of dynamism and growth, where several activity and residential centers have transformed, and new activity and residential centers have emerged.

Finally, we discussed and recommended modeling, development, land use, transportation, socioeconomic, and environmental implications of land use change model results in the DFW Metropolitan Area. Understanding the causes and effects of land use changes helps planners, policymakers, and local legislators observe the impacts of such development policies and regulations. Furthermore, via urban and regional policies, they can promote policies to improve the attractiveness of urban or suburban areas as locations for investment and intervene with policies to mitigate the negative consequences of these changes.

Keywords: Land Use Change Modeling, Spatial Statistics, Spatial Panel Data Regression, DFW Metropolitan Area.

Table of Contents

List of Tables

List of Figures

List of Maps

Chapter 1: Introduction

1.1 Research Background

This research explores land use changes in large US Metropolitan Areas. It selects the census-defined urban area of the Dallas-Fort Worth Metropolitan Area (DFW) as the study area, called the "DFW Urban Area." The study area encompasses physically connected cities around the Dallas-Arlington-Fort Worth corridor, modified to fit the census 2020 boundaries. This area was chosen due to substantial population centers that have economic and social integration with each other. It uses spatial statistics and Spatial Panel Data Regression (SPDR) to analyze data related to research purposes.

According to the 2020 U.S. Census Bureau, the Dallas-Fort Worth Metropolitan Statistical Area (DFW-MSA) is the fourth largest Metropolitan Area with a population of about 7.7 million. Between 2010 and 2020, while DFW-MSA scored the second-highest population growth among the top 11 most populous U.S. MSAs, it only ranked 8th regarding population density. This low population density, which resulted from an extensive amount of vacant, undeveloped land, and low-density developments, means DFW-MSA consumes more land for development than is usually required for a Metropolitan Area of a similar size. The lack of physical barriers, large water bodies, development policy and regulations, and abundant land availability accelerates such development in Metropolitan Areas (Ghosh et al., 2017).

Table 1-1 provides population-related metrics for the DFW-MSA compared to the top 11 MSAs in the country based on census data. They show that DFW-MSA ranked 2nd, 4th, and 8th for population growth, total population, and population density, respectively. These metrics could mutually impact the Metropolitan Area's land use (type, diversity, and distribution), urban structure (compactness vs. sprawl; mono-centricity vs. multi-centricity), and traffic flow, among other impacts.

Table 1-1: Most populous U.S. Metropolitan Areas

Metropolitan Statistical Area	Population 2010 (M)	Population 2020 (M)	Population Growth 2010 – 2020	Population Density 2020	Population Growth Rank	Population 2020 Rank	Population Density Rank
Atlanta-Sandy Springs-Alpharetta	5.4	6.0	13.3%	689	6	9	10
Boston-Cambridge-Newton	4.5	4.9	9.3%	1095	7	10	4
Chicago-Naperville-Elgin	9.6	9.6	0.5%	1004	11	3	6
Dallas-Fort Worth-Arlington	**6.4**	**7.6**	**20.0%**	**848**	**2**	**4**	**8**
Houston-The Woodlands-Sugar Land	5.7	7.1	24.3%	754	1	5	9
Los Angeles-Long Beach-Anaheim	12.9	13.2	2.5%	2316	10	2	1
Miami-Fort Lauderdale-Pompano Beach	5.5	6.1	13.4%	1009	5	8	5
New York-Newark-Jersey City	19.0	20.1	6.0%	2186	9	1	2
Philadelphia-Camden-Wilmington	5.8	6.2	7.0%	1282	8	7	3
Phoenix-Mesa-Chandler	4.2	4.8	15.6%	332	4	11	11
Washington-Arlington-Alexandria	5.3	6.4	19.2%	912	3	6	7

Analysis of 30 years of population and land use data from 1990 to 2020 shows a more than 91 percent increase in the population in the DFW Urban Area. On the other hand, the developed land area grew by 77 percent in the same period, which shows that the DFW Metropolitan Area is becoming denser. Table 1-2, Figure 1-1, and Map 1-1 show population and developed land area growth in the DFW Urban Area from 1990 to 2020.

Table 1-2: Population growth vs. developed land area growth in the DFW Urban Area (1990 to 2020)

Year	Land (Developed Area)			Population (Study Area)		
	Area (sq. miles)	Growth (%)	Cumulative Growth (%)	Population	Growth (%)	Cumulative Growth (%)
1990	782	0.00	0.00	3,525,000	0.00	0.00
1995	800	2.30	2.30	3,900,000	10.40	10.4
2000	969	21.00	23.85	4,550,000	16.90	29.00
2005	1071	10.55	36.90	5,050,000	10.65	42.75
2010	1301	21.50	66.35	5,600,000	10.80	58.20
2015	1317	1.20	68.35	6,200,000	10.80	75.30
2020	1385	5.20	77.00	6,750,000	9.00	91.20

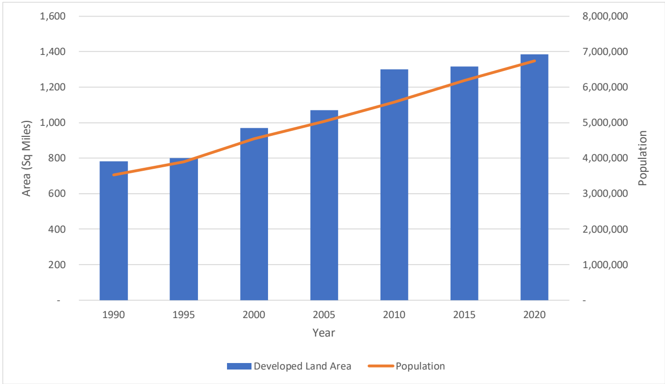
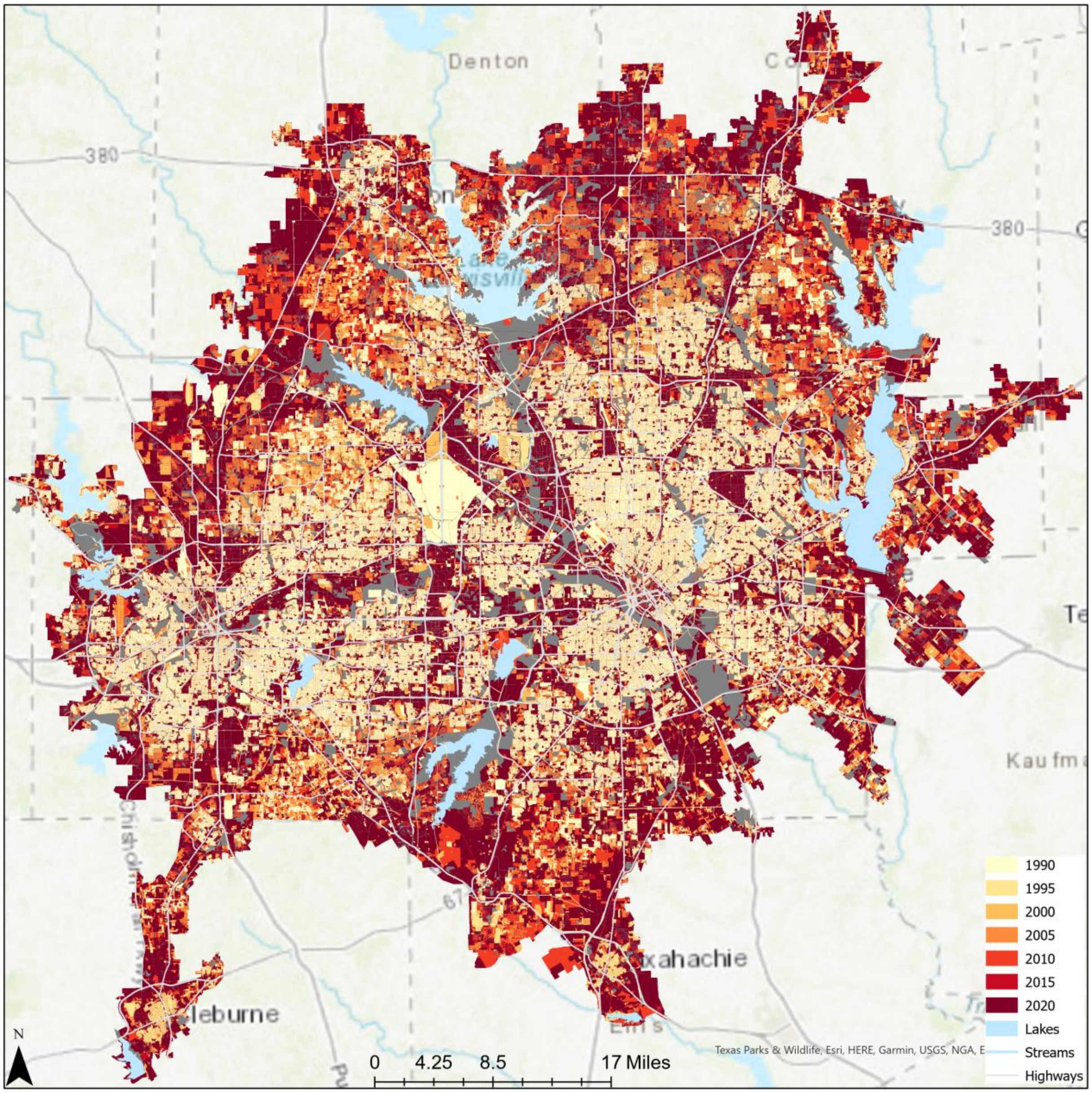


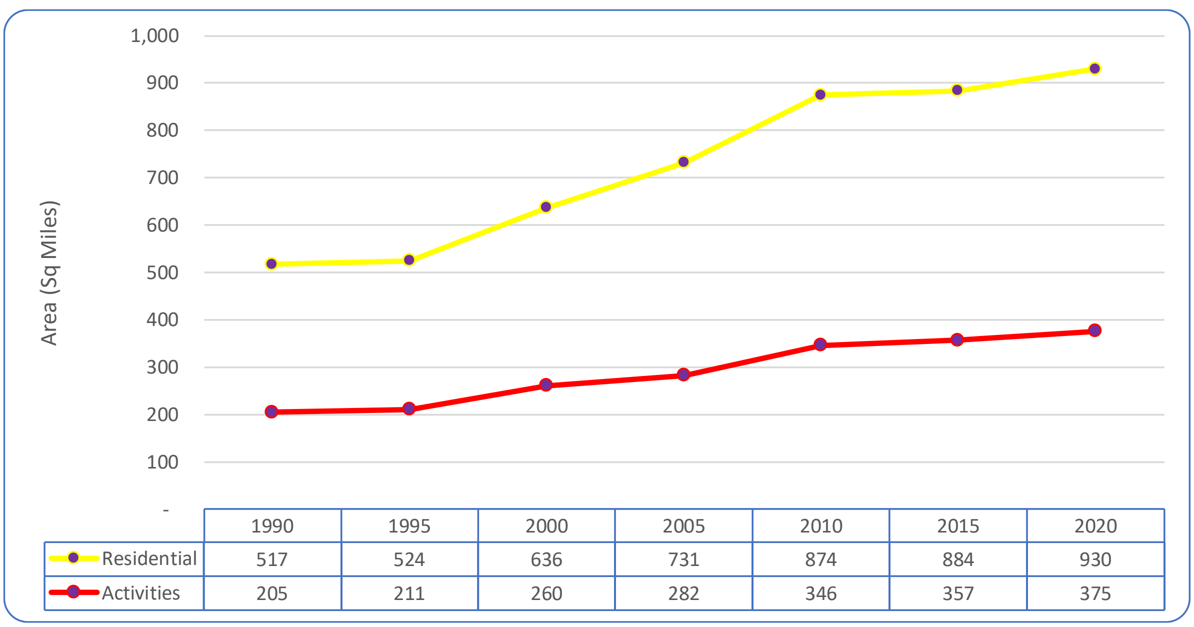
Figure 1-1: Population vs. developed land area change in the DFW Urban Area (1990 – 2020)



Map 1-1: Developed land area change in the DFW Urban Area (1990 to 2020)

The analysis of land uses in the DFW Urban Area between 1990 and 2020 shows that while the developed land area grew by 77 percent, Residential land uses increased by 80 percent, and Activity[1] land uses (industrial, retail, office, mixed-use, institutional/semi-public, education, and hotel/motel) grew by 83 percent. It indicates that various land uses are growing at a different pace over time and space; therefore, they might impact planning-related matters differently.

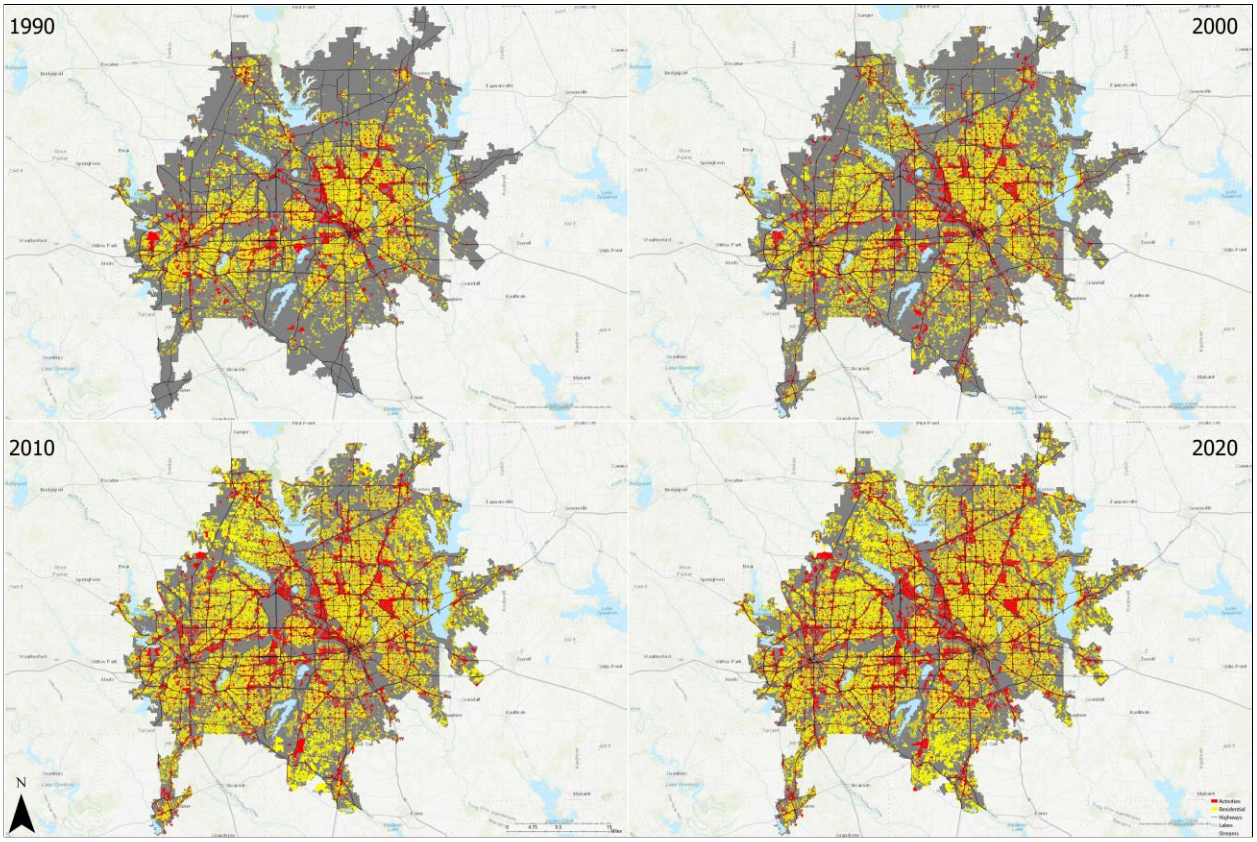
As Figure 1-2 shows, Residential land use grew at a higher rate between 1995 and 2010, then slowed down afterward. On the other hand, Activity land uses grew steadily at a slower pace, even though they had a slightly higher growth rate than Residential land uses (82.9 vs. 79.9% for Residential growth). However, as expected, Residential land uses growth was more than the Activity land uses growth. As Map 1-2 shows, the distribution of land use growth for these two land uses is not similar; some areas gained more Residential land uses, and others attracted more Activity land uses based on physical and socioeconomic factors that impacted their distribution.



	1990	1995	2000	2005	2010	2015	2020
Residential	517	524	636	731	874	884	930
Activities	205	211	260	282	346	357	375

Figure 1-2: Area of Activity and Residential land uses in the DFW Urban Area (1990 - 2020)

[1] According to the American Planning Association (APA), Activity land uses can be classified according to different categories, such as economic activities, social activities, cultural activities, or environmental activities.



Map 1-2: Activity and Residential land uses change in the DFW Urban Area (1990 - 2020)

Studying land use change and distribution patterns has always interested urban spatial modelers, planners, and policymakers (Sleeter et al., 2012). Spatial modelers can model the land use change to understand the growth trends in land use types (e.g., commercial, residential, and industrial), the distribution pattern of land use types, and potential anomalies in a specific land use type. More importantly, they can explain the cause and effect of these growth, distribution patterns, and incongruities.

Planners, on the other hand, can relate land use changes to other planning-related areas like housing, income, transportation, etc., to evaluate the mutual impact of land use changes and these planning matters. For instance, establishing a gigafactory in a low-density industrialized area can cause transportation improvements and bring related shopping centers and housing complexes. Eventually, it could cause land use changes in the area by creating agglomeration and attracting other firms and services.

Finally, in the bigger picture, policymakers and decision-makers can consider changes in land use patterns and the causes of these changes in imposing future development policies. These policies could include specific incentives (e.g., tax exemption) for a particular business (or business type) in a specific area to encourage certain activities. They also can limit the establishment of certain businesses in designated areas by imposing restrictions (e.g., solid environmental considerations or higher taxes). These policies could impact the development in desired areas, eventually affecting a Metropolitan Area's overall structure.

1.2 Research Problem

According to 2020 census data, the population of the DFW-MSA has grown by over 20 percent over a decade since 2010, ranking it second highest population growth rate among the country's top 11 most populous MSAs. Regarding population density, however, it is one of the least dense Metropolitan Areas in the country; it ranks 8th among the 11 most populous MSAs in the country. In a longer time, however, analysis of the population and developed land area from 1990 to 2020 shows that while the DFW Urban Area population increased by over 91 percent, the developed land area increased by 77 percent for the same period.

This growth imbalance between population and developed land area could indicate that the DFW Metropolitan Area has become denser in the past 30 years. Ideally, there should be a balance between population growth and developed land area (population density) in a Metropolitan Area to reduce the negative impacts of an imbalanced development. Otherwise, such an imbalance creates dense or scattered urban structures in the Metropolitan Areas, creating environmental issues[2] and increasing travel time[3] (Alig et al., 1987; Banzhaf et al., 2013; Bhatta, 2009; Lambin et al., 2000; Lambin et al., 2001; Lo et al., 2002; Wu, 2008) to name a few.

However, due to physical, social, economic, land use, and transportation forces (Živković, 2019), urban structures are far from ideal; they are either dense or scattered[4]. Such an imbalanced urban structure has two-sided causes and effects; the forces that cause changes in urban structure are also impacted by it. The urban structure's fundamental cause and effect is land use change, which consequently impacts other aspects of a Metropolitan Area, like the transportation system and travel time (Meyer et al., 2000), as well as the environmental system, which could further cause sustainability challenges.

According to the initial analysis and observation of land use change from 1990 to 2020, the DFW Metropolitan Area is going through an interesting period of development impacting its structure and social and economic changes that deserve close attention. While the centers are growing and multiplying, it is becoming a scattered Metropolitan Area; therefore, what caused land use change in the past 30 years and how it impacted the urban structure of the DFW Metropolitan Area is the problem on which this research will shed light.

1.3 Research Purpose

In land use change modeling, the main goal is to simulate the internal and external land dynamics and conditions that lead to land use change and assess the impact of land use change

[2] For instance, while scattered cities convert more land to developed land areas via land use conversion, a dense city can create urban heat islands.
[3] Travel time increases via congested streets in a dense Metropolitan Area; also, it can increase by creating long travel distances in a scattered Metropolitan Area.
[4] Since the density level in a Metropolitan Area is relative, it is not an easy task to measure it.

(Briassoulis, 2000; Tepe, 2023). According to the land economics theory, which is based on the neoclassical economics assumption of utility maximization for individuals and profit maximization for firms (Weintraub, 2002), "the current land use status on a given parcel changes if the expected benefits from changing conditions are larger than the benefits under current conditions" (Tepe, 2023, p. 1).

From Alonso's theory of location point of view, residential, commercial, and industrial land uses compete for location according to the bid rent curve and their requirements for access to the city center (Alonso, 1964). Residential land uses prefer to be located closer to the employment centers, which provide better access to jobs and minimize transportation costs. Commercial land uses are market-oriented; they try to maximize their market size and minimize travel costs, which is achievable via access to a larger market by accessing more concentrated urban centers (Hoover & Giarratani, 1984). Industrial land uses are input-oriented; they prefer to be located closer to production source rather than their output market (McCann, 2013). Also, land cost does not significantly affect their location (Hoover & Giarratani, 1984).

This competition for location among various land use types leads to land use change, which has socioeconomic and environmental impacts. The consequences of land use change include but are not limited to lengthy travel time, water quality degradation, air pollution, biodiversity loss, urban heat island effects, socioeconomic disparities, social fragmentation, and infrastructure costs (Maimaitijiang et al., 2015).

This dissertation aims to explore land use change for Activity and Residential land uses from 1990 to 2020 in the DFW Urban Area in association with spatial and socioeconomic factors. To achieve this goal, it will employ spatial statistics and Spatial Panel Data Regression (SPDR) to

- Identify the spatial cluster pattern of Activity and Residential land uses.
- Discover the main spatial distribution growth/decline for Activity and Residential land uses.
- Explore relationships between Activity and Residential land uses changes and spatial and socioeconomic factors.
- Show the changes in the urban structure of the DFW Metropolitan Area as the result of land use change.

Considering the land use change modeling goal and the research purpose mentioned above, we will study three aspects of land use changes in the DFW Urban Area: *dynamics* of land use change, *drivers* of land use change, and *impacts* of land use change.

First, we will study the *dynamics* of land use change in the DFW Urban Area by finding the location and size of new hot spots (or clusters) of Activity and Residential land uses, as well as gain or loss in existing hot spots (or clusters) of such land uses. It will reveal the location preferences of Activity and Residential land uses in each period. Regarding this aspect of land use change, the research hypothesis is that the economy of scale (aka agglomeration economy) and spatial dependency are weakening in explaining the distribution of land use change.

Next, we will study the *drivers* of land use change for Activity and Residential land uses during each period. This objective is achievable by associating land use change data with the theoretically and empirically related factors such as proximity/accessibility (e.g., distance from CBDs, proximity to large water bodies and major streams), socioeconomic (e.g., income, population, and jobs), natural and environmental (e.g., slope, precipitation, flood hazard), site-specific (e.g., vacant land), and development policies. In this regard, the research hypothesis is that land use change is autoregressive; the ratio of Activity or Residential land uses in a block at a previous time (t-1) is a significant driver of land use change, and the second law of geography applies to land use change as a spatial phenomenon.

Last, we will study the *impact* of land use change on the urban structure of the DFW Metropolitan Area by analyzing the changes in the distribution of Activity and Residential land uses with the distance from CBDs, transportation networks, and other amenities that such land uses prefer. In this regard, the research hypothesis is that multiple hot spots of Activity and Residential land uses have changed, and new clusters of Activity and Residential land uses have emerged beyond the Dallas-Fort Worth corridor because of land use change.

The model simulates land use changes every five years for 30 years. This would help to predict the next five years of land use. These predictions are the cornerstone of urban planning, which affects many aspects of society, such as the environment, economy, and transportation.

Understanding the causes and effects of land use changes in the DFW Metropolitan Area helps local and regional legislators, policymakers, and planners to observe the impacts of such development policies and regulations. Furthermore, via urban and regional policies (McCann, 2013), they can promote policies to improve the attractiveness of urban or suburban areas as locations for investment and intervene with policies to mitigate the negative consequences of these changes.

In a similar but limited approach, Al-Shammari (2007) shed light on the factors influencing the growth of employment sub-centers in major urban areas. He used a set of variables to identify and characterize employment sub-centers in the DFW Metropolitan Area and explained their growth rate variation over ten years. The research findings recommended several policies for enhancing sub-center development in the DFW Metropolitan Area, which is crucial for long-term economic and social sustainability.

1.4 Research Questions

As mentioned, the purpose of this dissertation is to study the "dynamics," "drivers," and "impacts" of land use change using spatial statistics and Spatial Panel Data Regression (SPDR) in the DFW Metropolitan Area. To achieve the research goals, we must answer the following research questions.

1. Regarding the dynamics of land use:
    - 1.1. How have the Activity and Residential land use distribution patterns changed over time?
    - 1.2. Where are the new hot spots of Activity and Residential land uses?
    - 1.3. Which hot spots grow, and which ones decline in size?
2. Regarding drivers of land use change:
    - 2.1. What are the significant factors (site-specific, physical, proximity, socioeconomic) of land use change?
3. Regarding the impact of land use change:
    - 3.1. How did land use change impact the urban structure of the DFW Metropolitan Area?
    - 3.2. What are the impacts on the pattern of the Activity and Residential land uses in terms of loss and gain and intensity changes?
    - 3.3. What are the impacts on the transportation, particularly traffic flow?

1.5 Research Significance

Land use change happens due to several factors, like characteristics of a parcel, structure of the neighborhood, historical trend, amenities, services, and infrastructure, socioeconomic factors, zoning regulations, and development policies (Tepe et al., 2020). Land use change studies attempt to explain the what, where, when, how, and why of changes to the use of the land (Sleeter, 2012), and land use change modeling is the approach to answer these questions.

Land use change modeling is challenging due to the complexity of formulating the relationship between what, where, when, how, and why of a land use change. Such system complexity and lack of historical land use data availability create significant challenges in the methodology and computation of land use change models (Tepe et al., 2020). Several modeling approaches can be used to formulate such a complex relationship; these models include but are not limited to remote sensing, cellular automata, and statistical models (Baker, 1989; Briassoulis, 2000; Gore et al., 1991; Irwin et al., 2001; Lambina et al., 2001; Rosa et al., 2016; Verburg et al., 2004; Wegener, 1995).

Even though these models are generally known as land use-land cover (LULC) change models, they deal with land cover change because the (so-called) land use data used is satellite images or aerial photos, which are suitable for extracting land cover, not land use. By definition, land cover is how much land is covered by developed areas, bare land, forest, agriculture, wetlands, and water. On the other hand, land use shows how people use land for residential, commercial, industrial, educational, office, mixed use, etc.

There are various reasons that some researchers have used land cover change rather than land use change in their modeling efforts. Among these reasons are the extensive availability of satellite images and areal photographs, lack of historical parcel-level land use data, and lack of software to model land use changes (Lo et al., 2002; Maimaitijiang et al., 2015; Malczewski, 2004). Today, however, thanks to Geodatabases, software, and the availability of historical socioeconomic and land use data (Berke et al., 2006), researchers can model land use change in parcel level or any aggregation like blocks, census boundaries, zip codes, TAZs, etc.

The significance of this dissertation research is four-fold. First and foremost, we will use spatial statistics and Spatial Panel Data Regression (SPDR) for the first time to model land use changes[5]. The results of this research can be a road map for urban modelers to replicate a similar approach in modeling land use change in other Metropolitan Areas (or even a smaller urban area), which would immensely enhance our understanding of the growth/decline of urban areas and related socioeconomic changes. Next, exploring changes in a central Metropolitan Area's Activity and Residential land uses patterns for three decades will enhance our understanding of the growth/decline and spatial dynamics of North American urban regions. Third, we will associate land use change data with spatial and socioeconomic data to analyze Residential and Activity land uses change in the DFW Metropolitan Area from 1990 to 2020 to find the drivers of land use change and the significance of each factor in each period. Finally, we evaluate the impact of land use change on urban structure in the DFW Metropolitan Area.

Understanding the points mentioned above helps local/regional legislators, policymakers, and planners mitigate or eliminate these negative consequences and enhance the positive impacts in the DFW Metropolitan Area. The results also help planning and policymaking at regional, sub-regional, city, and neighborhood levels in various arenas, including urban issues, economic development, environmental planning, etc.

[5] Potential application of spatial statistics in land use change modeling is not new; however, (to my knowledge) it has not been applied in case of studies yet.

1.6 Research Report Outline

This dissertation research is comprised of Seven chapters. In Chapter 1, we talked about the research background, problem, and objectives. In Chapter 2, we reviewed the literature relevant to various aspects of LULC modeling. In Chapter 3, we discussed the research methodology, including the study area, the unit of analysis, the data source, the software and applications used, and how the research is conducted.

In Chapters 4, 5, and 6, we covered each of the three aspects of the research purpose. In Chapter 4, we discussed the dynamics of land use change using spatial statistics; in Chapter 5, we covered the drivers of land use change using Spatial Panel Data Regression (SPDR); in Chapter 6, we deliberated the impacts of land use change on urban growth, traffic flow, and urban structure in the DFW Metropolitan Area.

In Chapter 7, we summarized the research findings, policy implications, challenges and limitations of the research, and potential areas for future research that we did not (or could not) cover in this research due to the challenges and limitations of this research.

Chapter 2: Literature Review

Land use is a crucial element of urban and regional planning; it encompasses more than just the basic categories of land use. It involves various aspects and elements related to land, such as the functions and purposes of different land uses, the activities that take place on land, the role of land in environmental systems, the value of land as a marketable asset, the planning and regulation of land by public authorities, and the visual and symbolic significance of land for orientation and social identity (Berke et al., 2006).

Due to the abovementioned characteristics and importance, land use has been the subject of many types of research in planning and related fields. Despite the interest among many academic disciplines like urban and regional planning, geography, and environmental science to study land use, limited researchers have studied land use as a spatial phenomenon that is the result of physical and socioeconomic factors (Briassoulis, 2000; Lo et al., 2002; Maimaitijiang et al., 2015), that impacts and impacted by other land uses in its surrounding.

Researchers interested in land use change have studied, theorized, and modeled land use from their academic lens. For instance, environmental, GIS, and R.S. experts have mostly modeled land cover (as the biophysical earth surface) changes. They have been looking for more sophisticated remote sensing techniques to extract land cover using aerial images as accurately as possible (Anderson et al., 1979). Planners and geographers, among other social scientists, view land use and its changes over time as a phenomenon formed on land due to people's behavior, socioeconomic interactions, and political influences. However, their lack of spatial knowledge and related modeling techniques thwarted them from using the right tools for modeling land use change (Briassoulis, 2000; Lambina et al., 2001).

Part of the problem in land use modeling is due to the confusion between "land use" and "land cover," which is used interchangeably in the land use/land cover (LULC) literature. Land cover can be extracted by analyzing satellite and aerial imagery, while land use is the result of gathering data via surveys and field trips, among other tools. Since these two terms are different but used interchangeably, it is required to define them to make a clear distinction between the two terms.

Turner et al. (1995) define land cover as "…the biophysical state of the earth's surface and immediate subsurface…" and land use "… involves both the manner in which the biophysical attributes of the land are manipulated and the intent underlying that manipulation – the purpose for which the land is used…" (p. 20). According to Meyer (as cited in Moser, 1996), land cover "…describes the physical state of the land surface: as in cropland, mountains, or forests…" and land use "…denotes the human employment of land…" (p. 247). Finally, FAO[6] (1995) states that "land use concerns the function or purpose for which the land is used by the local human population and can be defined as the human activities which are directly related to land, making use of its resources or having an impact on them…" (p. 21).

To summarize, land cover is how much a region is covered by developed land, bare land, forest, agriculture, wetlands, and water. On the other hand, land use shows how people use land for residential, commercial, industrial, educational, office, mixed use, etc. Therefore, while exploring and modeling land cover (and land cover change) is in the interest of environmental scientists, for Planners, land use (and land use change) is one of the crucial research topics. Planning for land use in a fast-growing and uncertain urban context is difficult, as many factors influence how land is used and transformed. Fortunately, modern planners can rely on powerful

[6] Food and Agriculture Organization

tools such as geographic information systems (GIS) and planning support systems (PSS) that provide rich data and analysis capabilities for land use planning (Berke et al., 2006).

In this regard, the significance of exploring and modeling land use (and land use change) is threefold. On the one hand, it represents people's behavior and regulations (in the form of zoning and ordinances) in space. On the other hand, we can study the drivers of land use change (why land use changes). Finally, we look at the impacts of land use change (how and where) on urban structure, traffic, sprawl, housing, and economy, to name a few.

The reviewed literature has discussed several aspects of land use/land cover change depending on the research lens used. Since this research aims to study the dynamics, drivers, and impacts of land use change, in this chapter, we shed light on the following aspects of relevant literature:

- Theories of LULC change
- Models of LULC change
- Methods of modeling LULC change
- Drivers of LULC change
- Impacts of LULC change

2.1 Theories of LULC Change

It is a common practice for modelers to address underlying theories that can be used to characterize LULC change. Before discussing theories of land use change, we need to answer the question: What is a theory of land use change? In its simplest explanation, a land use change theory explains all aspects of land use change, including the dynamism of land use changes, the mechanisms of the land use change, and what causes the land use change (i.e., drivers of land use change). Even though it is difficult to find a theory that is explicitly being formulated for land use change, in this section, we try to shed light on theories that are explicitly or implicitly related to LULC change.

The most comprehensive study that covers most theories and models of LULC change is the work of Dr. Helen Briassoulis. In her web book "Analysis of Land Use Change: Theoretical and Modeling Approaches[7]," she distinguishes three major theorization traditions that implicitly or explicitly cover LULC changes. These theorization traditions are "urban and regional economics," "sociological and political economy," and "nature-society (or human-nature)."

We can rank these three theorization traditions based on the degree of implicit/explicit LULC change discussed in each tradition. While sociological and nature-society traditions discuss LULC change in a broader context, some urban and regional economic traditions explicitly (even though limitedly) discuss LULC change. In other words, theories in urban and regional economics theorization tradition discuss LULC change more explicitly, theories in the nature-society tradition discuss LULC change more implicitly, and theories in the sociological tradition are in the middle. We will briefly explain these three theorization traditions and some critical theories belonging to each. A summary of these theorization traditions and related theories that discuss LULC change is provided in Table 2-1 at the end of this section.

As its name implies, LULC change theories in urban and regional economics theorization tradition lean towards using economics concepts to explain the LULC changes. The cost of production factors, transportation cost, externalities, economies of scale, and utility are among those economic concepts. Micro-economic-based and macro-economic-based theories belong to

[7] The newer version (2020) of the book can be found https://researchrepository.wvu.edu/cgi/viewcontent.cgi?article=1000&context=rri-web-book

this tradition, which deals with the impact of individual consumer behavior and aggregate behavior on LULC change patterns, respectively. Von Thünen's theory of agricultural land rent and Alonso's [neoclassical economy] urban land market theory are among the most popular micro-economic-based theories. For macro-economic-based theories, on the other hand, we can name spatial equilibrium, regional disequilibrium, and Keynesian development theory (Briassoulis, 2000; Verburg et al., 2004).

Neo-classical economics is one of the most prominent economic theoretical models that explain the location of land use. This theoretical model, in which supply and demand play critical roles in the distribution of goods and services in the market, rests on three assumptions (Weintraub, 2002): A) People have rational preferences among outcomes; they can rank different scenarios according to their benefits and costs. B) Individuals maximize utility, or satisfaction, from their choices, and that firms maximize profits, or revenues minus costs. C) People act independently based on full and relevant information; they have access to all the data they need to make optimal decisions and that they are not influenced by external factors such as peer pressure or advertising.

Even though all three assumptions may apply to the location preference of each land use type, the assumption that individuals seek to maximize utility (For residential land uses) and firms look for maximizing profit (For commercial and industrial land use) plays a crucial role in the location of each activity in the urban setting. Among neoclassical economic theories that explain the location preference of land use types is William Alonso's theory of location and land use (Alonso, 1964), in which residential, commercial, and industrial land uses compete for location according to the bid rent curve and access to the central locations within the city.

According to Alonso's theory of location and land use, a spatial pattern of urban land use emerges from the trade-off between land consumption and accessibility. Households with high incomes and low demand for city center services can afford to live in spacious lots at the urban periphery, where land is cheaper. Conversely, households with low incomes and high demand for city center services have to settle for smaller lots near the urban core, where land is more expensive but also more accessible. This creates a competition for central land with commercial and industrial activities that also benefit from accessibility. To maximize their utilities, residential land uses, ceteris paribus, prefer to be located closer to the employment centers, which provide better access to jobs and minimize transportation costs[8]. As a result of these locational preferences, "the overall pattern of rents and land values appears to be shaped to a greater extent by access to jobs, and high densities of urban population occur in areas close to major job concentrations" (Hoover & Giarratani, 1984, p. 76).

On the other hand, commercial land uses are market-oriented because their transferable outputs are more valuable than their transferable inputs. To maximize their profit, firms try to maximize their market size and minimize travel costs achievable via access to a larger market by accessing more concentrated urban centers (i.e., CBDs). In other words, business firms are incentivized to locate with good access to their local suppliers and customers. As a result of this location preference, the rent gradients rise in the direction of markets, generally toward the center of the urban concentration (Hoover & Giarratani, 1984). As CBDs and concentrated urban centers become more desirable to firms and businesses, land rent increases (due to competition for land) and travel costs (due to traffic flow to access CBDs). This situation may negatively impact some location preferences; that is, firms (especially small firms and startups) can afford the cost of

[8] In reality, a whole host of factors impact the location preference of residential land use, including but not limited to topography, climate, availability of clean water and air, privacy, quietness, and aesthetic appearances.

rising land prices and transit costs to a certain point. Beyond the resistance point, they look for new places to take advantage of economies of scale through the concentration of businesses. This search for new places eventually led to new commercial centers.

Finally, Industrial land uses (specially manufacturing industries) are input-oriented; they would rather be close to the source of production than their output market[9]. Most industrial activities are not evenly distributed in space, but rather tend to form spatial clusters of different sizes. These clusters vary in the range of activities that they host, depending on their location, resources, infrastructure, and other factors. (McCann, 2013). They are more dependent to their suppliers, which operate in specific regions, rather than competing in broader markets; therefore, the price of land does not have a significant impact on their location choices (Hoover & Giarratani, 1984). Another factor that impacts the location preference of industrial land uses is regulation regarding their environmental impacts (e.g., air, water, and noise prolusion), which prevent them from being located close to highly concentrated urban residential and commercial areas.

In spatial economic studies, two intertwined theoretical frameworks explain the dynamics of land use change more explicitly: Economies of scale (aggregation economy) and spatial dependence. Alfred Marshall (1920) developed a theoretical framework for understanding the formation and growth of industrial clusters called economies of scale (aka, agglomeration economies), which has a lasting impact on the contemporary economic studies of agglomerations and clusters (Pászto et al., 2019). Economies of scale refer to the cost benefits that companies can obtain by expanding production and reducing expenses. This occurs because costs are distributed over a greater amount of goods. For instance, a business might achieve an economy of scale regarding

[9] There are exceptions for industries that produce perishable products (e.g., dairy, food, etc.), which, with today's advancements in cooling system technology, is not a big problem.

its bulk purchasing. Buying many products at once could bargain for a lower price per unit than its competitors. Economies of scale can be place-specific; if a large group of Activities are in the same place, this clustering will result in considerable investment at that particular location (McCann, 2013). It encourages the concentration of economic activities in specific locations, such as urban areas, industrial zones, or agricultural regions, where the benefits of scale economies are higher. This may lead to changes in land use patterns, such as urbanization, deforestation, or monoculture, with environmental and social impacts (Barmelgy et al., 2014; Bonye et al., 2021; Gnedenko, 2020).

According to Tobler's first law of geography, "Everything is related to everything else, but near things are more related (similar) than distant things." (Tobler, 1970, p. 236). This law helps understand and analyze spatial patterns and processes, such as climate, vegetation, population, and urbanization (Zheng et al., 2023). The implication of it in land use is that we expect neighboring parcels to have similar land use types and densities, which causes centralization of similar land uses in one area (e.g., CBDs). The CBD has different functions in the city, mainly related to high land value, concentrated similar land use, and easy accessibility, as it serves as a central marketplace, major transportation node, administrative center, and high-level producer services and command and control center (Kaplan et al., 2014).

Even though the principles of the neoclassical economy for location preferences of Residential and Activity land uses are still valid, some of these preferences may have changed because of new technologies in transportation, information, and communication[10]. These technologies reduce the distance effect, make urban activities more cost-effective, and allow urban activities,

[10] Dominance of online shopping, social media, telecommunications, telemedicine, and work-from-home opportunities (thanks to the Covid-19 lockdowns) are a few innovations that would impact the location preference for many land uses.

families, and firms to move to the land around cities. They also increase the sprawl by providing accessibility between Residential and Activity land uses (Kim et al., 2009).

The second theoretical tradition is sociological and political economy, which “emphasizes the importance of human agency, social relationships, social networks, and socio-cultural change in bringing spatial, political, economic, and other changes.” (Briassoulis, 2000, p. 83). The Chicago School of Human Ecology proposed an ecological perspective to explain the spatial structure of the emerging American industrial city. They drew parallels between the urban dynamics and the natural processes, such as the competition for space among different land uses. This led to the invasion and succession of the most attractive areas of a city by a more dominant activity, for example, the growth of the central business district (CBD) into the nearby transition zone (Pacion, 2005).

Contrary to the economic-based theorization tradition, in which all theories rely upon a form of economy and all the justifications are economic-based, theories in sociological and political economy theorization tradition rooted in the various disciplines including but not limited to Anthropology, Psychology, Political Science, Planning, and Geography. Therefore, theories in this category are expected to be diverse, making their categorization challenging. Human ecology, concentric zone, radial sector theory, and multiple nuclei theory are among the theories that fit in this theorization tradition.

The nature-society is the broadest and the most diverse tradition because “it embeds the analysis of land use change within the broader discourse on global environmental change” (Briassoulis, 2000, p. 110). In other words, theories in this tradition try to explain human beings’ impact on the environment, economy, society, and culture (which eventually leads to land use change) through, for example, deforestation, air and water pollution, and climate change. Cultural

Ecology, Environmental Psychology, and Environmental Determinism are some of the theories that belong to this tradition (Briassoulis, 2000; Verburg et al., 2004).

Even though many of the theories being discussed, among many more that are not being discussed, can be [somehow] related to land use and land use change, they suffer from the following drawbacks:

- Land use change needs to be viewed in a broader context of many other social, economic, environmental, and spatial disciplines; therefore, finding a theory that exclusively deals with land use change is difficult. This multi-disciplinary nature of land use makes it difficult for researchers from various disciplines to find common ground regarding covering all aspects of land use change.
- These theories are formulated around a single aspect of land use. For instance, Alonso's theory of location and land use aims to explain the impact of residential location on the urban structure. It disregards the [mutual] impact of other land uses and socioeconomic factors on the location preference of residential land use.
- Some of these theories are based on assumptions that are not the case anymore in the current modern era of communication. For example, three critical assumptions of Von Thünen's model of land rent are the isolation of the central city (i.e., no competing cities in the area), the flatness of land (i.e., there are no physical barriers), and the ability to transport goods in all direction (i.e., having access to roads in any direction). Even though Von Thünen's theory provided the basis for other theories later, these assumptions are no longer valid in the real world and need re-evaluation.
- Even though the spatial level of analysis (e.g., neighborhood, urban, regional, and national) plays a vital role in the elaboration of land use, in most theories (especially theories in

Sociological and Nature-Society theorization traditions), land use change is considered abstractly, and isolated from such considerations.

- Very few theories explain land use change; instead, they focus more on particular types of change like industrialization and urbanization. Alonso and Von Thünen are among the few theories explicitly mentioning land use change, even though they are limited to particular land use types.

Because of the multidisciplinary nature of land use change, it is difficult to theorize it based on a single theory or theorization tradition; therefore, we cannot argue that the land use change model approach for the DFW Metropolitan Area fits within a specific theory. However, we can claim that it contains more theories within "the urban and regional economics theorization tradition," with highlighted elements of the "neoclassical economic" theoretical model. A summary of these theorization traditions and related theories that discuss LULC change is provided in Table 2-1.

Table 2-1: Theories of land use land and cover change (adapted from Briassoulis, 2000)

Tradition	Approaches	Related Theories
Urban and regional economics theoretical approaches	Micro-economic theoretical approaches	• Agricultural land rent theory (Von Thünen, 1826) • Urban land market theory (Alonso, 1964) • Agent-based theories of urban and regional spatial structure • Economy of scale (aka agglomeration economy)
	Macro-economic theoretical approaches	• Spatial economic equilibrium theory (Weber, 1929; Losch, 1954) • Regional disequilibrium theories (cumulative causation theory, growth pole theory) • Keynesian regional theory • Development theories (Harrod-Domar models, export-based model, factor-based models)
	Other theoretical approaches in regional science	• Social physics • Urban and regional ecology • Spatial dependency
Sociological theoretical approaches to land use change	Functionalist-behaviorist theoretical approaches	• Human ecological theories (concentric zone theory, radial sector theory, multiple nuclei theory, sequence occupancy concept) • Planning theories (Foley's theory of metropolitan spatial structure, the urban place, and non-place urban realm theory) • Activity system theory
	Institutionalist-structuralist theoretical approaches	• Urban social movements • Urban land nexus theory • Crisis theory of land capitalism
	Core-periphery theories	• Modernization theories • Stage theory of economic growth • Core-periphery model • Internal colonialism • World system theory
	Unequal exchange and dependency theories	• Unequal exchange • Unequal development • Dependency theory
	Other sociological theories	• Unequal regional exchange • The theory of the spatial divisions of labor • Uneven development
Nature-society theories of land use change	Humanities-based theories	• Frontier theories • Environmental/cultural anthropology and geography (structuralism, cognitive anthropology) • Environmental psychology
	Natural science-based theories	• Environmental determinism • Cultural (or human) ecology • The Berkley school of geography
	Social science-based theories	• Culture of mass consumption theory • Ecological revolution • Multi-disciplinary approaches-ecological equilibrium concept

2.2 Models of LULC Change

Land use modeling has evolved through three main stages (Meyer et al., 2000). The first stage took place in the 1960s and early 1970s when various methods of land use modeling were explored as part of the broader effort to develop comprehensive, long-term planning models using mainframe computers. The Lowry model was the most influential of these methods in the field. The second stage of land use modeling emerged in the 1970s; they were mainly large-scale, aggregate, mainframe-based simulation models that mimicked the changes in the urban system over time using discrete time steps. These models improved the first-generation models by incorporating a more explicit microeconomic theoretical basis and a more complex database and modeling structure. The third and current stage of operational models began to appear in the 1980s, based on the experiences gained in the previous two decades and taking advantage of the advances in computer hardware and software that have been happening continuously since the introduction of the microprocessor and the personal computer in the early eighties.

The models of land use change should be able to address at least one of these three questions: where (the locations where land use change happened), when (the rate of the change in different periods), and why (the contributing factors of land use change). Depending on the model's capability to address one or all these questions, Lambin et al. (2000) provided four categories of LULC change models: empirical-statistical models, stochastic models, optimization models, and dynamic simulation (aka process-based) models.

In its most straightforward account, empirical-statistical land use change models use multiple linear regression to explain the causes of LULC change. Stochastic models aim to calculate the probability that a LULC category changes to another category based on a "sample of transitions occurring during some time intervals" (Lambing et al., 2000, p. 324). Monte Carlo simulation and Markov Chains are abundantly used in stochastic models (Hägerstrand, 1968; Thornton et al., 1998). Developed mainly based on Von Thünen's theory of land rent, optimization models of LULC change use linear programming (at the microeconomic level) or general equilibrium model (at the macro-economic scale) to optimize land use change. Lastly, dynamic simulation models have been developed to simulate interactions between biophysical and socioeconomic processes (as the drivers of land use change) by simplifying the complex land use change system into simple mathematical equations.

Founded on seven factors of purpose, underlying theory, spatial scale, explicitness, land use types, temporal dimension, and modeling techniques, Briassoulis (2000) shed light on models that implicitly or explicitly deal with LULC change. These models are statistical and econometric models, spatial interaction models, optimization models, integrated models, and other modeling approaches.

Another broad classification of LULC models is descriptive and prescriptive models (Bockstael et al., 2000; Briassoulis, 2000; Kaimowitz et al., 1998; Lambin, 1997; Miller et al., 1999). Descriptive models aim to explain and simulate the function of land use at its current status. Prescriptive land use models, on the other hand, model the future status of land use based on the past and current trends in the land use status.

Having descriptive/prescriptive groups of land use models in mind, Verburg et al. (2004) provide theories, rationale, and implementations of the following factors in the classification of LULC change models: level of analysis, cross-scale dynamics, driving factors, spatial interaction, temporal dynamics, and level of integrations. The level of analysis deals with the micro- or macro-level of land use change analysis. Two examples in this class of models are micro-economic and macro-economic models. Cross-scale dynamics refer to the scale (local, regional, and national) of land use change analysis. Drivers of land use change touches on the exogeneity and endogeneity nature of land use change drivers, including socioeconomic, biophysical, and proximate factors. Spatial interaction considers if the LULC change model includes interaction between objects (i.e., land use type succession in a parcel) and the effect of its neighbors into account; CA models fit in this category. Temporal dynamic alludes to the inclusion and significance of time in the LULC change model. Eventually, the level of integration discusses how comprehensive a land use change model is and includes a variety of aspects of land use in the modeling. UrbanSim and Anjomani's Integrated Land Use-Transportation Model are classified as comprehensive LULC change models.

Overall, the microeconomic effects of different factors on the location decisions of firms can be analyzed using various models. For instance, the Hotelling model suggested that firms will cluster spatially only when they do not compete on price but on spatial factors. On the other hand, behavioral models of uncertainty indicated that spatial clustering could be a rational strategy to cope with spatial competition under uncertain and heterogeneous information conditions. The optimal location of a firm can vary depending on the transport rates and the production relationships. The firm may be located at an intermediate point or near the market or the suppliers (McCann, 2013).

Reviewing LULC change models reveals the complexity and interdisciplinary nature of land use and related models. Disciplinary background and the purpose of the study play critical roles in picking the base for the classification of these models. That is a well-accepted classification base for LULC change models that meaningfully reflect all classification factors that do not exist. As a result, there is no consensus on LULC change model classification.

Considering the complexity of land use change model classification, this dissertation research will use a descriptive spatial statistics model, which combines statistics and space (i.e., spatial interaction between neighboring units of analysis) to model the dynamics of land use change in the DFW Metropolitan Area. We test various aspects of land use change models (i.e., level of analysis, spatial interaction, and temporal dynamics) as described by Verburg et al. (2004) and mentioned above.

Also, since the data for land use and drivers of land use change are available in a panel data format (time-space cross-section) from 1990 to 2020 (seven 5-year intervals are: 1990, 1995, 2000, 2005, 2010, 2015, 2020) at census block and block groups, we will apply a variant of panel data regression model called Spatial Panel Data Regression (SPDR) to model the drivers of land use change and spatial interaction in the DFW Metropolitan Area. This part of the model helps us to understand the important factors of land use change.

2.3 Methods of LULC Change

Several modeling approaches can be used to model land use changes on the metropolitan scale. Each of these modeling approaches has limitations and does not satisfy the expectations of land use change research in explaining all aspects of land use change. The methods used to evaluate LULC change are Remote Sensing, conventional statistics, cellular automata (C.A.), or a combination of these methods. In this section, we will review the literature explaining these methods, how they are being used to model LULC change, and the limitations/drawbacks of each method in modeling land use (not land cover) change.

Remote sensing is the most commonly used approach for modeling LULC change, which provides powerful tools to study urban environments, urban growth modeling, and project socio-economical, environmental, and ecological effects of urban development (Maimaitijiang et al., 2015). This method is more common among researchers in areas other than urban planning. Remote sensing has been used for land use change modeling since the early 1970s when the first Landsat satellite was launched; however, recently, researchers have been trying to look for the drivers of land use change as well. For instance, Maimaitijiang et al. (2015) studied the spatial and temporal dynamics of urban growth in the St. Louis metropolitan region over the last 40 years based on remote sensing imagery and socioeconomic factors. Lo et al. (2002) integrated Landsat images and census data to analyze Atlanta Metropolitan Area land use/land cover changes. Overmars (2003) used aerial imagery data to model land use change at a multi-scale level to determine the correlation between land use change and the scale of analysis. The common theme of these studies is that the following steps have been taken to model land cover changes (Anderson et al., 1976):

1. Extract land cover from aerial/satellite images taken at different intervals.
2. Differentiate images to identify changes in land cover in those time intervals.
3. Examine the drivers of change (i.e., the factors that cause those changes) in each period.
4. Use conventional statistics (i.e., regression) to measure the effect and significance of each driver on the land cover change in the form of regression coefficients.
5. Use calculated regression coefficients in the previous step to predict the future of the developed area (regardless of the land use type) in the study area.

Remote sensing modeling techniques for LULC change have two drawbacks; therefore, they cannot be used for modeling land use change. First and foremost, they model land cover change, not land use change. With recent progress in remote sensing techniques and advanced extraction algorithms, the land cover can be extracted with a high degree of certainty. Since land use results from human activity on the land, satellite images and aerial photos cannot capture those activities; hence, it is impossible to extract land uses by remote sensing data extraction techniques. Also, in RS-based modeling techniques, the unit of analysis is either an individual pixel or some aggregation of pixels, called regions, rather than an individual parcel level, which is the very nature of every land use analysis and modeling approach (Li et al., 2014).

In the form of regression (i.e., logit, linear, non-linear), Conventional Statistics is another method in LULC changes. In this modeling approach, the objective is to run the regression to find the significance of the effect of independent variables (physical, socioeconomic, proximity, etc.) on land use change as the dependent variable. To model the relationship between LULC change, a most common approach is to apply global level statistics like Ordinary Least Square (OLS).

These statistical methods rely on two main assumptions: there is no correlation between model variables (Pearson correlation is mostly used to capture such a relationship), and the variance of the error terms is constant across all levels of the predictor variables (aka homoscedasticity). This means that we assume the natural and social characteristics are spatially homogeneous, while, in fact, they are not; such characteristics are constantly changing over time and space. Furthermore, these data often present some magnitudes of spatial autocorrelation. Also, global-level statistics cannot reveal the changing relationships over space (Maimaitijiang et al., 2015).

Mertens et al. (2002) described three common statistical problems in spatial explicit regression analysis for LULC change as multicollinearity between factors of land use change, spatial autocorrelation between various land use types[11], and endogeneity, in which an explanatory variable is correlated with the error term[12]. Empirical-statistical models could be misinterpreted due to these problems they face. (Mertens et al., 2002 as cited in Mitsuda et al., 2011).

To sum up, the problem with conventional statistics in modeling land use change is that they do not directly incorporate space (proximity, area, connectivity, and/or other spatial relationships) into their mathematics (Briassoulis, 2000).

The third commonly used LULC change methods are Cellular Automata (CA) based methods. CA, in its very conventional approach, is a cell-based modeling method in which the state of each cell (as the unit of analysis) evolves according to its neighborhood to other cells and a simple transition rule defined by the researcher (Koomen et al., 2011; Singh, 2003). In other words, the state of a cell depends on its initial state, the states of its neighboring cells, and a set of rules that determine how the state changes over time. (Verburg et al., 2004). Since its evolution in the 1940s in computer science and its widespread applications in spatial modeling in the 1970s (Pinto, 2015), many variants of CA models have been developed in spatial modeling. These models include but are not limited to SLEUTH CA, fuzzy CA, ANN CA, MCE CA, Multi-CA, Statistic-Based CA, and stochastic CA (Singh, 2003).

Each variant of the CA models has its applications and has been used in different situations, but they have common drawbacks. First, Transition rules are applied to the entire region, which means that the interactions occurring at different scales are not significantly; however, land use change factors impact the land use change at several spatial scales. Second, CA models usually

[11] Spatial autocorrelation in land use change is a concept that describes the degree of similarity or dissimilarity between land use types in neighboring locations. It reflects the spatial patterns and processes of land use change, such as clustering, dispersion, diffusion, or contagion.
[12] This means that the variable is not truly independent of the outcome and may be influenced by some unobserved factors that also affect the outcome.

consider external drivers of change (like accessibility) are considered as external attributes of cells, without taking into account they are interdependent. Third, CA models are a popular tool for modeling land use change at a large scale, such as regions or metropolitan areas. However, land use change is influenced by multiple factors and processes that operate at different spatial scales, from neighborhoods to localities to regions. Therefore, CA models need to account for the interactions and feedback among these scales to capture the complexity and dynamics of land use change (Pinto, 2015). Finally, the unit of analysis is a pixel, not a parcel, as is the case for RS techniques discussed earlier—however, land use changes at the parcel level with various shapes, areas, and sizes.

The fourth LULC change approach, which is becoming more common nowadays, is to take advantage of combining GIS and Statistics, called spatial statistics. Thanks to GIS, Geodatabase (which enables us to keep historical spatial data, including land use data), and various spatial statistics software/tools, there are efforts (even though in the infancy stages of development) to combine statistics and GIS to model land use change to find the land use distribution patterns and the causes and effects of these changes. These spatial statistics methods contain unique statistical tools to analyze patterns, relationships, and trends of phenomena that occur in space (Anselin, 1988).

Tepe (2023) developed a novel approach to analyze land use change based on dynamic spatial panel data, which accounts for spatiotemporal dependencies for a continuous variable. This approach reveals new insights (spatial autocorrelation, spillover effects, heterogeneity across regions, and temporal lag effects) into the patterns and drivers of urban growth in Florida from 2010 to 2019, which achieved high accuracy and can predict future urban growth.

While there may be similarities between spatial and conventional statistics regarding concepts and objectives, spatial statistics differ from conventional statistics in that they are designed for geographic data. They take into account spatial concepts such as distance, area, and neighborhood in their calculations, unlike non-spatial methods that ignore them. For instance, Geographically Weighted Regression (GWR) can capture spatial heterogeneity and relationship variations and account for spatial autocorrelation (Maimaitijianga et al., 2015). Spatial statistics models cannot only include various aspects of space in modeling but also include time (Burkey, 2018). Depending on what aspect of space/time is included in the model, several variations of spatial statistics are explained in Chapter 4.

Considering the drawbacks of other land use change methods and capabilities and advantages of spatial statistics, which are applicability to parcel level data, incorporating space concepts directly into the model, and statistical analysis capacity in modeling spatial phenomena, they are better techniques to model land use change. Therefore, this research uses a variation of spatial statistics methods called Spatial Panel Data Regression (SPDR) to model land use change in the DFW Metropolitan Area.

2.4 Drivers of LULC Change

Land use change is the result of contributing drivers, as one of the three pillars of a land use change theory; these factors may vary in time and space (Lo et al., 2002). For instance, technology may significantly impact land use change in the 21st century[13], while its effect was minimal (if any) in the early 20th century. Also, the impact of technology on land use change in a Metropolitan Area

[13] For instance, the Internet provides a telework option for many white-collar workers who can live far away from work without commuting. Consequently, it may change locational preference for residential and activity land uses.

in a developed country is different from its impact on land use change in a similar Metropolitan Area in a developing country.

Even if the driving factors of land use change are similar from one place/period to another, their significance might differ. The significance of the population on land use change as a contributing factor being used in most land use change studies (Banzhaf et al., 2013; Brueckner et al., 1983; Deng et al., 2008; Ghosh et al., 2017; Lo et al., 2002; Maimaitijiang et al., 2015; Mitsuda et al., 2011; Overmars et al., 2003) might be different in the DFW Metropolitan Area than in the LA Metropolitan Area. Alternatively, the significance of population on land use change in the DFW Metropolitan Area in 1990 might differ from its significance in 2015.

With this changeability nature of the drivers of land use change, the question is, what are the driving factors of land use change in the DFW Metropolitan Area? We must review the literature from different perspectives, including time, place, discipline, modeling technique, and data availability to answer this question. To quantify the relations between land use change and the deriving forces of these changes, researchers use statistical methods, mainly regression, based on historical data on land use change. Most of these approaches describe historical land use conversions as a function of the changes in drivers and location characteristics (Verburg et al., 2004).

According to Berke et al. (2006), the concept of land use change is affected by three main factors from the planner’s point of view:

- Actions of developers who respond to the real estate market demand.
- Community values and interests that seek to maintain and enhance the quality of life.
- Plans, policies, decisions, capital investments, and regulations of the government that aim to

manage the development of the community.

McHarg (1969), in his book "Design with Nature," explores the relationship between human civilization and the natural environment, arguing that humans should design and plan their settlements and activities in harmony with nature rather than imposing their will on it. He provides a framework and methodology for ecological planning based on understanding the natural processes and systems that shape the landscape by presenting several case studies that illustrate the application of ecological planning in different contexts, such as urban development, watershed management, coastal protection, and climate adaptation. Even though his work does not focus specifically on land use change, his methodology framework includes physical factors that could impact land use change under various time/space circumstances. These factor categories along each measure are:

- Climate: air pollution and tidal inundation.
- Geology: features with unique, scientific, and educational value; and foundation conditions.
- Physiography: features with unique, scientific, and educational value, land features of scenic value, water features of scenic value, riparian lands of water features, beaches along the bay, surface drainage, and slope.
- Hydrology: marine commercial craft, marine pleasure craft, fresh water, stream-side recreation, watersheds for stream quality protection, aquifers, and aquifer recharge zones.
- Pedology: erosion, soil drainage, foundation conditions.
- Vegetation: existing forest, forest type, existing marshes.
- Wildlife: existing habitats, intertidal species, water-associated species, field and forest species, and urban-related species.

- Land use: features of unique educational and historical value, features of scenic value, and existing and potential recreation resources.

McHarg's methodology is based on overlaying different layers of environmental data to determine the suitability of land uses. His method is unsuitable for planning in areas with existing development since it does not consider social and economic factors that may affect land use decisions. Overall, his method is too simplistic and idealistic and needs to be updated to address the complexities and challenges of contemporary planning (Daniels, 2019).

Marsh (1983) provides a comprehensive and practical approach to dealing with environmental problems associated with land planning, landscape design, and land use. His approach and methodology apply several factors to illustrate how landscape planning can be applied to various environmental issues and challenges. These factors are topography, slopes, soil, wastewater disposal systems, groundwater (aquifer), watersheds/ drainage nets, streamflow/floodplains/flood hazard, wetlands, habitat, water quality, runoff, stream sedimentation, watersheds; riparian landscape (streams, channel forms, and valley floors); coastal landscape (shoreline systems, landforms); solar climate near the ground; microclimate, climate change; ground frost, permafrost, and vegetation. As for McHarg's methodology, Marsh's does not consider enough socioeconomic and proximity factors influencing land use change (Yang et al., 2021).

Maimaitijiang et al. (2015) provided spatial and temporal dynamics of urban growth in the St. Louis (STL) Metropolitan Region over 40 years based on land cover extracted from Landsat images and socioeconomic data of population, race, and housing units. They integrated remote sensing and census data using global OLS and local GWR statistical methods to analyze the spatial and temporal patterns and trends of urban growth in the STL. They concluded that urban sprawl positively correlates with population change during the study period (i.e., 1970 – 2010).

Also, it is positively correlated with population changes in outer suburbs and negatively correlated in the central city and inner suburbs. These changes eventually led to the growth of vacant houses in the central city and inner suburbs and environmental and racial segregation.

To study the impact of scale on spatial autocorrelation between drivers of land use change, Overmars et al. (2003) applied Moran's I model in Ecuador. Their study grouped the drivers of land use change into bio-geophysical and socioeconomic. The bio-geophysical factors include soils with texture, soils with slope, soil fertility, altitude, and total annual precipitation. Socioeconomic data, on the other hand, are distance to the urban center, distance to roads, distance to rivers, total population, rural population, urban population, total population living in poverty, the rural population living in poverty, total illiterate population, rural illiterate population, total population working in agriculture, and rural population working in agriculture. They applied these data sets in three grid scales (1x1, 3x3, and 6x6) to examine the scale's impact. They concluded that the Moran's I increased with higher aggregation levels; the bigger the cell size, the more the Moran's I index.

Another LULC change model in the Metropolitan scale is the work of Lo et al. (2002), which is one of the key studies that combines spatial and socioeconomic data to not only model LULC change but also measure the impact of scale on LULC change. They used a 60-meter cell size MSS[14], TM[15], and ETM+[16] Landsat images and census data to analyze LULC changes in the Atlanta Metropolitan Area from 1973 to 1999. To evaluate the impact of spatial scale on land use change, they used 17 factors at the county and census tract level as their unit of analysis. These factors are high-density urban use, low-density urban use, cropland, forest, grassland, and water

[14] Multi-Spectral Scanner
[15] Thematic Mapper
[16] Enhanced Thematic Mapper Plus

bodies (as dependent variables), greenness, fragmentation, elevation, slope, population density, per capita income, proximity to urban centers, proximity to shopping malls, proximity to highways, and proximity to transit nodes (as independent variables). They have found a larger number of correlation coefficients that were more statistically significant at the census tract level than at the county level, supporting the impact of the modifiable areal unit problem. Even though they considered "high-density urban use" as commercial, industrial, transportation, and infrastructure, and "low-density urban uses" as single-family and multi-family residential land uses, they analyzed the impact of physical and socioeconomic factors on the land cover change as their classification of dependent variable shows.

Mitsuda et al. (2011) focused on empirical–statistical models (in which regression models were usually applied for the spatial distribution data of LULC change) and reviewed factors that affect land use change patterns (Table 2-2). In their in-depth literature review of LULC change factors, they have listed the factors that could impact deforestation and the conversion of agricultural lands to the developed area, regardless of the type of development (i.e., land use type). In other words, they summarized the drivers of land cover change even though many factors (primarily socioeconomic factors) also contribute to land use change.

The paper concludes that "almost all recent studies investigating [LULC] pattern is based on remote sensing image processing data" (Mitsuda et al., 2011, p. 122), which supports the findings in other research being reviewed in this research. Also, to overcome some of the limitations of empirical–statistical models (i.e., multicollinearity, spatial autocorrelation, and endogeneity) and for more effective implications derived from empirical–statistical models, they recommended "…to reveal the changes in LULC using time-series data and to relate them to economic, demographic, and social development" (Mitsuda et al., 2011, p. 123).

Table 2-2: Factors affecting land use change pattern (Mitsuda et al., 2011, p. 119)

Category	Sub-category	Variable
Socioeconomic factors	Accessibility	• Distance to road • Connectivity to Market • Distance to Settlements
	Development of Local community	• Population • Labor • Technology
	Spatial Configuration	• Distance to LUC front • Surrounding land use • Fragmentation
	Political Restriction	• Protection
Natural/Environmental factors	Topography	• Elevation • Slope
	Productivity	• Soil • Climate

Banzhaf et al. (2013) shed light on the consequences of demographic changes on land use changes and the related pressure on the urban environment in the Santiago Metropolitan Area in Chile. In their study, which concentrates on the urban and suburban dichotomy and the different impact that each factor has on these two, they picked population, population density, percentage of built-up area, population density per built-up area, the proportion of new built-up area per flood-hazard zone, percentage of public green spaces, percentage of public green spaces per flood-hazard zone, and public green spaces per capita as their drivers behind land use change in urban and suburban areas. Their research aimed “to detect and evaluate the physical structure and composition of urban areas, especially built-up and green spaces” (Banzhaf et al., 2013, p. 180). They have seen a decrease in urban population figures and density versus an increase in suburbanization processes. On the other hand, according to the public green space indicators, the urban development process is not linear or simple, but rather complex and multifaceted. It involves both high-density and low-density developments in various directions, creating a diverse and dynamic urban landscape (Banzhaf et al., 2013).

Muth-Mils' theory of urban sprawl explains urban sprawl from an economist's point of view, believing that income, population, agricultural rent, and commuting costs play the most significant role in urban sprawl (Mills, 1972; Muth, 1969). To prove this theory empirically, Brueckner et al. (1983) applied census 1970 data to 40 U.S. urbanized areas. In their research, the dependent variable is the area of urbanized land (in square miles); the independent variables are population, income, median agricultural land value per acre in counties with urbanized area, public transit commuters, and automobile ownership (the last two variables measure commuting costs). They conclude that the economist view of urban sprawl is justifiable, meaning that "urban sizes are the result of an orderly market equilibrium where competing claims to the land are appropriately balanced" (Brueckner et al., 1983, p. 479).

In an attempt to find the effect of establishing a single industrial activity on large-scale land use change, Ghosh et al. (2017) modeled the changes in land use before and after a new Toyota plant was established in Kentucky. The change in land use resulting from the new factory is compared in two situations: after the Toyota plant came to the area (actual) and what may have occurred without Toyota in the region (estimate). For the actual land use change, land cover maps (provided by Landsat TM satellite images) for 1989, 1992, 1998, and 2006 are used. For the estimate of land use changes in the region without Toyota (simulate), they applied a CA-Markov[17] model, which needs "suitability maps for every land use category to determine the location suitable for land use change likely to occur" (Ghosh et al., 2017, p. 292). To fulfill the

[17] "CA-Markov models are spatially explicit and hybrid LU change models that integrate CA for computing spatial dynamics and Markov Chain Analysis for computing temporal dynamics of LU change" (Ghosh et al., 2017).

CA-Markov suitability maps requirements, the following factors are used as the inputs for the Multi-Criteria Evaluation (MCE) approach: proximity to roads, schools, urban areas, lakes, and employment locations, employment density, population growth, population density, per capita income, slope, and hydrology.

As mentioned earlier, there have been recent attempts to model land use change using spatial statistics and panel data regression. Deng et al. (2008) used a three-period panel data of satellite imagery and socioeconomic data from 1980 to 2000 to analyze the extent and the driving factors of urban expansion in China. Land use data, extracted from satellite remote sensing data provided by the U.S. Landsat TM/ETM images, consists of six land cover classes: cultivated land, forestry area, grassland, water area, built-up area[18], and unused land. The dependent variable is the urban core area (in hectares); the independent variables are GDP (income), population density, agriculture land investment, highway density, port distance, province/county capital, rainfall, slope, temperature, and elevation. Since the scope of the study is the country of China, some of their variable (e.g., rainfall and temperature) are applicable in such large-scale analysis. Even though they examined the impact of all independent factors, they concluded that income plays the most significant role in urban expansion in China.

The most recent use of panel data regression at the parcel level is the work of Tepe (2023) and Tepe et al. (2017), in which spatiotemporal approaches are applied to model urban growth (in the form of the ratio of the number of developed parcels). By applying dynamic spatial panel data to

[18] Their data set includes three classifications of the built-up area (Deng et al., 2008, p. 101):

- Urban core: defined as all built-up area that is contiguous to urban settlements.
- Rural settlements: all built-up area in small towns and villages.
- Other built-up area: roads, mines and development zones that are not contiguous with the urban core.

a balanced spatial panel data at the block level in Florida between 2010 and 2019, they investigated and found their significant impact on urban growth:

- Site-specific Factors: average lot area in a block group and the variance of lot area in the block group.
- Neighborhood Factors: share of the single residential parcel, the share of recreational area, and the share of agricultural land.
- Socioeconomic Factors: population density, the ratio of non-white population, the ratio of students, the employment ratio, the ratio of finance units, and the ratio of workers commuting by car.
- Transportation Factors: distance to roads and annual average daily traffic (AADT).

Regardless of the scale of the analysis (i.e., parcel, neighborhood, city, region, or country), most of the cases being reviewed in this section deal with urban growth and land cover (not land use) change, in which the modeling topic is how other land cover types (e.g., agriculture, forest) are changing toward the developed area and impacting urban growth. From a land use change point of view, the missing point is that they do not mention which land use within the developed area changes to another land use. Finally, several geographic, socioeconomic, and biophysical factors are considered drivers of LULC change, denoting the diversity of factors of land use change in time and space.

Considering the literature being reviewed, data availability, and time/space considerations, a list of the theoretical-empirical drivers of land use change, the theoretical basis for each driver, and underlying assumptions for each theory are provided in Table 2-3. For each driver of land use change, the name of the theory is mentioned; otherwise, if it is hard to relate the driver of land use change to a specific theory, the theoretical tradition[19] is mentioned. The explanations provided for each driver of land use change are not necessarily related to the corresponding theories. The list and description of drivers of land use change in the DFW Metropolitan Area, along with the selection process, is provided in Chapter 5 of this research.

Table 2-3: Theoretical-Empirical drivers of land use change

Driver	Supporting Theories	Explanation
Topography	Urban and regional economics	Slope and elevation are correlated with the difficulties of constructing urban infrastructure and buildings. High altitude and steep slopes increase the cost of construction; hence, they are less desirable for development.
Proximity to lakes and water reservoirs	Von Thünen, Weber, Christaller, Alonso	Lakes and reservoirs are attractive for development, especially for residential land uses, because people can take advantage of having access to cleaner air and scenic views of their living location.
Soil quality	Nature-Society	Suitable soils are for agricultural purposes; they are less desirable for development because they are reserved lands and prohibited for further development other than agricultural activities.
Air quality	Nature-Society	Areas with cleaner air and less noise in the suburbs are more attractive to residential land uses.
Proximity to highways	Von Thünen, Weber, Christaller, Alonso	By building highways, travel becomes faster and easier, which lowers the cost of commuting and allows people to live in cheaper houses in the suburbs. This increases the demand for suburban areas, reduces the cost of commuting, and causes urban areas to expand physically.
Proximity to transportation nodes	Von Thünen, Weber, Christaller, Alonso	Transportation nodes, where two or more highways intersect, provide higher accessibility; therefore, commercial and industrial land uses tend to be close to transit nodes.
Highway Density	Von Thünen, Weber, Christaller, Alonso	Since there is a mutual relationship between accessibility and physical growth, census tracts with a higher density of highways attract more growth.
Proximity to landmarks	Urban and regional economics	Landmarks attract more people, especially visitors and tourists. Therefore, commercial land uses prefer to be closer to these landmarks.
Distance from CBDs	Von Thünen, Weber, Christaller, Alonso	Locational preference for residential, commercial, and industrial land uses is based on bid rent curve and access to the city center.
Population density	Urban and regional economics	Ceteris paribus, more population, is an engine for new development, especially for commercial and

[19] A theoretical tradition contains a broader range of similar theories.

		residential land uses.
Means of Transportation to Work	Urban and regional economics	A high value of public transit use indicates a high commuting cost, which could result from a mono-centric metropolitan area in which opportunities are not evenly distributed. A high value of automobile use is associated with ease of automobile usage (reflecting low congestion level); it would indicate a low commuting cost, which could result from an automobile-dominant, multi-centric metropolitan area.
Travel time	Von Thünen, Weber, Christaller, Alonso	Longer travel time could indicate congested roads or a low-density (vs. compact) metropolitan area. Congested roads during peak hours could mean a mono-centric metropolitan area because people drive toward opportunities during morning peak hours and drive toward home during evening peak hours. On the other hand, a low-density metropolitan area could have its opportunities dispersed all over the metropolitan area, which could indicate a multi-centricity.
Proximity to public transit facilities	Urban and Regional Economics	Public transit, in general, and rail-based public transit play a critical role in land use change. It provides good accessibility to opportunities, especially for low- and middle-income groups. Rail stations are good TOD opportunities, including high-rise residential apartments, commercial land uses, and mixed uses.
Employment	Sociological & political economy; Urban and regional economics	Higher employment in a census tract means more jobs in that census tract, which indicates more commercial and industrial land uses in the census tract.
Land value	Von Thünen, Weber, Christaller, Alonso	Higher land value means more opportunities (i.e., commercial land uses) in the census tracts.
Household income	Urban and regional economics	Income positively correlates with automobile ownership and commuting costs, eventually impacting households' location.
Policy factors	Nature-society, Sociological & political economy	Land use change does not always occur only by natural and socioeconomic factors; sometimes, political decisions in establishing a specific establishment could have a domino impact, eventually leading to land use change. These policies impact zoning and related ordinances, ultimately impacting location preferences for various land uses.

2.5 Impacts of LULC Change

There is no question that land use change has environmental and socioeconomic impacts in a Metropolitan Area, which could cause further land use change (cyclical impact). Wu (2008) and Briassoulis (2000) categorize the impact of land use change into socioeconomic and environmental impacts. Briassoulis argues that the environmental impacts of land use change (e.g., soil erosion and degradation, water quality and quantity, reducing biodiversity, increasing greenhouse gas emissions, etc.) are being discussed and got more attention due to the visibility of its impacts. Socioeconomic impacts of land use change (housing, income, education, well-being of people, etc.) are more complex and become visible in the long run.

Socioeconomic and environmental impacts of land use change are interdependent, meaning one affects the other and vice versa. For example, environmental impacts can lead to socioeconomic effects, such as loss of income or livelihoods, triggering more land use change and creating a cycle of degradation and poverty (Briassoulis, 2000).

From the socioeconomic perspective, Wu (2008) analyzes how land use change affects food and timber production, water and soil conservation, open space preservation, rural communities, housing affordability, and private property rights. From an environmental standpoint, he examines how land use change influences water pollution, soil degradation, carbon emissions, climate change, and biodiversity loss.

The first economics-based land use theory, proposed by Von Thünen, applied Ricardo's idea that profit will lead to reinvestment in land use change. Alonso extended this theory by adding land use suitability and a bid-price curve for households or firms. Sinclair further developed Von Thünen's theory to account for urban sprawl (Koomen et al., 2011). According to the Urban Institute, local land use regulations that are overly restrictive can limit economic opportunities for workers and narrow the housing supply, affecting housing prices and affordability. One of the factors that affects the affordability of housing is the regulation of land use and density. When reforms are implemented that limit the amount and type of development that can occur on a given parcel of land, they tend to increase the median rent and reduce the supply of units that are within the reach of middle-income renters. This is because such reforms reduce the profitability and feasibility of building new housing, especially multifamily housing, and create artificial scarcity in the market. (Stacy et al., 2023).

Another impact of land use change in a Metropolitan Area is that it causes urban growth and sprawl, which is defined as "the rapid expansion of the geographic extent of cities and towns, often characterized by low-density residential housing, single-use zoning, and increased reliance on the private automobile for transportation" (Rafferty, 2023). Urban sprawl, conversely, influences the accessibility and mobility of urban residents (via increased automobile dependency), the distribution and equity of urban amenities and opportunities, and the well-being of urban populations (via reducing physical activity levels) and exacerbates spatial social segregation.

One economic theory closely related to the impacts of land use change on urban growth and sprawl is the growth pole theory, outlined by Perroux in the 1950s. He argued that a specific pole (or cluster) of economic activity attracts more growth than other areas due to its regional potential, key infrastructure facilities (e.g., transportation network), and agglomeration economies, which leads to an unbalanced regional development. However, this unbalanced growth may also stimulate the creation of secondary growth poles, which can eventually enhance regional economic diversity. (Rodrigue, 2020).

To assist urban planners in comprehending the factors that drive land use change, Deng et al. (2010) examined how the urban core area was influenced by economic growth. Their analysis revealed the complex and dynamic interactions between the spatial and economic dimensions of urban development, and concluded that land use change can have various economic impacts on urban growth, such as affecting the costs and benefits of providing public services and infrastructure, the productivity and competitiveness of urban sectors, and the value and use of land resources. For example, low-density and dispersed urban development can be more cost-intensive than compact development patterns.

Hoover et al. (1984) discuss problems related to changes in spatial patterns of urban activities in four categories: declining activity in CBDs, transportation, urban poverty, and the fiscal disparity between central cities and their surrounding suburbs. They relate these issues to underlying changes in land use, location, or locational advantage. Therefore, Land use change has a cause and effect on many aspects of a Metropolitan Area. It can lead to changes in the spatial distribution of urban activities and the location of urban infrastructure, which can lead to the creation of new urban areas or the expansion or shrinkage of existing ones.

With an emphasis on transportation, Meyer et al. (2000) summarized some of the transportation impacts on land use and land use-related issues. These impacts are:

- Creating the first suburbs.
- Decentralizing urban areas.
- Reinforcing the downtown role.
- Population and employment growth.
- Increase in land value near highways.
- Higher property value near rail stations.
- Impact on urban growth and urban form.
- Expansion of cities along the rail corridors.
- Increase in the value of lands close to expressways.
- Increase in population density near subway stations.
- Development of employment centers in inner suburbs.
- Increase in residential construction and highway expansion.

- Lower residential densities in relation to major employment centers.
- New home buyers in the commuter rail service area (new residential developments).
- Development of land from one part of the region to another part (relocation of development).
- Increase in ridership due to mixed-use developments and accessibility to transit services.
- Major urban development at highway interchanges results from market conditions and financing arrangements.
- Consistent growth and raising property value in corridors with better accessibility and service to public transit services.

Overall, and regardless of the categorization of the impacts, land use change has a range of positive and negative impacts. Some positive impacts include environmental improvement and restoration, enhanced social interaction, economic growth and development, and improved access to services and amenities. Some negative impacts of land use change are social inequality and exclusion, climate change and greenhouse gas emissions, environmental degradation and pollution, excessive land consumption and loss, and increased traffic congestion that could cause urban sprawl (Nuissl et al., 2021).

Based on the literature reviewed in this section, one can discuss the impacts of LULC change at three scales: global, regional, and local (Briassoulis, 2000). In this research, we are interested in the local impacts of land use change, in which the way land is used can shape how cities grow and shrink, resulting in different degrees of compactness, dispersion, fragmentation, and complexity. These degrees affect how efficient, livable, and sustainable urban areas are. For example, compact cities may have lower greenhouse gas emissions and higher social cohesion, while dispersed cities may have more traffic congestion and lower accessibility. We will discuss the impact of land use change in the DFW Metropolitan Area on the relocation of Activity and Residential land uses, which could lead to the formation of new activity and residential hubs, traffic distribution, and overall structure of the Metropolitan Area.

Chapter 3: Methodology

3.1 Introduction

Research methodology is a systematic plan that outlines the approach, techniques, and interpretation of data gathered by a researcher to resolve a research problem. It is a blueprint for conducting research that ensures reliable and valid results that address the research purpose. The methodology chapter in a dissertation (thesis or research paper) explains what and how the research is done. It includes the type of research conducted, how data was collected and analyzed, what tools or materials were used, how research biases were mitigated or avoided, and why these methods were chosen.

To fulfill this mission, in this chapter, we introduce the study area, the unit of analysis, data sources, extraction and harmonization techniques, data analysis process and methods, and modeling software.

3.2 Study Area

The study area is the “DFW Urban Area” within the DFW Metropolitan Statistical Area (DFW-MSA). It is designed based on the census-defined urban area of the Dallas-Fort Worth Metropolitan Area (DFW Metropolitan Area). It contains 128 physically connected cities, towns, and census-designed places (CDPs). Since the unit of analysis in this research is census boundaries 2020 (as discussed in the next section of this chapter), the boundary of the study area is modified to fit the census boundaries.

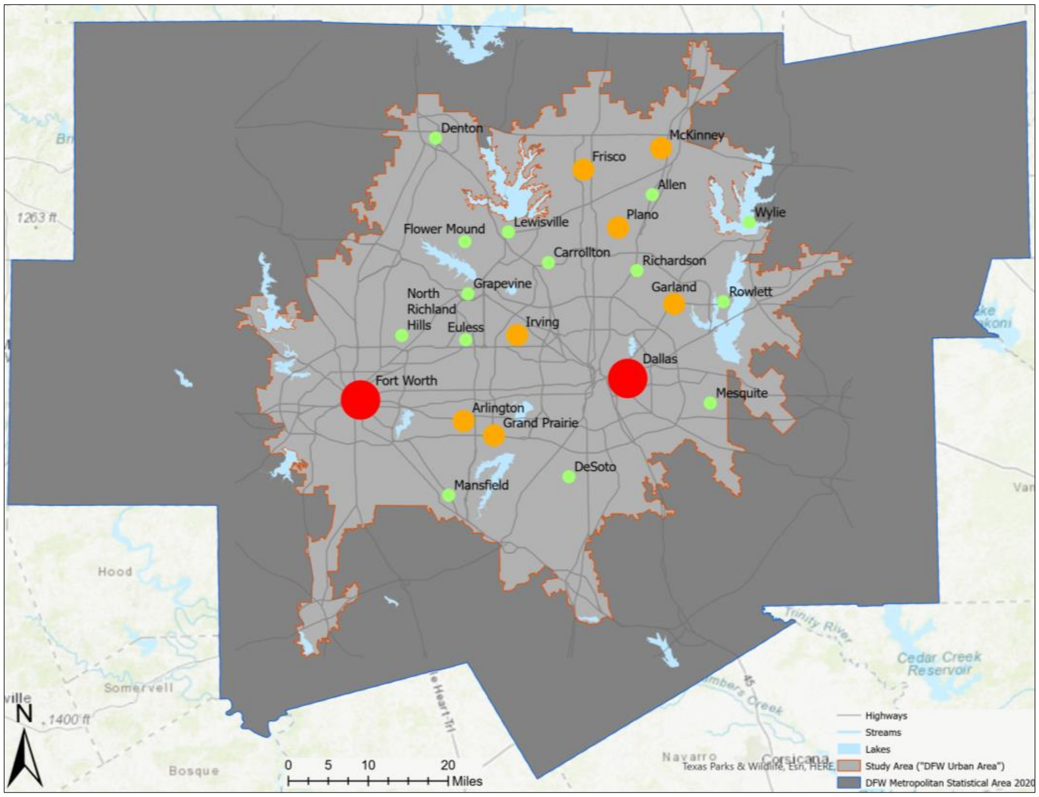
Encyclopedia Britannica defines the Metropolitan Area as “a major city with its suburbs and nearby cities, towns, and environs over which the major city exercises a commanding economic and social influence" (Britannica, 2023). Therefore, this study area was named DFW Urban Area within the DFW Metropolitan Area because of the close social, economic, and interaction ties between its entities; it does not bear any political or statistical definitions.

According to the 2020 census, the population and the area of the DFW-MSA are more than 7.6 million and 9,000 square miles, respectively. However, the population and area of the study area are more than 6.7 million and 2,740 square miles, respectively. Table 3-1 shows cities, towns, and CDPs, and Map 3-1 shows the study area in the DFW-MSA and major cities with a population of more than 50,000 people.

In 2020, the area of the developed land area (which includes all land use, except roads, cemeteries, farmland, flood control, landfill, parks/recreation, ranch, timberland, vacant, and water bodies) in the DFW Urban Area is 1,385 square miles (more than 50 percent of the study area). The Area of the Activity and Residential land uses (which this research is looking to model their changes) is 375 and 930 square miles, respectively, about 48 percent of the study area.

Table 3-1: Cities, towns, and CDPs within the study area

Name	Population 2020	Area (Sq Mi)	Name	Population 2020	Area (Sq Mi)	Name	Population 2020	Area (Sq Mi)	Name	Population 2020	Area (Sq Mi)	Name	Population 2020	Area (Sq Mi)
Addison	16,661	4	Dalworthington Gardens	2,293	2	Hebron	803	1	Melissa	13,901	12	Royse City	13,508	18
Allen	104,627	26	Denton	139,869	97	Hickory Creek	4,718	4	Mesquite	150,108	49	Sachse	27,103	10
Anna	16,896	16	DeSoto	56,145	22	Highland Park	8,864	2	Midlothian	35,125	64	Saginaw	23,890	8
Argyle	4,403	12	Double Oak	3,054	2	Highland Village	15,899	6	Mobile City	142	1	Sansom Park	5,454	1
Arlington	394,266	99	Draper	33	1	Hurst	40,413	10	Murphy	21,013	6	Savannah	6,529	1
Azle	13,369	9	Duncanville	40,706	11	Hutchins	5,607	9	North Richland Hills	69,917	18	Seagoville	18,446	19
Balch Springs	27,685	9	Edgecliff Village	3,788	1	Irving	256,684	68	Northlake	5,201	17	Seis Lagos	1,450	1
Bartonville	1,725	7	Euless	61,032	16	Joshua	7,891	9	Oak Leaf	1,552	2	Shady Shores	2,764	3
Bear Creek Ranch	1,787	1	Everman	6,154	2	Justin	4,409	3	Oak Point	4,357	5	Southlake	31,265	22
Bedford	49,928	10	Fairview	10,372	9	Keene	6,387	5	Ovilla	4,304	6	St. Paul	992	1
Benbrook	24,520	11	Farmers Branch	35,991	12	Keller	45,776	18	Paloma Creek	3,177	1	Sunnyvale	7,893	17
Blue Mound	2,393	1	Fate	17,958	12	Kennedale	8,517	7	Paloma Creek South	9,539	1	Talty	2,500	3
Burleson	47,641	28	Ferris	2,788	4	Krum	5,483	2	Pantego	2,568	1	The Colony	44,534	16
Carrollton	133,434	37	Flower Mound	75,956	45	Lake Dallas	7,708	3	Parker	5,462	9	Travis Ranch	7,324	1
Cedar Hill	49,148	36	Forest Hill	13,955	4	Lake Worth	4,711	2	Pelican Bay	2,049	1	Trophy Club	13,688	4
Celina	16,739	32	Forney	23,455	15	Lakeside	1,649	2	Plano	285,494	72	University Park	25,278	4
Cleburne	31,352	35	Fort Worth	918,915	343	Lakewood Village	635	1	Princeton	17,027	10	Watauga	23,650	4
Cockrell Hill	3,815	1	Frisco	200,509	69	Lancaster	41,275	33	Prosper	30,174	25	Waxahachie	41,140	51
Colleyville	26,057	13	Garland	246,018	57	Lantana	10,785	2	Providence Village	7,691	2	Westlake	1,623	7
Combine	2,245	8	Glenn Heights	15,819	7	Lavon	4,469	3	Rendon	13,533	25	Westover Hills	641	1
Coppell	42,983	15	Grand Prairie	196,100	81	Lewisville	111,822	43	Richardson	119,469	29	Westworth Village	2,585	2
Copper Canyon	1,731	5	Grapevine	50,631	36	Lowry Crossing	1,689	3	Richland Hills	8,621	3	White Settlement	18,269	5
Corinth	22,634	8	Hackberry	2,973	1	Lucas	7,612	16	River Oaks	7,646	2	Wilmer	4,974	8
Cross Roads	1,744	7	Haltom City	46,073	12	Mansfield	72,602	37	Roanoke	9,665	7	Wylie	57,526	37
Crowley	18,070	8	Haslet	1,952	11	McKinney	195,308	68	Rockwall	47,251	30	**Total**	**6,738,365**	**2,740**
Dallas	1,304,379	385	Heath	9,769	12	McLendon-Chisholm	3,562	13	Rowlett	62,535	21			



Map 3-1: DFW-MSA, study area (DFW Urban Area), and cities with population > 50,000

3.3 Unit of Analysis

As mentioned in Chapter 1, this research aims to analyze the dynamics, drivers, and impacts of land use change in the DFW Metropolitan Area. The unit of analysis applies only to the dynamism and drivers of land use change. In this regard, this research is being conducted at two levels of analysis: census block and census block group.

The dynamics of land use development (and consequently the land use change) happen at the parcel level (Tepe, 2023); however, historical land use data for the study area is available at the block level. Therefore, the dynamic of land use change is being analyzed and modeled at the census block level.[20] There are 103,220 blocks within the study area.

[20]Census blocks are like urban blocks (at least within cities' boundaries), which are "defined as the space within the street pattern of a city that is subdivided into land lots for the construction of buildings" (Ghisleni, 2023).

On the other hand, most of the historical socioeconomic data (mostly retrieved from Census and ACS data) is available at the block group level. Therefore, the unit of analysis (i.e., entities or individuals) for drivers of land use change is the census block group; other factors (site-specific, proximity, and natural) are aggregated at the block group level to match the unit of analysis for socioeconomic data. There are 4,016 block groups within the study area.

Since we are modeling the distribution of land use change from the past to the present, it is critical to select an appropriate unit of analysis that can be applied to all periods and make logical and valid comparisons possible. On the other hand, the census boundaries change from one census period to another. Therefore, we picked census block and block groups 2020 (the most recent census boundaries available) as the basis of analysis, and the data for the previous periods are apportioned accordingly, as explained later in section 3.5 (data harmonization) of this chapter.

3.4 Data Source

The data source in this dissertation research has four major components: land use (site-specific), proximity, traffic, and socioeconomic data. In this section, we will provide the data source for these categories of data and, when needed, explain how we calculated, extracted, or aggregated the data (including missing data).

3.4.1 Land Use Data

GIS-ready land use data from 1990 to 2020 are provided by the North Central Texas Council of Governments (NCTCOG) in five-year intervals of 1990, 1995, 2000, 2010, 2015, and 2020, in the following categories:

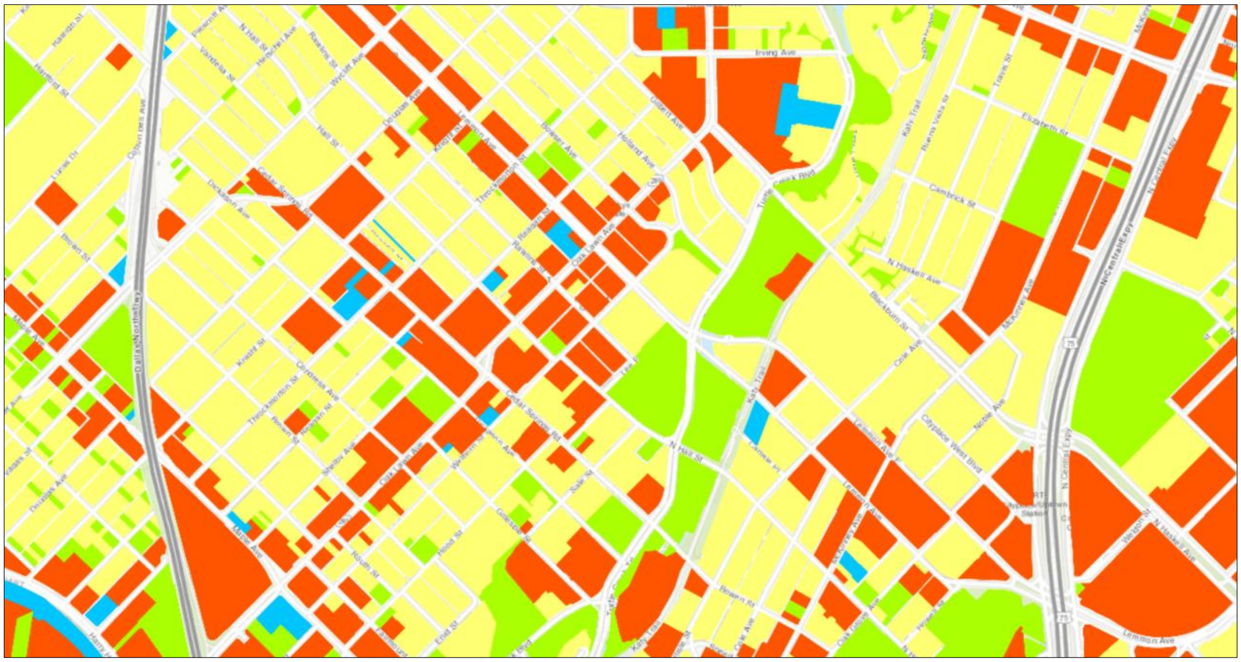
Education
Hotel/Motel
Industrial
Institutional/Semi-Public
Office
Retail
Venue
Airport
Communication
Parking
Railroad
Runway
Transit
Utilities
Cemeteries
Flood Control
Improved Acreage
Landfill
Parks/Recreation
Small Water Bodies
Vacant
Water
Group Quarters
Mobile Homes
Multi-Family
Single Family
Residential Acreage

However, since we are modeling the distribution of two broad categories of land uses, as explained previously, these land uses are categorized as "Activity," "Residential," "Vacant," "Infrastructure," and "Miscellaneous," as presented in Table 3-2. Map 3-2 shows a sample of land use data for the year 2020, which is aggregated in the census block.

Table 3-2: Land use categories for the research

	Land Use Category				
	Activity	**Residential**	**Vacant**	**Infrastructure**	**Miscellaneous**
Land Use	Education	Single family	Vacant	Airport	Cemeteries
	Hotel/motel	Multi-family		Communication	Flood control
	Industrial	Mobile home		Parking	Improved acreage
	Office	Group quarters		Railroad	Landfill
	Retail	Residential acreage		Runway	Parks/recreation
	Venue			Transit	Small water bodies
	Institutional/semi-public			Utilities	Water

Among these five land use categories, in this research, we are interested in three categories: Activity, Residential, and Vacant. Activity and Residential land uses are used to analyze the dynamics of land use change (Chapter 4). Also, they are considered dependent variables for the Activity Model and Residential Model, respectively; Vacant land (the area of vacant land in each block group) is a significant independent variable for the land use change model (Chapter 5). We discussed these land use categories, and how we used them extensively in the research, in Chapters 4 and 5.



Map 3-2: An excerpt of the land use map aggregated based on land use type in each block.

3.4.2 Proximity Data

Proximity data include the distance from several factors (e.g., activities centers, major airports, roads, lakes, rivers, etc.), calculated as Euclidean (straight line) distance from each objective. The base data for proximity analysis is also retrieved from NCTCOG; then, the distance is calculated and tabulated on the block group level in ArcGIS Pro.

3.4.3 Traffic Data

Traffic data is provided by the Texas Department of Transportation (TxDOT). The traffic stations collect short-term traffic count data to produce Annual Average Daily Traffic (AADT). Traffic counts are collected on an annual basis as a means of measuring the use of public roads in the state. AADTs are calculated using a volume count, axle factor, and seasonal factor. A general equation overview is as follows:

$$\text{AADT} = \text{Vehicle Axles} \times \text{Axle Factor} \times \text{Seasonal Factor}$$

3.4.4 Socioeconomic Data

Socioeconomic data is derived from the Census Bureau (both centennial census and ACS-5-year estimate), retrieved from Social Explorer and NHGIS. In this regard, for 1990 and 2000, we used centennial census data. For 2010, 2015, and 2020, the ACS (5-year estimate) was used.[21] For 1995 and 2005, socioeconomic data is not available at the block group level; therefore, for these two times, the Annual Average Growth Rate (AAGR) between lower and upper values (1990 and 2000 are references for 1995; 2000 and 2010 are references for 2005) are calculated. Then, the target years (1995 and 2005) are estimated based on the correspondent AAGR:

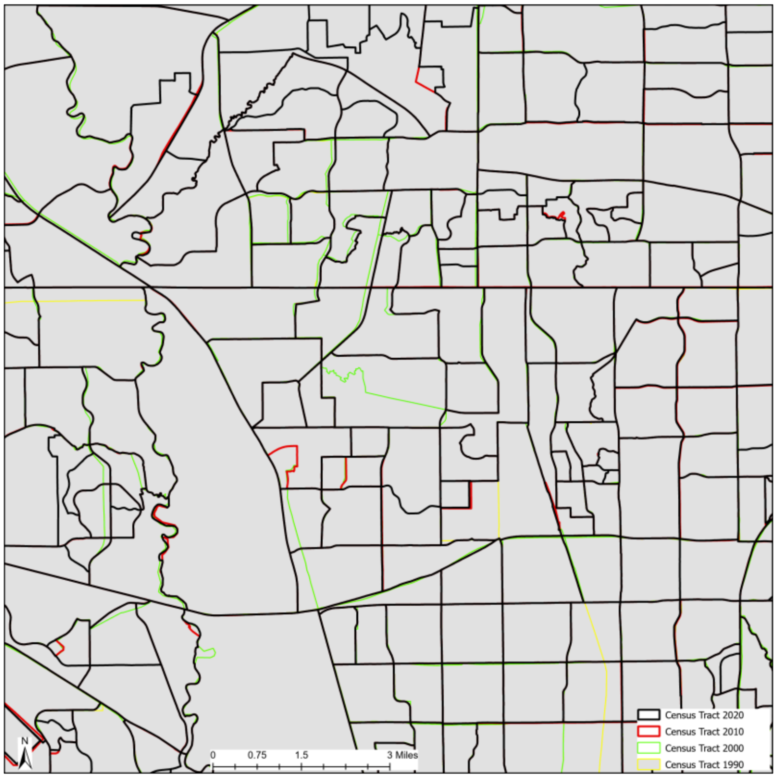
$$\text{AAGR} = (\text{Ending value} / \text{Beginning value}) - 1$$

3.5 Data Harmonization

In research involving panel data, which includes census data from several periods and individuals, the main obstacle is the changes in geographic statistical boundaries (e.g., block, block group, tract, etc.) in different census periods; it occurs because areas may split and/or merge from one census year to another (Deng et al., 2008). Map 3-3 shows an example of changes in the census tract in the study area from 1990 to 2020.

In such research, the same boundaries must be used for all periods to make the modeling and analysis consistent and logical. Converting data from different boundaries to one boundary is called data harmonization (TIBCO, n.d.). In spatial modeling involving various boundaries, there are several approaches to harmonizing data: upward harmonization, downward harmonization, interpolation, and apportioning. A brief description of each method is provided in the following sections, and the details of the method used for this research (i.e., apportioning) are explained.

[21] ACS replaced the long-form questionnaire used in the decennial censuses until 2000. ACS started in 2006, and the first 5-year estimate became available in 2009, covering 2006-2009.



Map 3-3: Boundary change in census tracts (1990, 2000, 2010, and 2020)

3.5.1 Upward Harmonization

Upward harmonization is helpful when the unit of analysis is the bigger unit (e.g., census tract), but the data is in the smaller unit (e.g., census block group). In this approach, the data from the smaller unit in one census year (e.g., 2000) is aggregated toward the larger unit in another year (e.g., 2010). Figure 3-1 shows the process of aggregating data from block groups 2000 to census tracts 2010; block groups are converted to centroids (i.e., the center point of each polygon-block group) with all associated attributes, then attributes are aggregated to the census tracts.

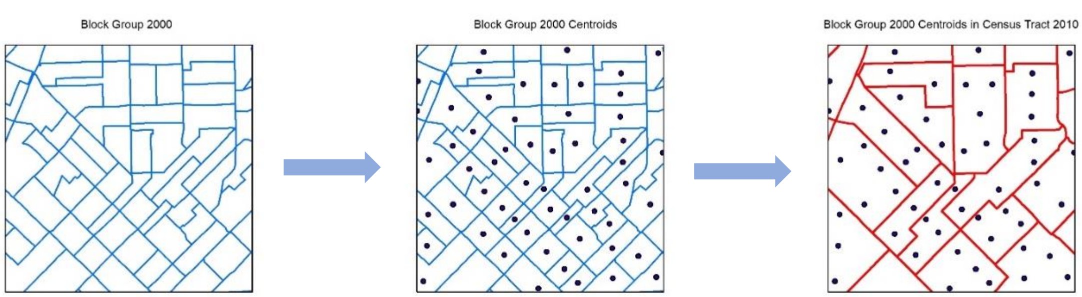


Figure 3-1: Upward harmonization to aggregate population in block groups 2000 toward census tracts 2010

3.5.2 Downward Harmonization

Downward harmonization is helpful when the unit of analysis is the smaller unit (e.g., block group), but the data is in the larger unit (i.e., census tracts). In this approach, the data from the larger unit in one census year (e.g., 2000) is aggregated toward the smaller unit in another year (e.g., 2020). Figure 3-2 shows the process of aggregating socioeconomic data at the census tract 1990 toward blocks 2010.

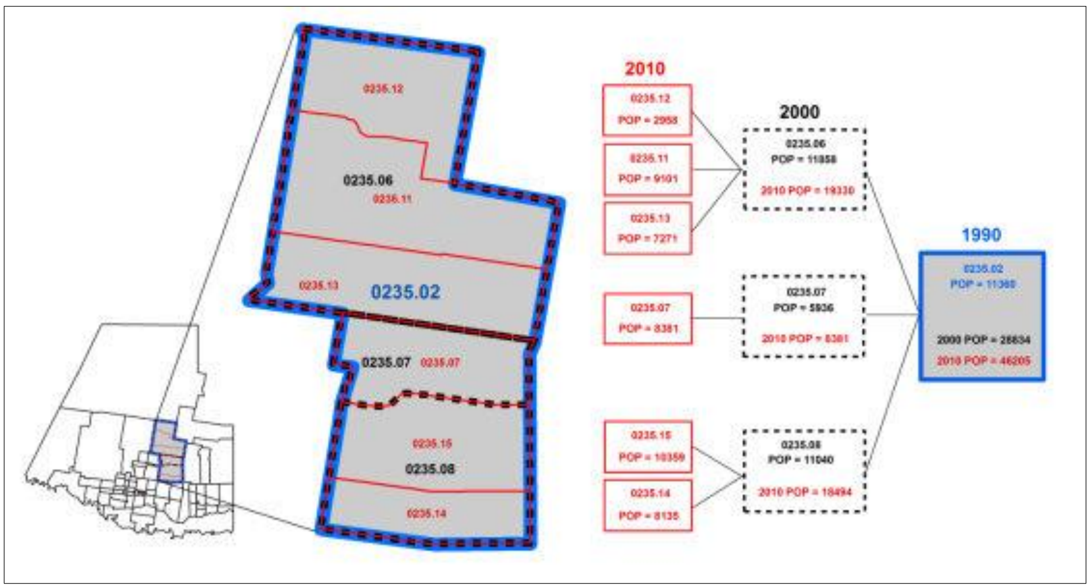


Figure 3-2: Downward harmonization for census tracts 1990 to census blocks 2010 (credit: Peña, 2012)

3.5.3 Interpolation Method

In this method, the new value for each polygon (e.g., a census block) can be estimated by interpolating centroids of surrounding polygons (Logan et al., 2014). The advantage of this method is that data from the surface can be aggregated to any desired unit of analysis. However, the downside is that the data, like population characteristics, are not likely to fit a smooth surface (Logan et al., 2014). Figure 3-3 shows the process of using interpolation to assign population in census tracts 1990 to census blocks 2010.

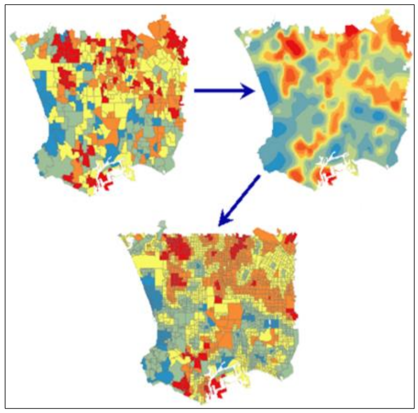


Figure 3-3: Interpolation method to assign population in census tracts 1990 to census blocks 2010.

3.5.4 Apportioning

Apportioning is useful when units of analysis and socioeconomic data are at the same level (e.g., both are block groups) but from different periods. In this method, the boundaries from one year (usually the year that most of the data is based on; otherwise, the most recent year) are set as the unit of analysis. Then, the data from other periods are proportionately summarized to the base unit according to the area of the block group, weighted by the area of residential (or any other land use type depending on the nature of analysis and the variable being analyzed) in the block group, as depicted in Figure 3-4.

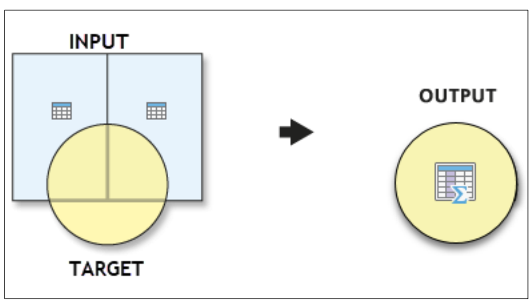


Figure 3-4: Apportion polygon (credit: Esri)

In this research, as discussed in Chapter 5, we will apply a Spatial Panel Data Regression (SPDR), for which the unit of analysis is the census block groups. Also, socioeconomic data (as regression model variables) is collected at the census block groups but at different times (1990, 1995, 2000, 2005, 2010, 2015, and 2020). Therefore, the Apportioning method is used to harmonize data from different periods to the census block groups 2020. To do so, the data from other periods are proportionately summarized to the block groups in 2020 based on the block group area weighted by the area of residential area in the block group at the origin time (e.g., 1990).

3.6 Modeling Software

In this research, we used three software and related packages: ArcGIS Pro is used to prepare spatial data, visualization, and modeling the dynamics of land use change (using spatial statistics Toolbox). R and R-Studio (Packages: PLM, SPDEP, and SPLM) are used for programming Spatial Panel Data Regression (SPDR), including variable tests, model selection, and model execution. GeoDa is used for calculating neighborhood, which is later used in the SPDR.

3.7 Data Analysis

There are different research methodologies, including quantitative, qualitative, and mixed methods (Streefkerk, 2023). Qualitative methods are used for collecting and analyzing non-numerical data (e.g., words, images) to understand the meanings and experiences of the participants. These research methodologies are mostly used to explore new topics, gain insights into poorly understood phenomena, or interpret complex social realities. Some examples of qualitative methods are interviews, observations, focus groups, and literature reviews (Mcleod, 2023).

Quantitative methods, on the other hand, involve collecting numerical data and analyzing them using statistical techniques. Quantitative methods are often used to test hypotheses, measure variables, identify patterns, and make predictions. Some examples of quantitative methods are surveys, experiments, content analysis, and meta-analysis (Mcleod, 2023; Streefkerk, 2023).

The choice of qualitative or quantitative methods depends on the research question, the purpose of the study, the resources available, and the researcher's preferences. Sometimes, a mixed-methods approach that combines qualitative and quantitative methods can be helpful to address complex research problems.

As mentioned in Chapter 1, this research aims to explore the dynamics, drivers, and impacts of land use change in the DFW Metropolitan Area. For dynamics of land use change, we used spatial statistics techniques of Global Moran's I for spatial autocorrelation, Getis-Ord Gi* for hot spot analysis, and Anselin Local Moran's I for cluster/outlier analysis. For drivers of land use change, on the other hand, we used Spatial Panel Data Regression (SPDR) to model the factors that drive land use change. Finally, for impacts of land use change, we used the results from the spatial statistics and the Spatial Panel Data Regression (SPDR) to discover the impacts of land use change on urban growth (sprawl), traffic flow, and urban structure. For the relationship between the distribution of traffic flow and land use change, we used the QQ plot to assess the similarity between the two variables.

Therefore, in this dissertation research, we are dealing with a mixed methodology that includes (mostly) quantitative statistical methods to model land use change and qualitative methods to interpret the results of the quantitative methods. To fulfill the research's mission, we summarized all model variables (i.e., land use, socioeconomic, proximity, site-specific, and natural factors) at census blocks 2020 (for dynamics of land use change) and census block groups 2020 (for drivers of land use change). Figure 3-5 shows the overall data analysis process and modeling land use change as presented in this research (Chapters 4 - 6).

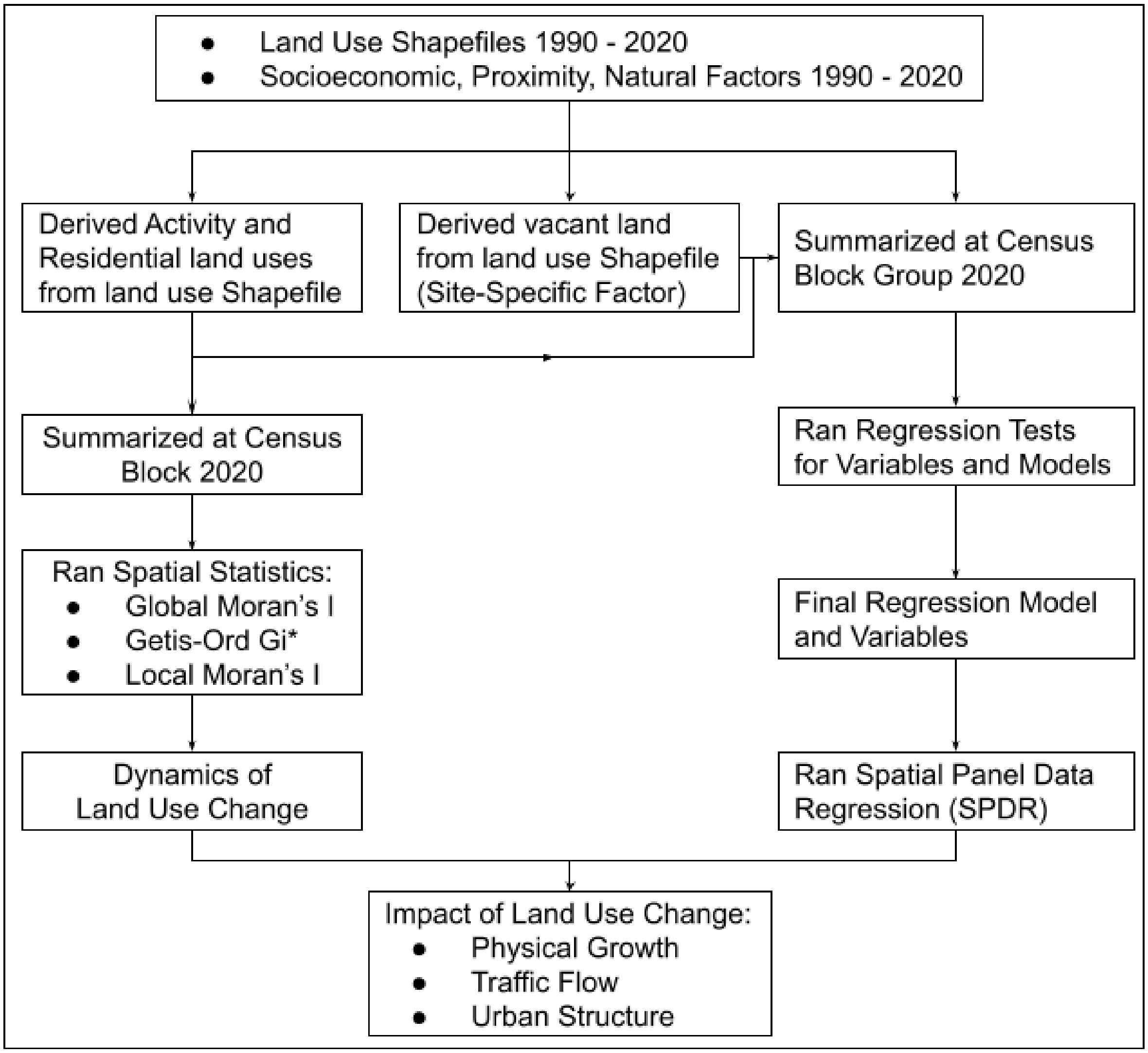


Figure 3-5: Schematic process of data analysis and modeling

Chapter 4: Dynamics of Land Use Change in DFW Metropolitan Area

4.1 Introduction

Human activities often transform natural landscapes or alter the management of lands already dominated by humans. These activities significantly impact the planet's surface and the spatial patterns that emerge and evolve. How resources are distributed across space and how economic activity is located and shapes spatial patterns is called land use, and its changes over time are called land use dynamics, which are the subject of spatial economics studies (Camacho et al., 2015). As discussed in Chapter 2, in spatial economic studies, two intertwined theoretical frameworks are essential to land use dynamics: Economies of Scale (aka Agglomeration Economy) and Spatial Dependency.

Alfred Marshall, in the 1920s, developed a theoretical framework for understanding the formation and growth of industrial clusters called economies of scale (agglomeration economies), which has a lasting impact on the contemporary economic studies of agglomerations and clusters (Pászto et al., 2019). Economies of scale refer to the cost benefits that companies can obtain by expanding production and reducing expenses. This occurs because costs are distributed over a greater amount of goods. For instance, a business might achieve an economy of scale regarding its bulk purchasing. Buying many products at once could bargain for a lower price per unit than its competitors. Economies of scale can be place-specific; if a large group of Activities are in the same place, this clustering will result in considerable investment at that particular location (McCann, 2013). It encourages the concentration of economic activities in specific locations, such as urban areas, industrial zones, or agricultural regions, where the benefits of scale economies are higher. This may lead to changes in land use patterns, such as urbanization, deforestation, or monoculture, with environmental and social impacts (Barmelgy et al., 2014; Bonye et al., 2021; Gnedenko, 2020).

On the other hand, according to Tobler's first law of geography, "everything is related to everything else, but near things are more related (similar) than distant things." (Tobler, 1970, p. 236). Known as the first law of geography, it describes the spatial autocorrelation of phenomena, meaning objects close to each other tend to be more similar or related than far apart things. This spatial dependency concept helps to understand and analyze spatial patterns and processes, such as climate, vegetation, population, and urbanization (Zheng et al., 2023). The implication of it in land use is that we expect neighboring parcels to have similar land use types and densities, which causes centralization of similar land uses in one area (e.g., CBDs). The CBD[22] has different functions in the city, mainly related to high land value, concentrated similar land use, and easy accessibility, as it serves as a central marketplace, major transportation node, administrative center, and high-level producer services and command and control center (Kaplan et al., 2014).

The emergence of the Internet (which enabled and emboldened distant communication), social media, online shopping, Internet of Things-IOTs (which makes device-to-device communications possible), and the Covid-19 outbreak (which made remote work an option not only during the pandemic but also afterward) have impacted these two theories' implications. In fact, because of these changes, Tobler recently updated his theory to consider the time factor and changed it to everything related to everything else, but near and recent things are more related than distant (both space and time) things. This most recent theory is a principle that describes phenomena's spatial and temporal autocorrelation (Amgalan et al., 2022; Griffith et al., 2018).

[22] According to Kaplan (2014), The CBD is a more accurate name for "downtown," but in large cities, the CBD can include several downtowns and uptowns, and the whole area or areas can cover a few square miles.

As mentioned in Chapter 1, one aspect of this research's goal is to study the dynamics of land use change in the DFW Metropolitan Area; in this chapter, we analyze this aspect of the research purpose by answering the following research questions to test the research hypothesis that the economies of scale and spatial dependency are weakening in explaining the distribution of land use change:

- How have the Activity and Residential land uses distribution patterns changed over time?
- Where are the new hot spots of Activity and Residential land uses?
- Which hot spots grow, and which ones decline in size?

Even though this mission can be done via map visualization, simply using map visualization is subjective in that by changing the classification of data presentation, a map can present different information about the same data. Spatial statistics helps us minimize maps' subjectivity by mapping the patterns. To examine this spatial dependency or similarity, we use three spatial statistics methods called "Spatial Autocorrelation," "Hot Spot Analysis," and "Cluster and Outlier Analysis."

First, we will discuss the spatiotemporal distribution pattern of Activity and Residential land uses at the macro level by analyzing the land use change trend in the overall DFW Urban Area, distance from downtown Dallas and Fort Worth, individual cities, and counties that the study area encompasses. It helps us to understand the impact of the scale of analysis on land use change. Then, we will use several spatial statistics methods (i.e., spatial autocorrelation, hot spot analysis, and cluster and outlier analysis) to examine land use change at the block level[23]. Finally, we will cross-examine the change at various times to determine the dynamics of land use change.

[23] The land use data until 2000 is available at the block level; also, there are computation limitations to process the data at the parcel level. Therefore, we aggregated the data at the block level for all periods.

4.2 Spatiotemporal Distribution Pattern of Land Use at Macro Scale

The analysis of Activity and Residential land uses in the DFW Urban Area between 1990 and 2020 shows that while the developed land area grew by 77 percent, Residential land uses increased by 80 percent, and Activity land uses grew by 83 percent. As Figure 4-1 shows, Residential land uses grew at a higher rate between 1995 and 2010, then slowed down afterward. On the other hand, Activity land uses grow steadily at a slower pace, even though they have a higher growth rate than Residential land uses. As Map 4-1 shows, the distribution of these two land use types is not similar; some areas gained more Residential land uses, and others attracted more Activity land uses based on physical, proximity, and socioeconomic factors that impacted the distribution of such land uses.

To explore the distribution of Activity and Residential land uses over time and space, we considered four levels of analysis: the overall study area, county, cities, and distance from downtown. For the overall pattern of land use change, we used the cut-fill method. For county, city, and the distance from downtown, however, the percentage of Activity and Residential land uses change from one period over the previous period is calculated.

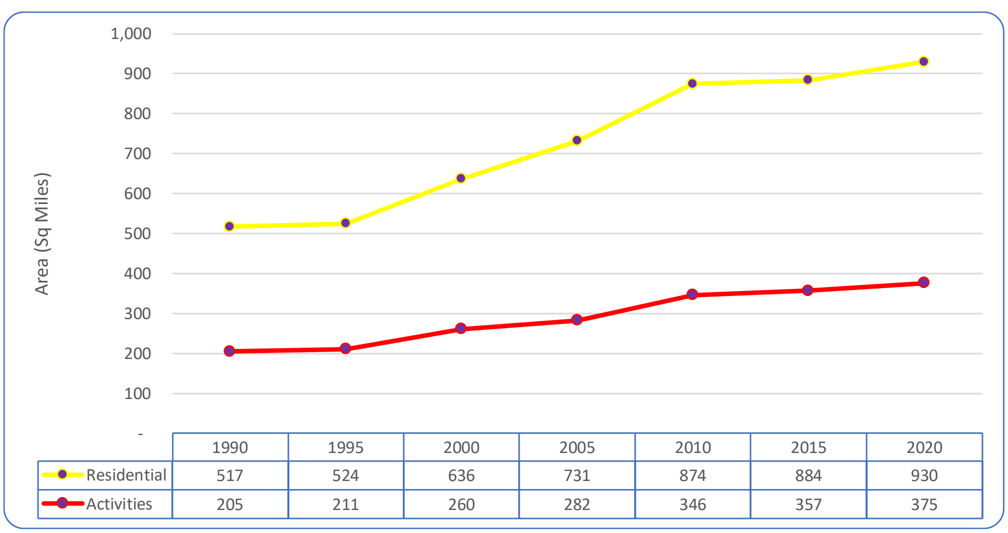
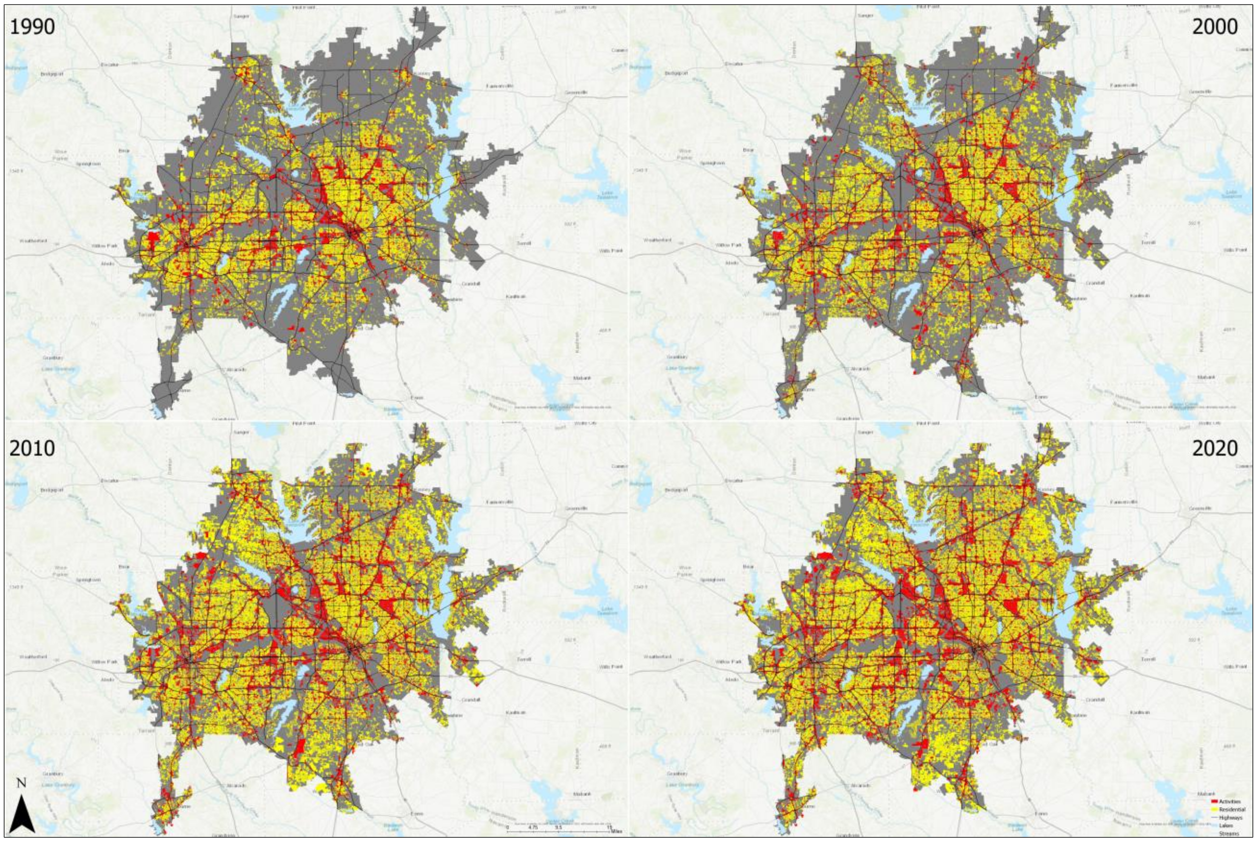


Figure 4-1: Area change for Activity and Residential land uses in the DFW Urban Area (1990 - 2020)



Map 4-1: Activity and Residential land uses change in the DFW Urban Area (1990 - 2020)

4.2.1 Overall Land Use Change Pattern in the Study Area

To calculate overall Activity and Residential land uses change, we used the Cut and Fill technique by calculating the area change between two time periods (e.g., 1990 vs. 2020). As illustrated in Figure 4-2, Cut and Fill takes surfaces of a given location at two different periods and identifies blocks that gained, lost, or had no change of Activity or Residential land uses.

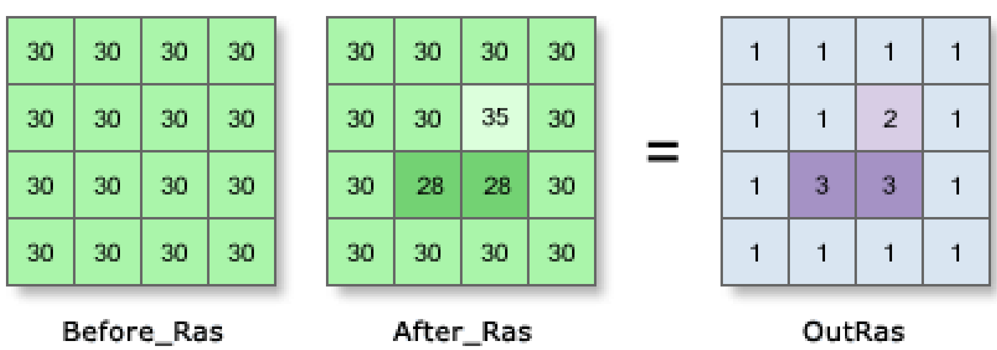
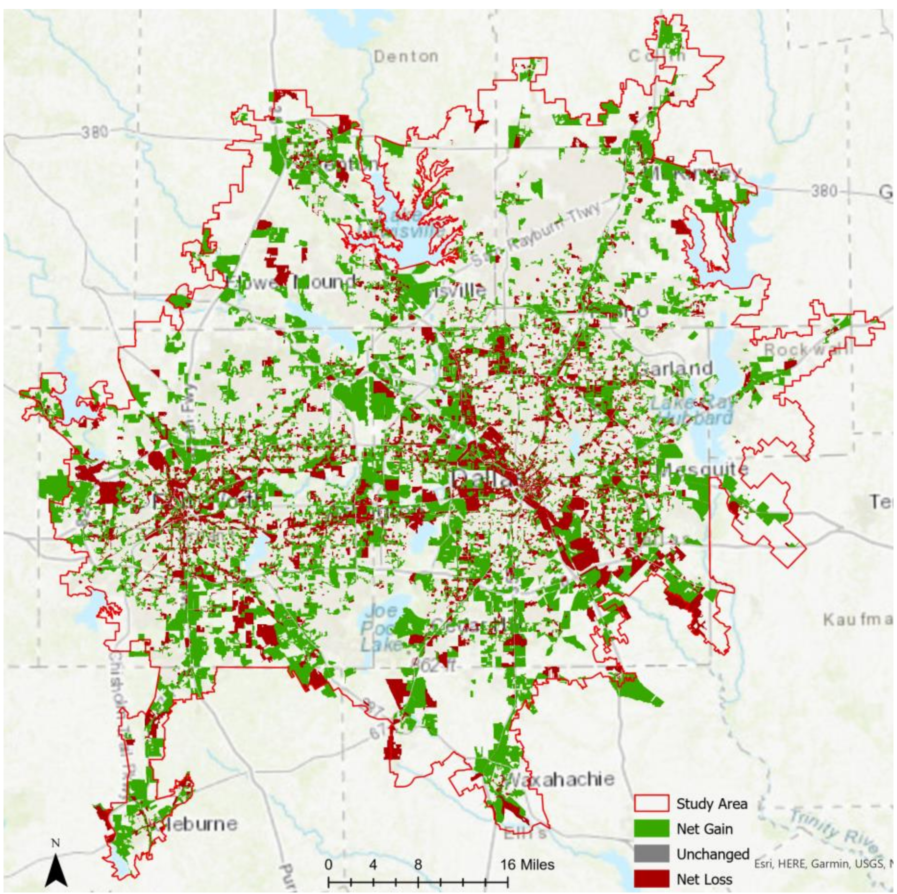


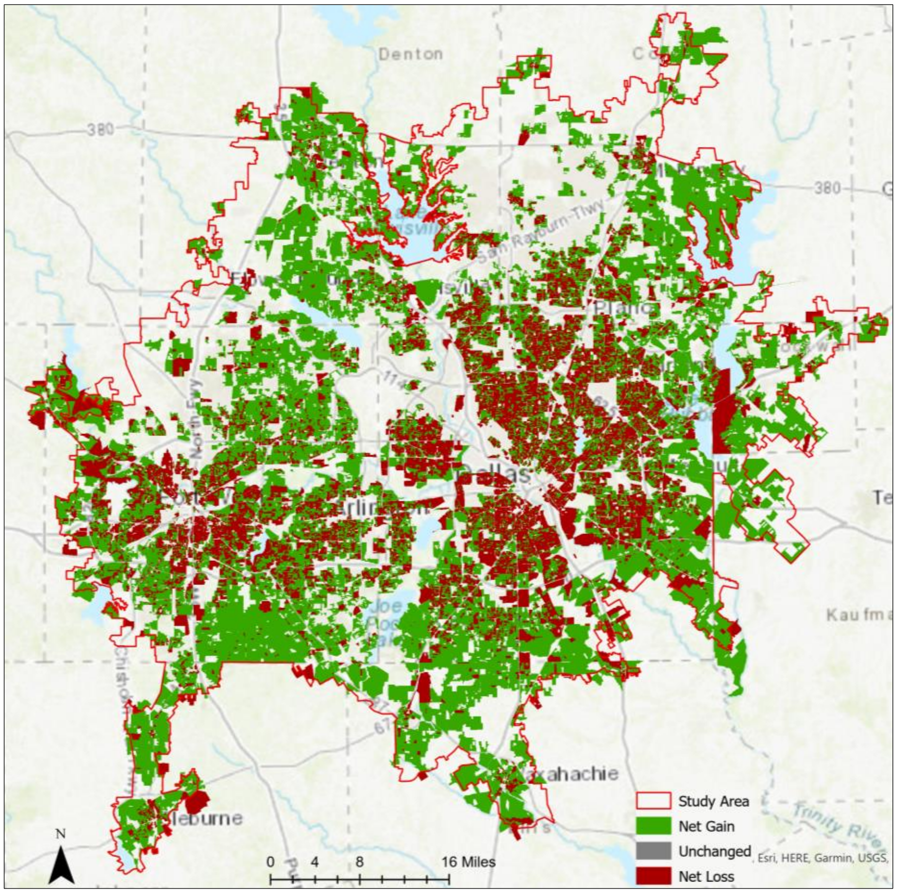
Figure 4-2: Cut and Fill illustration (credit: Esri)

The result of Cut and Fill operation depicted that the DFW Metropolitan Area's traditional centers and CBDs (i.e., downtown Dallas and Fort Worth and the corridor between them) are losing Activity and Residential land uses in favor of peripheral areas in the Metropolitan Area's north, east, and south. This trend is corroborated by previous studies that observed that “CBD densities peaked as a city grew, then they began to fall as the city expanded” (Kaplan et al., 2014, p. 125).

Interestingly, the west of the DFW Metropolitan Area did not gain much development; it could be due to lower economic opportunities, weaker infrastructure, less diverse demographics, or development policies, which are beyond the scope of this research. Map 4-2 and Map 4-3 show the gains and losses between 1990 and 2020 for Activity and Residential land uses, respectively.



Map 4-2: Activity land uses’ Gain/Loss between 1990 and 2020



Map 4-3: Residential land uses' Gain/Loss between 1990 and 2020

4.2.2 Land Use Change Trend by Distance from Downtowns

Since one of the overall assumptions and hypothesis of this research is decentralization from CBDs and the formation of new Activity and Residential land uses centers beyond the Dallas-Fort Worth corridor, we measured the spatial distribution of Activity and Residential land uses within the DFW Urban Area by distance from downtowns. We analyzed land use change in 1-mile buffers (Map 4-4) from downtown Dallas and Fort Worth. Overall, three spatial distribution patterns are observed; the trend results are presented in Figure 4-3, Figure 4-4, and Figure 4-5.

- Up to 2 miles from downtowns, as expected, the share of Activity land uses is higher than residential. The percentage of Activity land uses constantly decreased from 40% in 1990 to 25% in 2020. Residential land uses, however, were steady until 2015 (at about 10%); it increased slightly upward (1.5% to 11.5%) from 2015 to 2020. This upward trend in Residential land uses in CBDs could indicate millennials' willingness to live in downtown areas that appreciate their values of walkability, convenience, and environmental sustainability, which are often associated with downtown living lifestyles (Florida, 2019; Lee et al., 2019).

- In the distance between 2 to 8 miles from downtowns, the Residential share of land use is higher than Activity land uses; there is not much change in Activity and Residential land uses between 1990 and 2020. In this zone, while Residential land uses declined slightly from 36% to 34%, Activity land uses increased from 16% to 17.5%, which is not remarkable in either case.
- Beyond 8 miles from downtown and closer to the fringes of the Metropolitan Area and suburbs, the share of Residential land uses is much higher than Activity land uses during each period; however, the gap between the two increases faster starting in 2005. The increase rate for Activity and Residential land uses is more noticeable, indicating the dispersion of Activity land uses from downtown toward suburban areas. Studying the relationship between this land use pattern and demographic characteristics (age, race, gender, education, etc.) could reveal interesting results, which is beyond this research's scope.

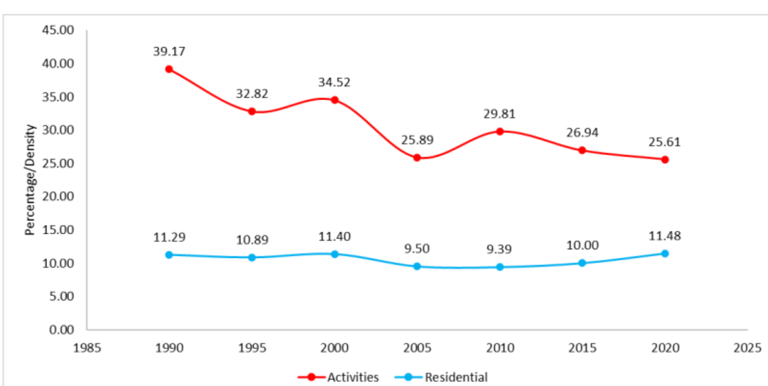


Figure 4-3: Land use change trend within <2 miles of downtown Dallas and Fort Worth

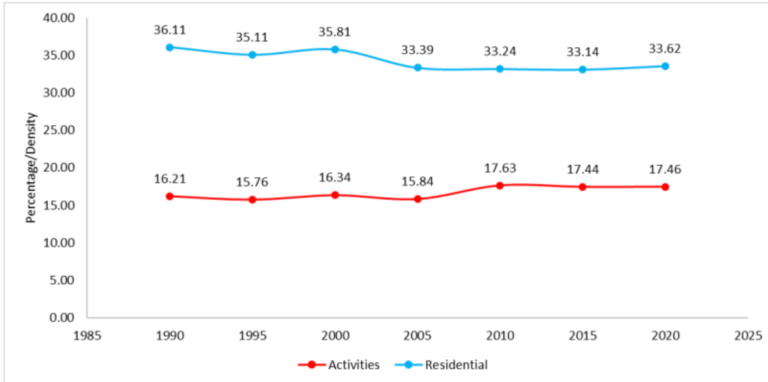


Figure 4-4: Land use change trend within 2 - 8 miles of downtown Dallas and Fort Worth

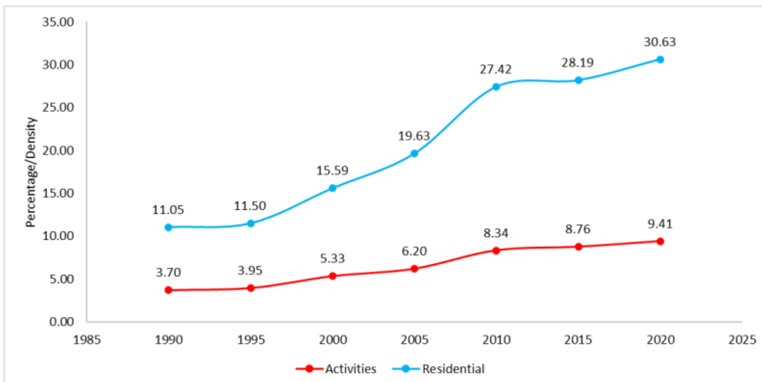
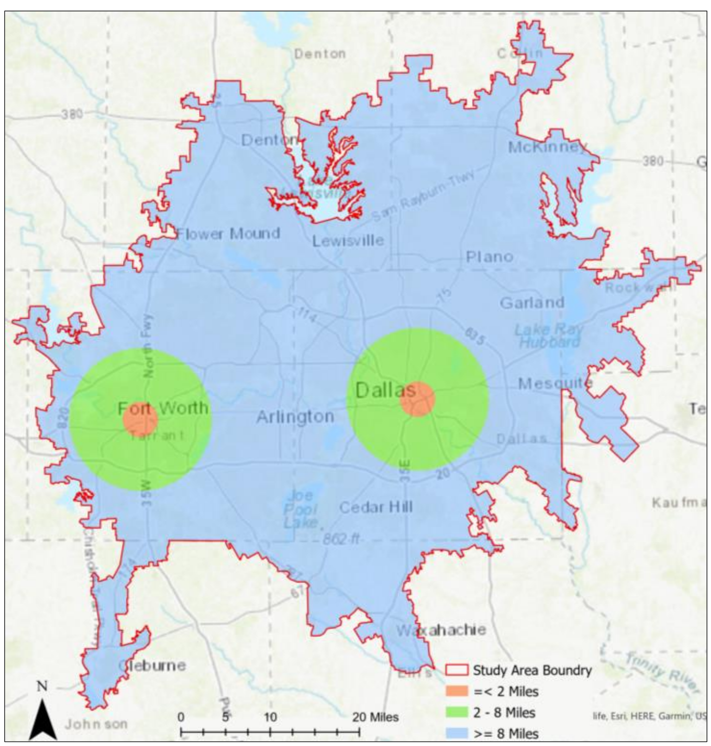


Figure 4-5: Land use change trend beyond > 8 miles of downtown Dallas and Fort Worth



Map 4-4: Distances from downtown Dallas and Fort Worth (<2 miles, 2 – 8 miles, >8 miles)

4.2.3 Land Use Change Trend by Cities

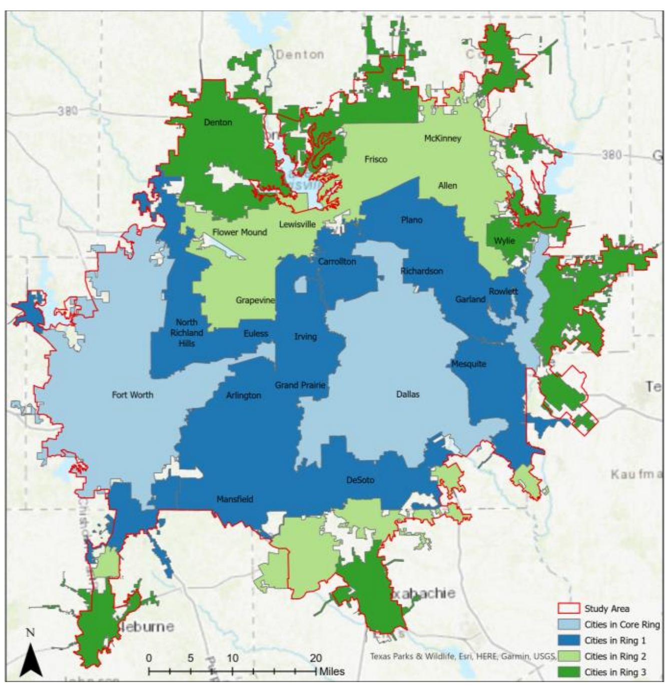
Analysis of land use change by city boundaries shows that while cities in the Dallas-Fort Worth corridor are growing slower, smaller cities in the fringes and suburbs of the Metropolitan Area are growing much faster. As mentioned in the previous section, since one of the overall assumptions and hypothesis of this research is decentralization and formation of new Activity and Residential centers beyond Dallas-Fort Worth corridor, and due to the large number of cities and places within the study area (about 130 cities), cities are categorized into four major categories based on their distance from Dallas and Fort Worth, called rings (Map 4-5); then, the change in Activity and Residential land uses is calculated cumulatively within each ring. Major cities (with a population of more than 50,000 people in 2020) associated with each ring are:

- Core Ring: includes Dallas and Fort Worth (along small towns within each city boundary (e.g., University Park and Cockrell Hill in Dallas, Lake Worth, Saginaw, and Forest Hill in Fort Worth).
- Ring 1: Includes cities that share a border with core ring cities, which are Arlington, Richardson, Carrollton, DeSoto, Euless, Mansfield, Mesquite, North Richland Hills, Plano, Rowlett, Garland, Grand Prairie, and Irving.
- Ring 2: Includes cities that share borders with the outskirts of cities in Ring 1, which are Lewisville, Allen, Flower Mound, Frisco, McKinney, and Grapevine.
- Ring 3: Include cities on the fringe of the DFW Metropolitan Area and share borders with the outskirts of cities in Ring 2, which are Denton and Wylie.

Figure 4-6, Figure 4-7, Figure 4-8, and Figure 4-9 show the trend of Activity and Residential land uses change in each ring from 1990 to 2020. They show that the overall share of Activity and Residential land uses for cities within the Core ring slightly increased by less than 4 percent.

For cities in Ring 1, the increase rate for Activity and Residential land uses accelerates; Residential land uses increased by more than 11 percent, and Activity land uses increased by more than 6 percent. For cities within Ring 2, while Activity land uses increased at the same 6 percent rate as in Ring 1, Residential land uses increased by 23 percent. Finally, for cities within Ring 3, Residential land uses increased by 28 percent; however, Activity land uses increased by more than 4 percent.

Overall, not much change is noticeable in the cities within the core ring for both Activity and Residential land uses. However, Activity land uses for cities in Rings 1 and 2 are growing more than two times faster than cities within the Core ring (6.3% vs. 2.8%). For Ring 3, although the growth of Activity land uses is slowing down, it is still higher than cities within the Core ring (4.3% vs. 2.8%), indicating spreading Activity land uses toward the fringes of the DFW Metropolitan Area. Residential land uses change, however, is constantly increasing toward the outskirt's rings: 4%, 11%, 23%, and 28% for Core ring, Ring 1, Ring 2, and Ring 3, respectively.



Map 4-5: City-based rings in the study area

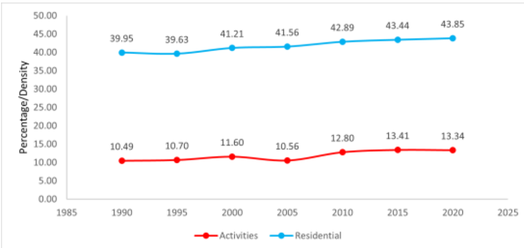


Figure 4-6: Land use change trend in Core Ring cities

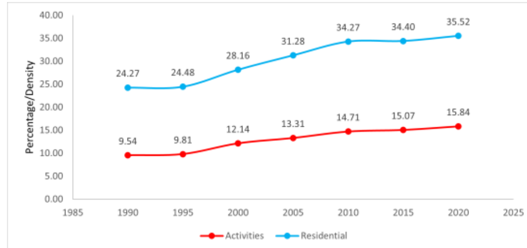


Figure 4-7: Land use change trend in Ring 1 cities

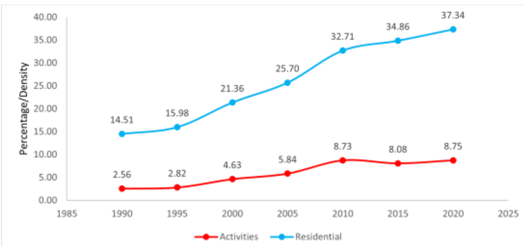


Figure 4-8: Land use change trend in Ring 2 cities

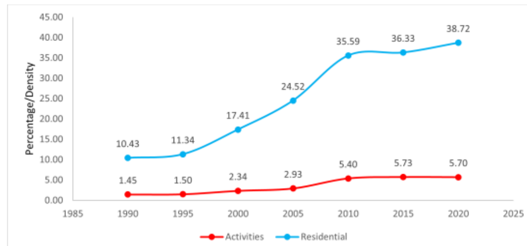
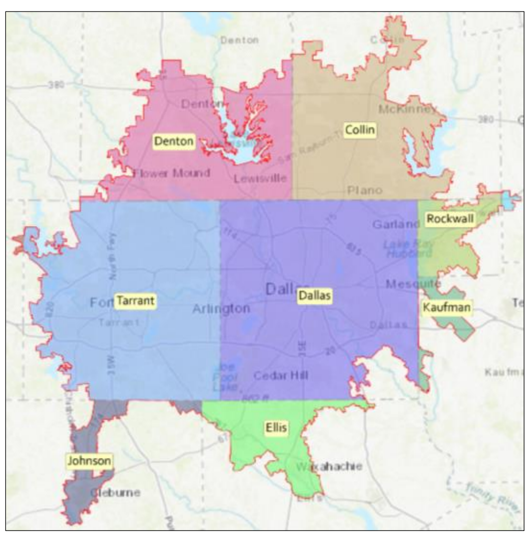


Figure 4-9: Land use change trend in Ring 3 cities

4.2.4 Land Use Change Trend by Counties

Land use change analysis by counties located within the study area shows that while Dallas and Tarrant counties (the two major and central counties in the region) are growing at a slower pace, counties located in the fringes of the study area (including Denton, Collin, Rockwall, Ellis, Kaufman, and Johnson) are growing at a much higher pace. This is another indication of sprawl and scatteredness in the DFW Metropolitan Area. Map 4-6 shows counties within the study area; Figure 4-10 through Figure 4-17 show the land use change trend based on counties.



Map 4-6: Counties in the study area

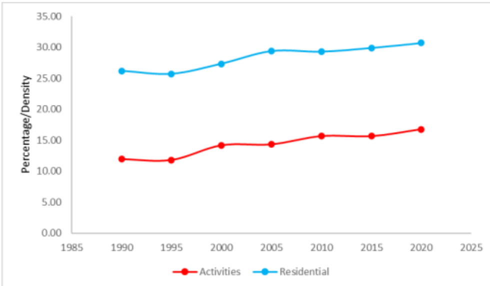


Figure 4-10: Land use change trend in Dallas

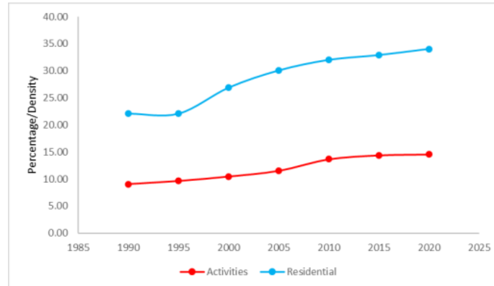


Figure 4-11: Land use change trend in Tarrant

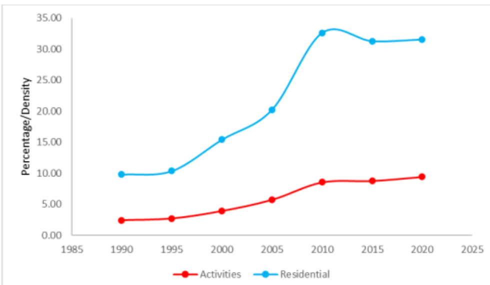


Figure 4-12: Land use change trend in Denton

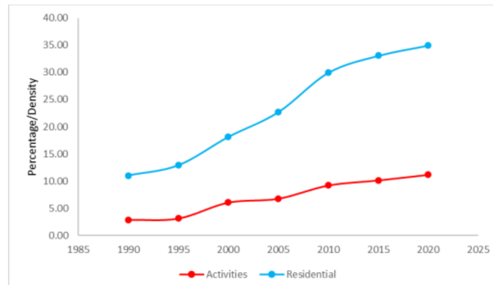


Figure 4-13: Land use change trend in Collin

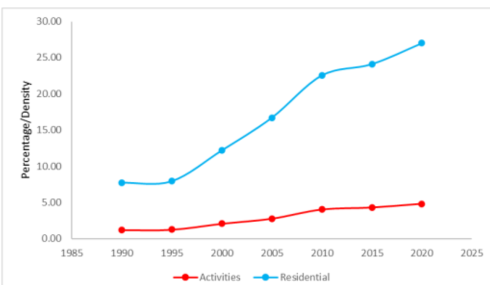


Figure 4-14: Land use change trend in Rockwall

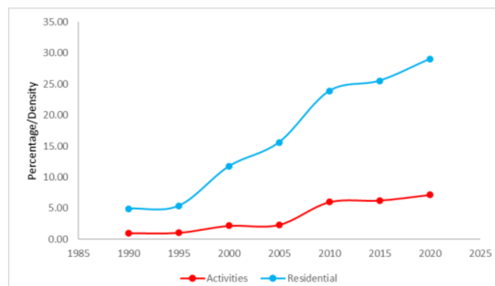


Figure 4-15: Land use change trend in Kaufman

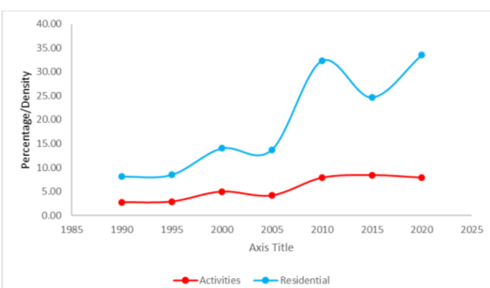


Figure 4-16: Land use change trend in Ellis

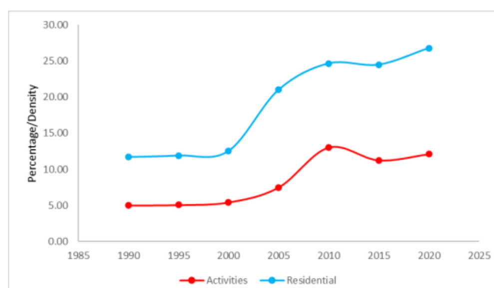


Figure 4-17: Land use change trend in Johnson

4.3 Spatiotemporal Distribution Pattern of Land Use at Micro Scale

This section will discuss the spatiotemporal pattern of Activity and Residential land uses in the DFW Urban Area from 1990 to 2020 at the micro level (i.e., census blocks). The goal is to examine whether the Activity and Residential land uses distribution pattern is clustered, scattered, or random. To do so, we will use three spatial statistical methods called Global Moran's I (for spatial autocorrelation), (Getis-Ord Gi* (for hot spot analysis), and (Anselin Local Moran's I (for cluster/outlier analysis). The unit of analysis is a census block, and the attribute that is being analyzed is the ratio of Activity and Residential land uses in a block, calculated as:

$$Activity(Residential)\ Ratio = \frac{Area\ of\ Activity\ (Residential)\ Land\ uses\ in\ Block}{Area\ of\ Block} * 100$$

Before presenting the results of spatial statistics analysis methods, it is necessary to explain the foundations of spatial statistics, as it helps us to understand its key characteristics and measures and interpret the outputs.

4.3.1 Spatial Statistics

Spatial statistics are developed for use with geographic data. While there may be similarities between spatial and conventional (nonspatial) statistics regarding concepts and objectives, spatial statistics differ from conventional statistics in that they are designed for geographic data. They take into account spatial concepts such as distance, area, and neighborhood in their calculations, unlike non-spatial methods that ignore them. With spatial statistics, to quantify the pattern of features and their associated values, we can calculate a statistic instead of just mapping them; it is easier to compare distribution patterns in different periods, as it accounts for the spatial relationships among the features and associated values.

Most "statistical tests begin by identifying a Null hypothesis. The null hypothesis for the pattern analysis is Complete Spatial Randomness (CSR) of the features themselves or the values associated with those features" (Esri, 2023b). Pattern analysis has two outputs of z-scores and p-values, by which we decide if we can reject the null hypothesis, as we hope. Therefore, If the z-score is high (numerically, regardless of the positive or negative sign) and the p-value is low, we can reject the null hypothesis; it means that the features or associated values with features (i.e., Activity or Residential Ratio) are not randomly distributed but show significant clustering (if z-score is positive) or dispersion (if z-score is negative).

The z-score and p-value are based on the standard normal distribution, as shown in Figure 4-18. They indicate how likely the observed spatial pattern is random or not. A very high or very low z-score (at the tales of the normal distribution graph) and a very small p-value mean that the pattern differs from the null hypothesis of CSR. Therefore, z-score and p-value are the basis of pattern analysis for spatial autocorrelation, hot spot analysis, and cluster/outlier analysis.

To test the null hypothesis of CSR, we need to decide how much risk we are willing to take in making a wrong decision, that is, rejecting the null hypothesis when it is true. This risk is determined by the confidence level, usually set at 90, 95, or 99 percent. The higher the confidence level, the more stringent the test and the less likely it is to reject the null hypothesis by mistake (Ebdon, 1991; Goodchild, 1986; Mitchell, 2005). Therefore, choosing the confidence level before performing the spatial statistic, based on our judgment and the context of analysis, is critical and significantly impacts the results.

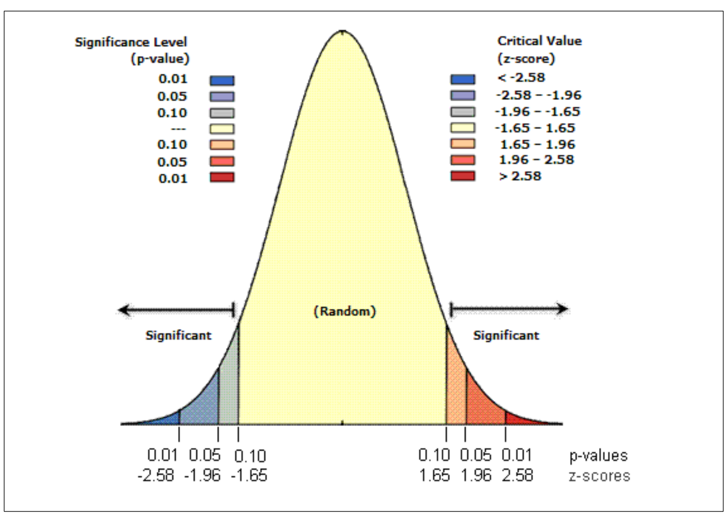


Figure 4-18: Standard normal distribution (credit: Esri)

4.3.2 Spatial Autocorrelation for Activity and Residential Land Uses

Spatial autocorrelation is a concept that measures the degree of similarity or dissimilarity between spatial units, such as regions, countries, or pixels (Goodchild, 1986; Griffith, 1987). It is based on the idea that spatial units close to each other tend to have similar values or attributes, while those far apart tend to have different values or attributes. This phenomenon is also known as Tobler's first law of geography, as described previously in this chapter.

The spatial autocorrelation index can be positive, negative, or zero. A positive spatial autocorrelation Index means that similar values cluster together in space, forming patterns of high-high or low-low values. A negative spatial autocorrelation index means dissimilar values tend to be adjacent in space, forming patterns of high-low values. Zero spatial autocorrelation index means no spatial pattern or relationship between the associated values of spatial units, indicating randomness in the distribution pattern of a phenomenon or associated value.

Spatial autocorrelation can be measured by various methods and indices, such as Moran's I, Geary's C, and Local Indicators of Spatial Association (LISA). In this research, we used the Global Moran' I index due to its popularity and ease of calculation in similar situations (Tepe, 2023). Calculated as in equations 1 through 5 below, Global Moran's I calculates an observed index value based on the data values for each feature and estimates the expected index value. It compares the observed and expected values of the indices to measure how different they are and tests whether the observed pattern is clustered, dispersed, or random using z-score and p-value statistics.

Since we are interested in the distribution pattern of Activity and Residential land uses, the goal of the research, as is the case for any distribution pattern analysis, is to reject the null hypothesis of Complete Spatial Randomness (CSR). According to the economies of scale and spatial dependency theories discussed earlier, Activity and Residential land uses are expected to be clustered (APA-PAS Report No. 135, 1960). However, we are interested in whether the degree of the clustering or dispersion for these two land uses in the DFW Metropolitan Area changes over time and, if yes, by what ratio.

$$I = \frac{n}{S_0} \frac{\sum_{i=1}^{n} \sum_{j=1}^{n} w_{i,j} z_i z_j}{\sum_{i=1}^{n} z_i^2} \tag{1}$$

where z_i is the deviation of an attribute for feature i from its mean $(x_i - \bar{X})$, $w_{i,j}$ is the spatial weight between feature i and j, n is equal to the total number of features, and S_0 is the aggregate of all the spatial weights:

$$S_0 = \sum_{i=1}^{n} \sum_{j=1}^{n} w_{i,j} \tag{2}$$

The z_I-score for the statistic is computed as:

$$z_I = \frac{I - \mathrm{E}[I]}{\sqrt{\mathrm{V}[I]}} \tag{3}$$

where:

$$\mathrm{E}[I] = -1/(n-1) \tag{4}$$

$$\mathrm{V}[I] = \mathrm{E}[I^2] - \mathrm{E}[I]^2 \tag{5}$$

Even though the unit of analysis is block, we calculated Moran's I for Activity and Residential land uses at block, block group, and census tracts to measure the impact of the scale of analysis. The results (Table 4-1) show that as the unit of analysis gets smaller (i.e., blocks), the z-score increases, resulting in a coarser dispersion trend, and vice versa; as the unit of analysis gets larger (i.e., tracts), the z-score decreases, resulting in a smoother dispersion trend. Figure 4-19, Figure 4-20, and Figure 4-21 show the z-score values at block, block group, and census tracts, respectively, which indicates that the distribution trend of Activity and Residential land uses in the DFW Urban Area is toward scatteredness regardless of the size of the unit of analysis. This is called scale neutrality, in which the scale or size of the unit of analysis does not change the overall course of the trend (Vagias, 2006).

Table 4-1: Global Moran I result for Activity and Residential land uses (1990 - 2020)

Year	Census Blocks 2020				Census Block Groups 2020				Census Tracts 2020			
	P-Value		Z-Score		P-Value		Z-Score		P-Value		Z-Score	
	Residential	Activities	Residential	Activities	Residential	Activities	Residential	Activities	Residential	Activities	Residential	Activities
1990	0.00	0.00	35.86	73.54	0.00	0.00	21.28	32.31	0.00	0.00	19.37	27.92
1995	0.00	0.00	36.04	70.30	0.00	0.00	21.39	31.85	0.00	0.00	18.44	26.70
2000	0.00	0.00	27.73	55.46	0.00	0.00	16.34	29.56	0.00	0.00	14.59	24.36
2005	0.00	0.00	19.94	50.27	0.00	0.00	8.27	28.43	0.00	0.00	8.94	23.29
2010	0.00	0.00	5.17	55.63	0.01	0.00	2.57	28.67	0.00	0.00	5.92	21.86
2015	0.02	0.00	-6.82	54.49	0.46	0.00	0.74	26.96	0.02	0.00	4.83	21.16
2020	0.00	0.00	-14.22	48.80	0.14	0.00	-1.47	28.14	0.00	0.00	3.73	21.83

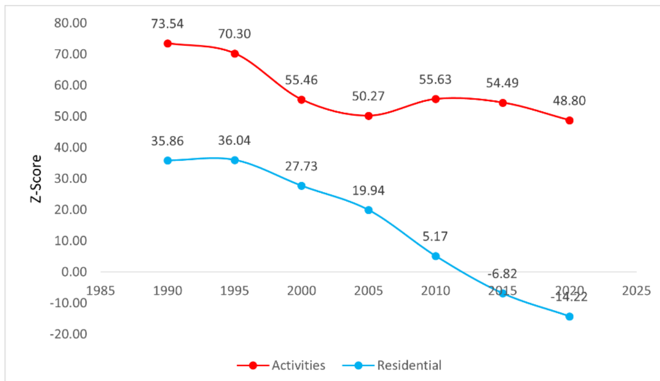


Figure 4-19: Distribution of Activity and Residential land uses at census block level

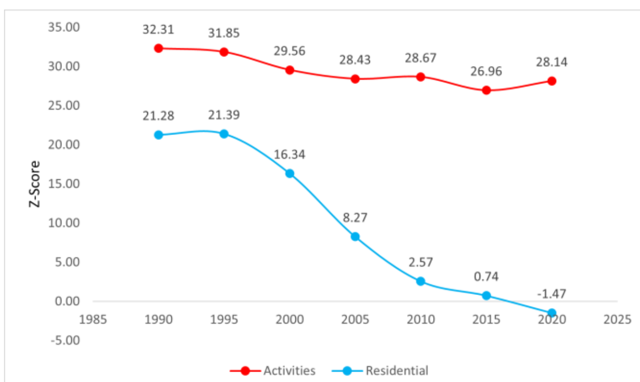


Figure 4-20: Distribution of Activity and Residential land uses at census block group level

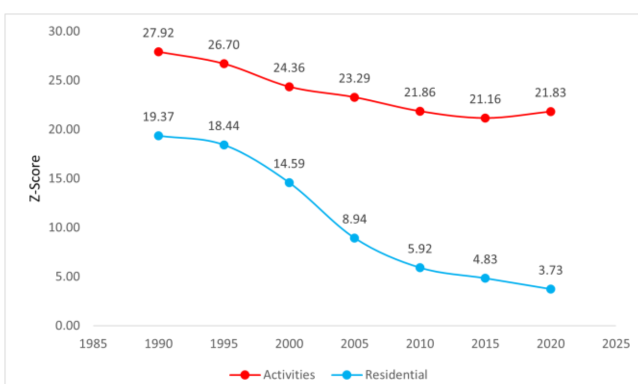


Figure 4-21: Distribution of Activity and Residential land uses at census tract level

Analysis results of Spatial Autocorrelation for land use change pattern show a very small p-value (close to zero) and large positive z-score for both Activity and Residential land uses. It is interpreted as blocks with a high ratio of Activity or Residential land uses are clustered together, and blocks with a low ratio of Activity or Residential land uses are clustered together. As a result, the null hypothesis of Complete Spatial Randomness (CSR) is rejected for both Activity and Residential land uses. However, the overall distribution trend of Activity and Residential land uses in the DFW Urban Area is toward scatteredness, as the z-score is getting smaller over time between 1990 and 2020.

4.3.3 Hot Spot Analysis for Activity and Residential Land Uses

In Hot Spot analysis, we compare the value associated with a feature (i.e., the ratio of Activity or Residential land use in a block) and its neighbors to the study area (refer to Map 4-7-A for the concept of neighborhood in hot spot analysis). If the value is significantly higher than the study area, that block is a hot spot; if the value is significantly lower than the study area, that block is a cold spot.

Both hot spots and cold spots are evaluated through confidence levels (typically 90, 95, or 99 percent). Hot spots and cold spots are not necessarily the same as blocks with high or low values of each land use type. Instead, it shows which high/low-value blocks are surrounded by other high/low-value blocks. The method used for Hot Spot analysis is called Getis-Ord Gi*, and it is calculated according to equations 1 through 3 below (Getis et al., 1992; Ord et al., 1995). This equation returns a z-score for each block, which indicates how significant the spatial clustering of high or low values is. Positive z-scores with high values show hot spots of features with high values surrounded by other high-value features. Negative z-scores with low values show cold spots of features with low values surrounded by other low-value features.

The Getis-Ord local statistic is given as:

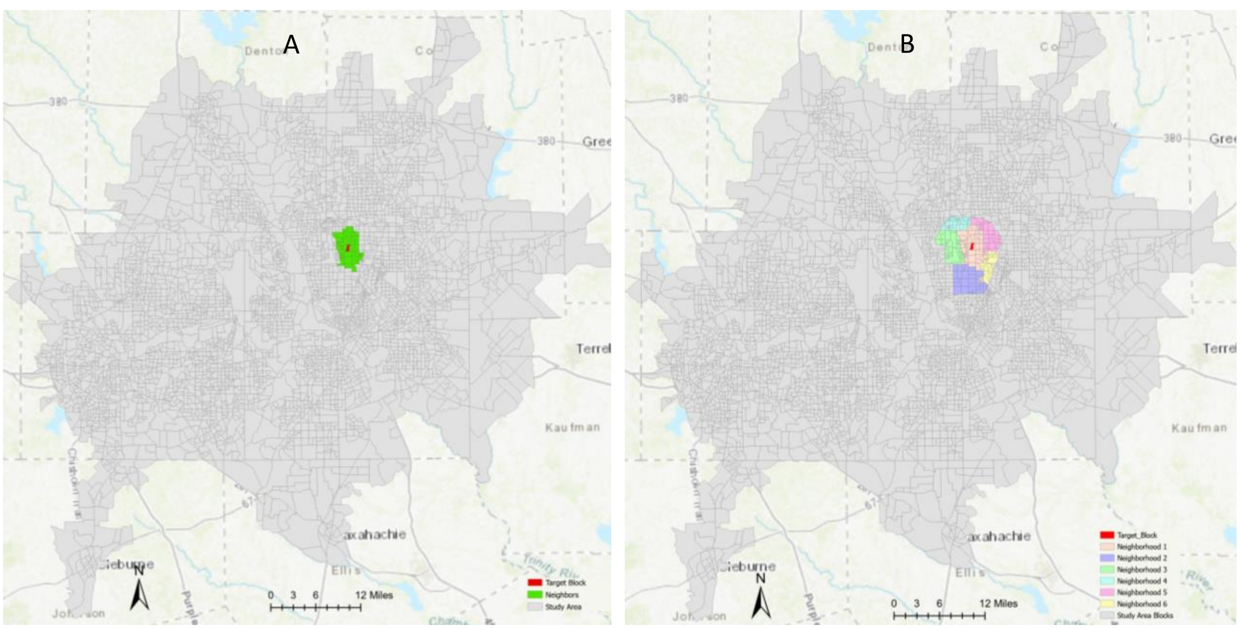
$$G_i^* = \frac{\sum_{j=1}^{n} w_{i,j} x_j - \bar{X} \sum_{j=1}^{n} w_{i,j}}{S \sqrt{\frac{\left[n \sum_{j=1}^{n} w_{i,j}^2 - \left(\sum_{j=1}^{n} w_{i,j} \right)^2 \right]}{n-1}}} \tag{1}$$

where x_j is the attribute value for feature j, $w_{i,j}$ is the spatial weight between feature i and j, n is equal to the total number of features and:

$$\bar{X} = \frac{\sum_{j=1}^{n} x_j}{n} \tag{2}$$

$$S = \sqrt{\frac{\sum_{j=1}^{n} x_j^2}{n} - \left(\bar{X}\right)^2} \tag{3}$$

The G_i^* statistic is a z-score so no further calculations are required.



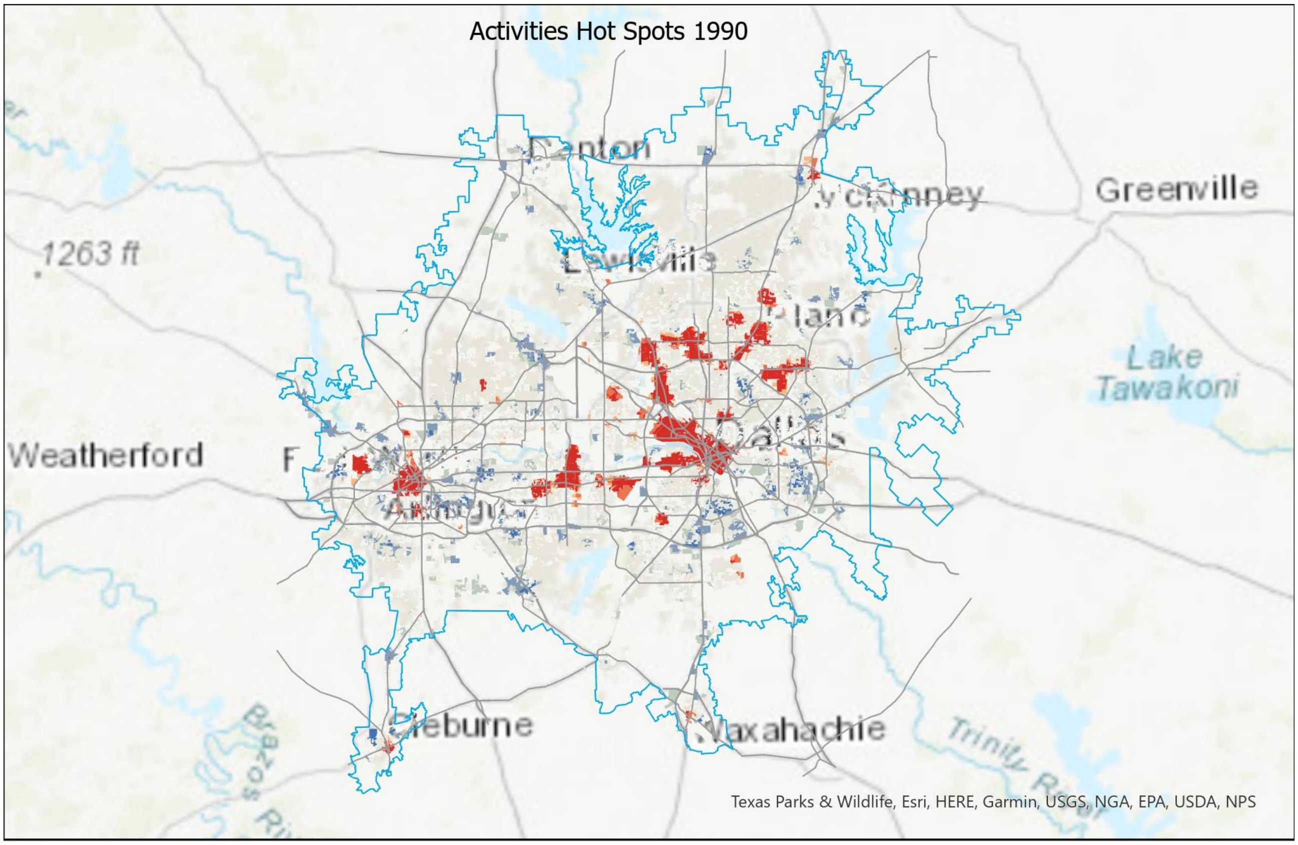
Map 4-7: Concept of neighborhood in spatial statistics

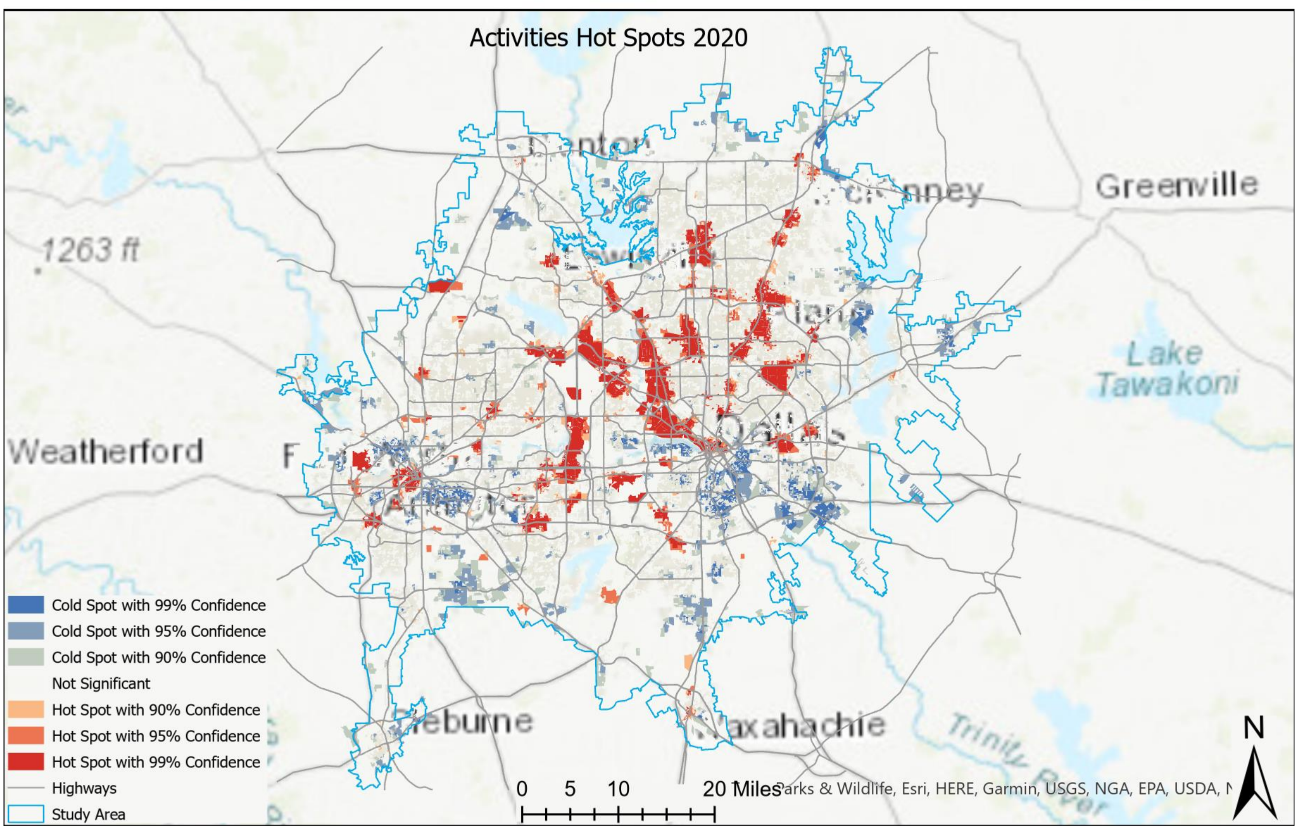
The results of hot spot analysis in the DFW Urban Area not only reveal the growth in the Activity and Residential land uses (as expected) between 1990 and 2020, but it also proved the formation of new centers and corridors of Activity land uses mainly in the north and north-east part of the Metropolitan Area (Map 4-8).

Pairing with the results from spatial autocorrelation, it reveals the decentralization of Activity land uses from central areas of the DFW Metropolitan Area. While the center of the Metropolitan Area (downtown Dallas, Fort Worth, and the corridor between the two) is losing Activity land uses, the peripheral areas in the north and northeast absorb these land uses. Reviewing the development policies for individual cities may reveal the underlying conditions that lead to such development patterns.

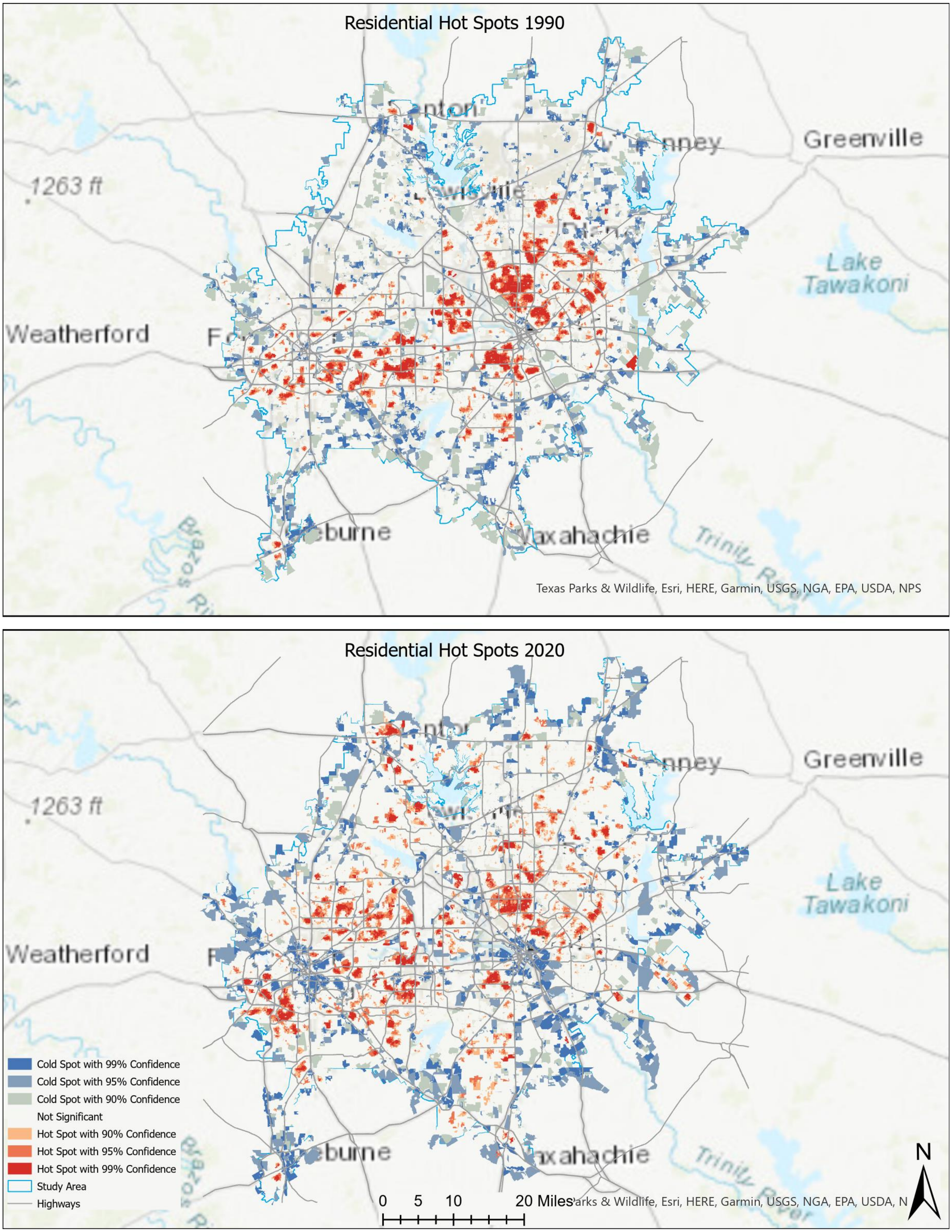
For Residential land uses, however, as shown in Map 4-9, we observe the expansion of Residential land uses all over the Metropolitan Area, including downtown Dallas and Fort Worth, where they are losing Activity land uses. It depicts the formation of new hot spots in the DFW Metropolitan Area. In contrast, new cold spots in the central parts of the Metropolitan Area indicate people's tendency to live (come back) in downtowns and CBDs.

One explanation for such a trend could be the intention of having mixed-use land uses in the central locations (e.g., uptown Dallas, where we have seen the expansion of high rises of mixed uses). Also, the urban revival that has taken place in the last 20 years owes much to the role of young people, who have shown a greater preference for living in central urban areas than previous generations did when they were at similar life stages (Lee, 2019). This generational shift in residential choices has implications for the future of cities and the economic, social, and environmental challenges and opportunities that urbanization presents.





Map 4-8: Hot spots change for Activity land uses (1990 vs. 2020)



Map 4-9: Hot spots change for Residential land uses (1990 vs. 2020)

4.3.4 Cluster and Outlier Analysis for Activity and Residential Land Uses

In Cluster/Outlier analysis, we analyze the associated value (i.e., the ratio of Activity and Residential land uses in a block) of the feature to its neighbors and the significance of neighbors to their neighbors (refer to Map 4-7-B for the concept of neighborhood in cluster/outlier analysis). The Cluster and Outlier Analysis method is Anselin Local Moran's I, calculated according to equations 1 through 5 below (Anselin, 1995).

The Local Moran's I statistic of spatial association is given as:

$$I_i = \frac{x_i - \bar{X}}{S_i^2} \sum_{j=1, j\neq i}^{n} w_{i,j}(x_j - \bar{X}) \qquad (1)$$

where x_i is an attribute for feature i, $\bar{X}$ is the mean of the corresponding attribute, $w_{i,j}$ is the spatial weight between feature i and j, and:

$$S_i^2 = \frac{\sum_{j=1, j\neq i}^{n} (x_j - \bar{X})^2}{n-1} \qquad (2)$$

with n equating to the total number of features.

The z_{I_i}-score for the statistics are computed as:

$$z_{I_i} = \frac{I_i - \mathrm{E}[I_i]}{\sqrt{\mathrm{V}[I_i]}} \qquad (3)$$

where:

$$\mathrm{E}[I_i] = -\frac{\sum_{j=1, j\neq i}^{n} w_{ij}}{n-1} \qquad (4)$$

$$\mathrm{V}[I_i] = \mathrm{E}[I_i^2] - \mathrm{E}[I_i]^2 \qquad (5)$$

Anselin Local Moran's I is based on the global Moran's I statistic, which calculates the overall spatial autocorrelation for the entire dataset. The difference is that the Local Moran's I computes a local index for each feature, which indicates how similar or dissimilar its value is to the values of its neighbors. The neighbors are defined by a spatial weights matrix (SWM), which specifies how much each feature influences or is influenced by other features. Cluster/Outlier Analysis results are one of the following options and illustrated in Figure 4-22.

- **High-value cluster:** If the value of the feature is higher than other features, and the value of the feature's neighbor is higher than the value of other neighbors.
- **Low-value cluster:** If the value of the feature is lower than other features, and the value of the feature's neighbor is lower than the value of other neighbors.
- **Outlier:** If the value of the feature is higher than other features, but the value of the feature's neighbor is lower than the value of other neighbors. Or if the value of the feature is lower than the value of other features, but the value of the feature's neighbor is higher than the value of other neighbors.

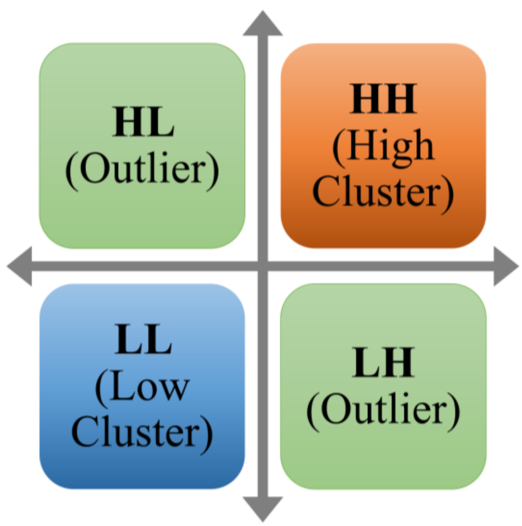
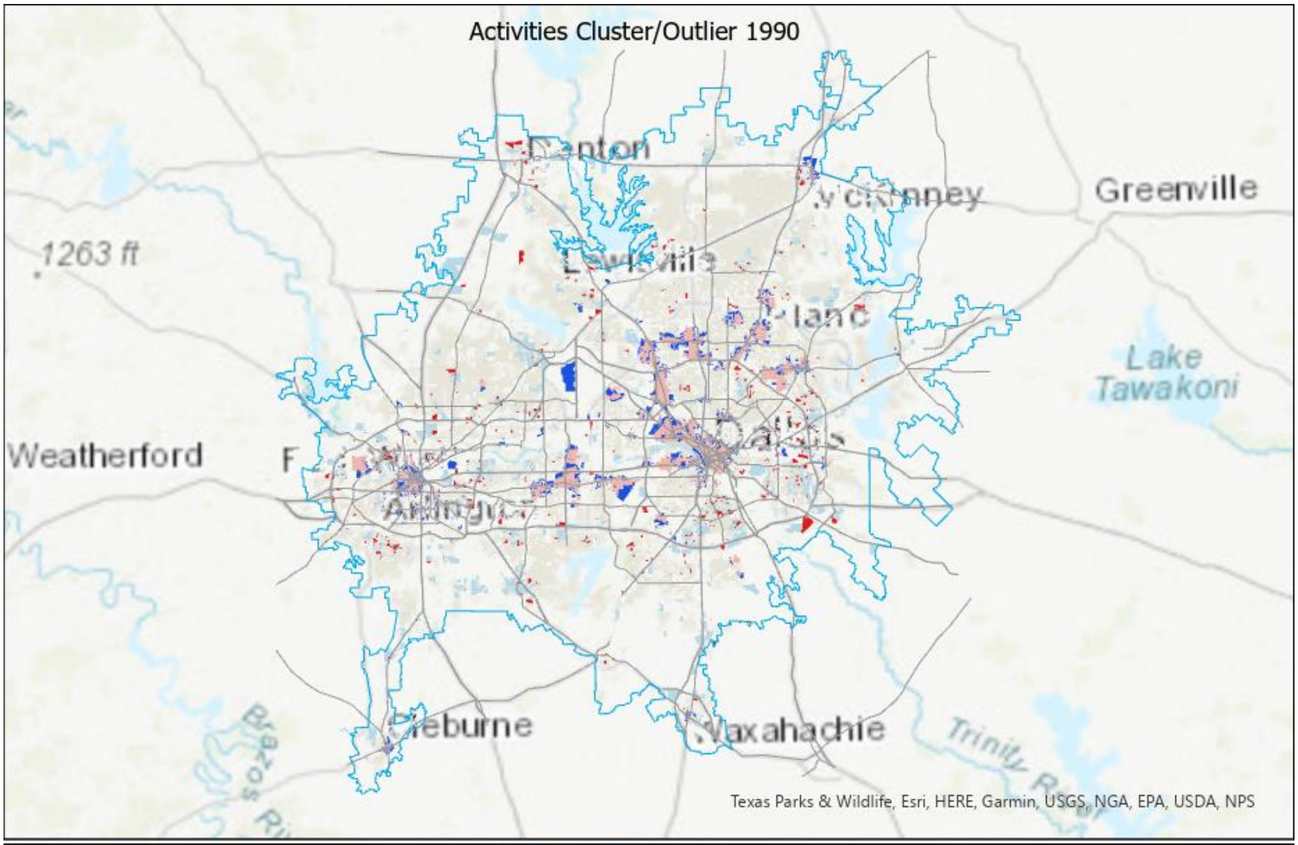


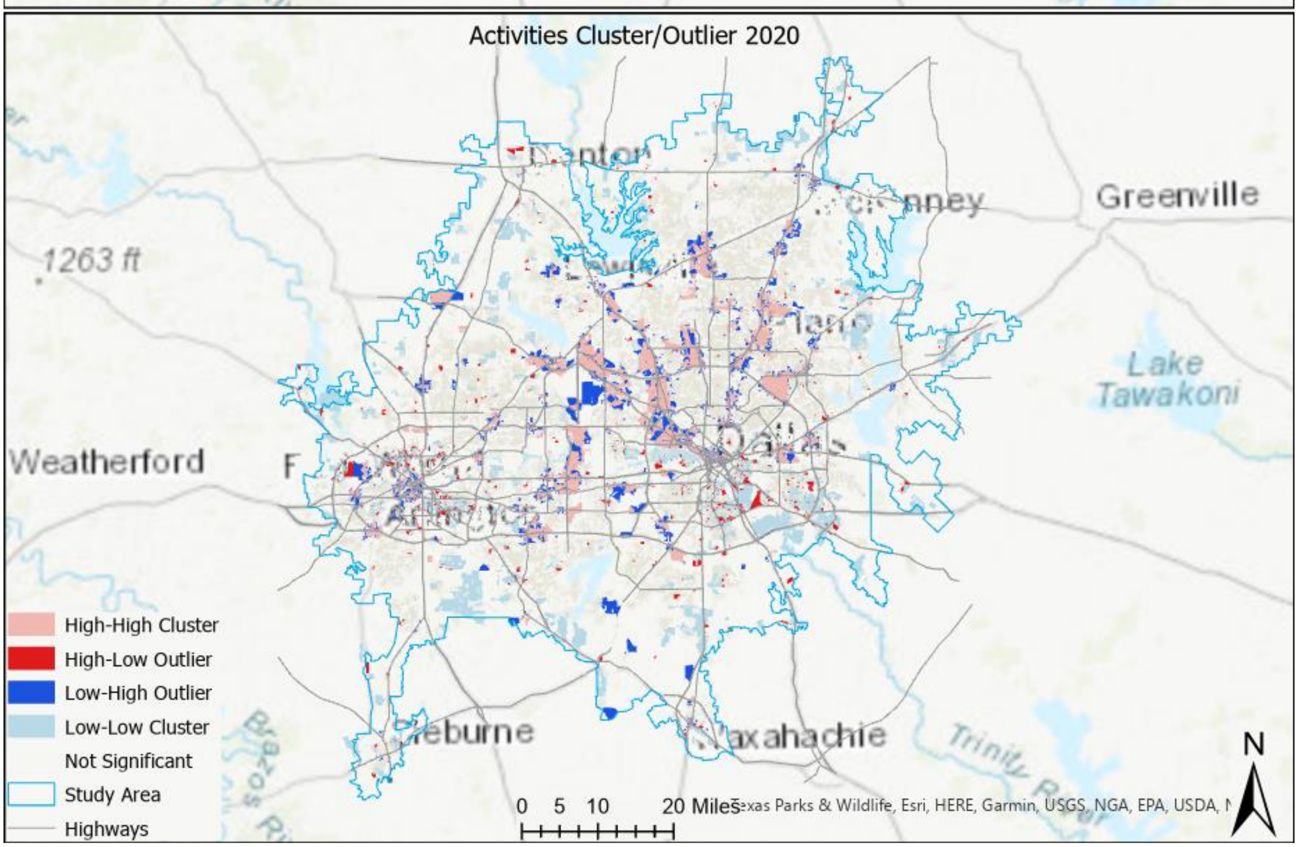
Figure 4-22: Schematic results of Cluster/Outlier Analysis

The Cluster/Outlier analysis results of Activity and Residential land uses change in the DFW Urban Area between 1990 and 2020 are shown in Map 4-10 and Map 4-11, respectively. A feature with a positive I value is surrounded by features with similar attribute values, either high (light red) or low (light blue). This means the feature belongs to a cluster of features with a common pattern. A feature with a negative I value is surrounded by features with different attribute values, either higher (dark red) or lower (dark blue). This means that the feature is an outlier that deviates from the general trend of its neighbors.

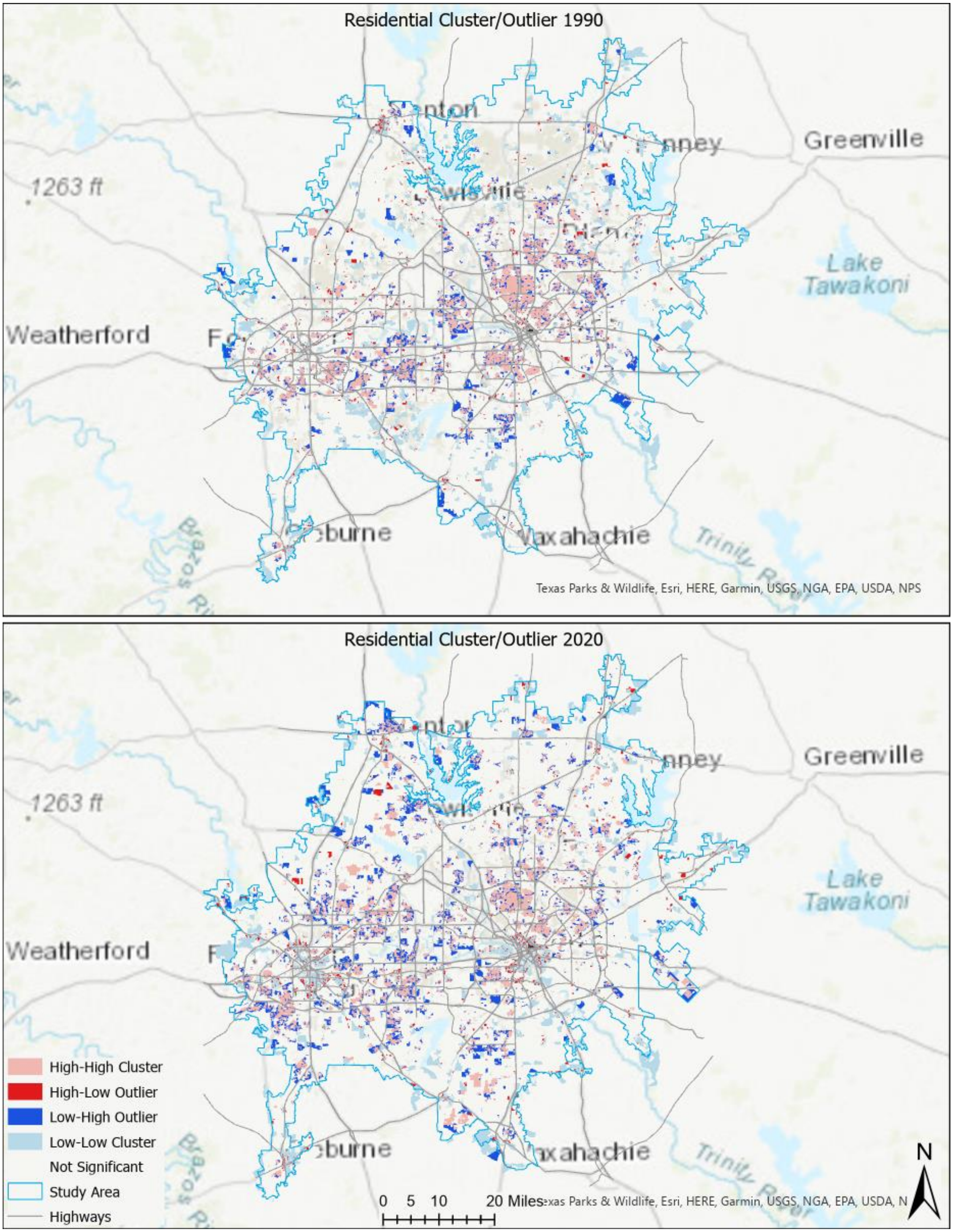
For Activity land uses, most of the clusters in the High-High categories are in the north and northwest of the Metropolitan Area, meaning not only the intensity of Activity land uses in a block is high, but also the high intensity is observable at the neighborhood level. On the other hand, most High-Low outliers of Activity land uses are in the south of downtown Dallas and Fort Worth, indicating the formation of new Activity hubs at the block level. However, they are not as dense as the northern part of the Metropolitan Area. Here, we can see the effect of trickle-down growth, outlined by Perroux's growth pole theory, as the formation of one cluster of Activity land uses brings more Activity land uses to the neighboring blocks. Also, we are observing many Low-High outliers for Activity land uses; these areas indicate blocks of low-intensity Activity land uses surrounded by neighborhoods with high-intensity Activity land uses. These areas have been losing Activity land uses lately.

For Residential land uses, while the number and intensity of High-High clusters are shrinking in 2020 vs. 1990, the number of Low-High outliers is growing more obviously. Neighborhoods with high-intensity Residential land uses surround these low-intensity Residential areas. In other words, these areas have been losing Residential land uses recently. Finally, there are High-Low outliers in both the center (Dallas – Fort Worth corridor, including their downtowns) and the Metropolitan Area's suburbs, indicating the recent formation of residential clusters in both areas, especially in downtown Dallas with many high-low outlier blocks.





Map 4-10: Cluster change for Activity land uses (1990 vs. 2020)



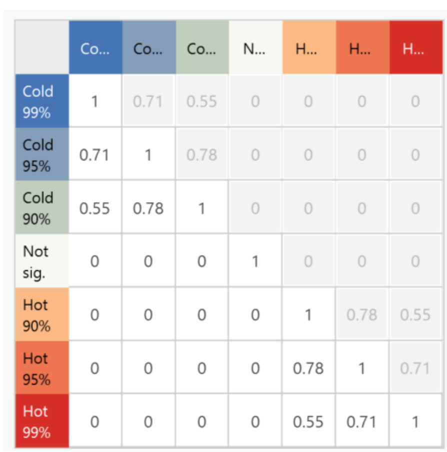
Map 4-11: Cluster change for Residential land uses (1990 vs. 2020)

4.4 Dynamics of Land Use Change

To analyze the dynamics of land use change, we calculated the similarity and association between the hot spots by comparing the results of hot spots for pairs of times (e.g., 1990 and 1995). To assess how the hot spots are similar to each other, we compare the categories of significance level (i.e., 99% hot, 95% hot, 90% hot, not significant, 90% cold, 95% cold, and 99% cold) for each pair of blocks (and their neighboring blocks). To assess the temporal consistency and correlation of the hot spots, we used a kappa statistic to quantify the degree of agreement between the hot spot locations at different time points (Esri, 2023c). For smooth cross-significance evaluation, a fuzzy similarity weight is specified for each significance level.

$$Kappa = \frac{Similarity - Expected\ Similarity}{1 - \ Expected\ Similarity}$$

The closeness of significance levels determines similarity weights in the fuzzy similarity weight. As depicted in Figure 4-23, “the weights between 90%, 95%, and 99% hot and cold spots are determined by ratios of critical values of upper one-sided rejection regions of the normal distribution”[24] (Esri, 2023c).



	Co...	Co...	Co...	N...	H...	H...	H...
Cold 99%	1	0.71	0.55	0	0	0	0
Cold 95%	0.71	1	0.78	0	0	0	0
Cold 90%	0.55	0.78	1	0	0	0	0
Not sig.	0	0	0	1	0	0	0
Hot 90%	0	0	0	0	1	0.78	0.55
Hot 95%	0	0	0	0	0.78	1	0.71
Hot 99%	0	0	0	0	0.55	0.71	1

Figure 4-23: Fuzzy category similarity weight

[24] For instance, the weight between 95% hot and 99% hot is 1.645/2.33 = 0.71

Figure 4-24 and Figure 4-25 for Activity and Residential Land uses, respectively, visualize the counts of each significance level category of the second hot spot result (2020) within categories of the first result (1990).

For Activity land uses, more than 3.5 percent of blocks that were in the cold spots (at various levels of significance) in 1990 became hot spots (at various levels of significance) in 2020. More than 18 percent of not significant blocks in 1990 became hot spots (at various levels of significance) in 2020. Finally, about 2 percent of blocks in the hot spot at 90 and 95 percent significance in 1990 moved to the hot spot at 99 percent hot spot significance in 2020. Overall, about 24 percent of blocks changed to hot spots with various significance levels.

Table 4-2: Hot spot significance level pair (%) for Activity land uses (1990 vs. 2020)

	Hot Spot 2 Significance Level						
Hot Spot 1 Significance Level	Cold 99%	Cold 95%	Cold 90%	Not Significant	Hot 90%	Hot 95%	Hot 99%
Cold 99%	50.00	8.29	4.14	36.74	0.00	0.00	0.83
Cold 95%	16.51	9.08	7.29	63.69	0.69	1.24	1.51
Cold 90%	12.01	8.62	7.91	68.79	0.85	1.27	0.56
Not Significant	4.72	4.88	3.48	71.81	3.37	4.82	6.92
Hot 90%	1.19	2.08	1.49	57.74	8.04	8.93	20.54
Hot 95%	0.58	1.74	1.74	41.47	7.56	17.44	29.46
Hot 99%	0.19	0.28	0.41	28.62	4.04	10.43	56.03

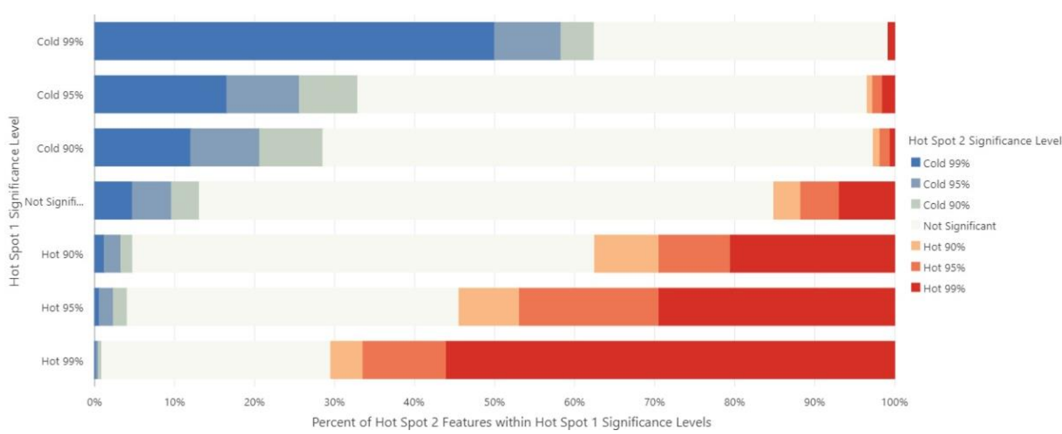


Figure 4-24: Hot spot 1 (Activity 1990) level counts within hot spot 2 (Activity 2020) level categories

For Residential land uses, more than 10 percent of blocks that were in the cold spots (at various levels of significance) in 1990 became hot spots (at various levels of significance) in 2020. More than 15 percent of not significant blocks in 1990 became hot spots (at various levels of significance) in 2020. Finally, more than 17 percent of blocks in the hot spot at 90 and 95 percent significance level in 1990 moved to the hot spot at 99 percent significance level in 2020. Overall, about 42 percent of blocks changed to hot spots with various significance levels.

Table 4-3: Hot spot significance level pair (%) for Residential land uses (1990 vs. 2020)

	Hot Spot 2 Significance Level						
Hot Spot 1 Significance Level	Cold 99%	Cold 95%	Cold 90%	Not Significant	Hot 90%	Hot 95%	Hot 99%
Cold 99%	36.32	7.82	4.77	46.85	1.72	2.24	0.29
Cold 95%	25.94	8.61	3.42	55.67	2.30	2.78	1.28
Cold 90%	22.21	9.85	5.04	54.76	2.77	3.34	2.03
Not Significant	10.91	3.45	2.11	68.36	5.30	6.33	3.54
Hot 90%	4.71	1.87	0.97	62.98	9.66	14.21	5.61
Hot 95%	3.07	1.45	1.04	51.41	14.38	18.49	10.16
Hot 99%	1.21	0.31	0.20	23.47	10.17	24.56	40.08

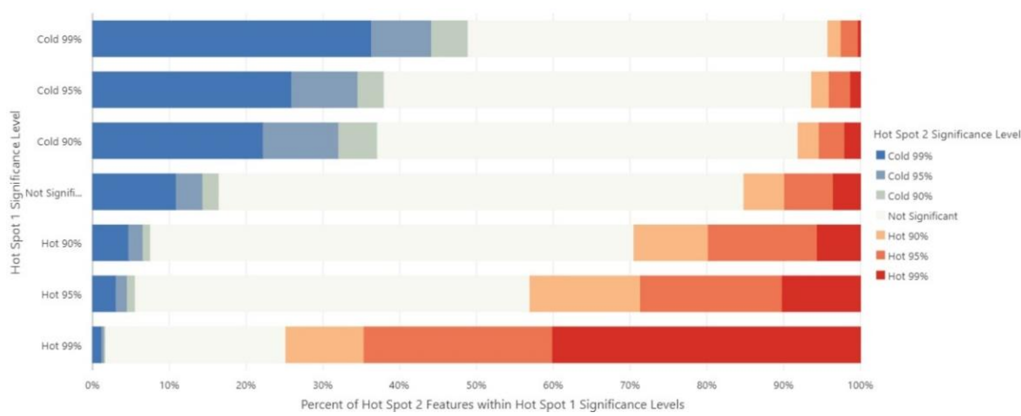
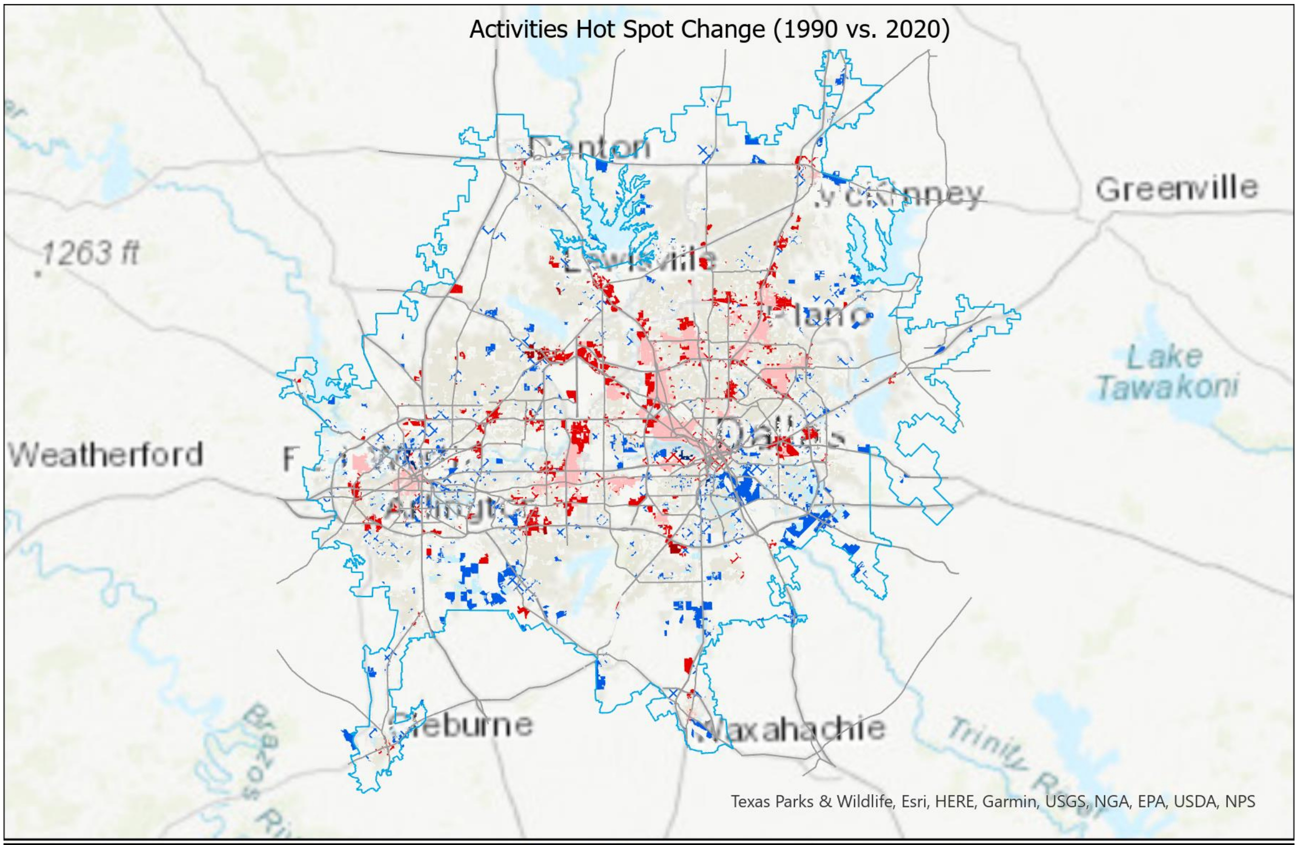
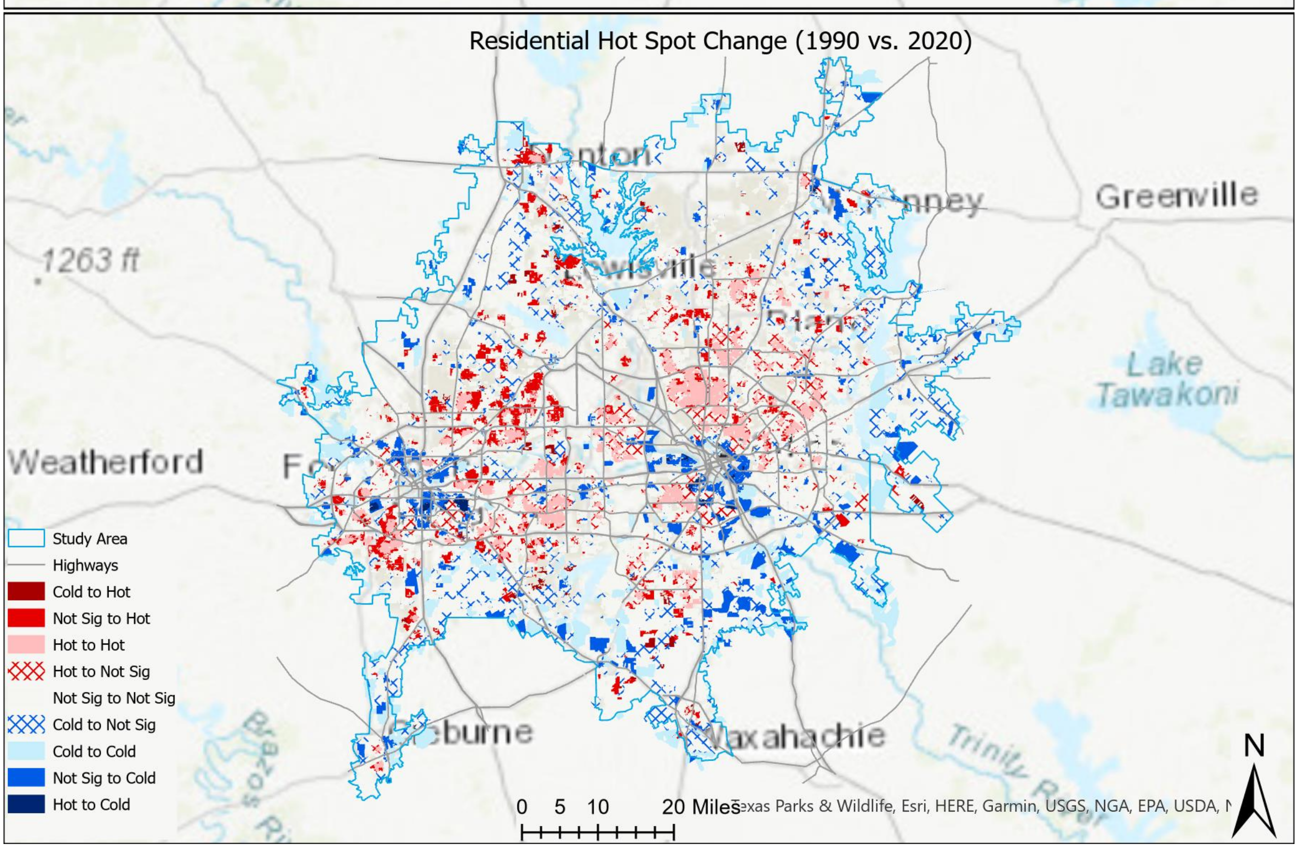


Figure 4-25: Hot spot 1 (Residential 1990) level counts within hot spot 2 (Residential 2020) level categories

The result of the change in hot spots for Activity and Residential land uses between 1990 and 2020 is presented in Map 4-12. In this map, the direction in which new hot spots (that is, these areas were either cold spots or not significant in 1990 but transformed into hot spots in 2020) and new cold spots (that is, these areas were hot spots or not significant in 1990 but transformed to cold spots in 2020) for Activity and Residential land uses are shown in the DFW Urban Area.





Map 4-12: Hot spots change for Activity and Residential land uses (1990 vs. 2020)

4.5 Chapter Summary and Conclusion

Spatial pattern analysis starts with the null hypothesis of Complete Spatial Randomness (CSR), and the goal is to reject it. Since the distribution pattern of Activity and Residential land uses follows the logic in "economies of scale" and "spatial dependency" theories, similar features or associated values tend to cluster, it is not unrealistic to assume that we can reject the null hypothesis. However, how clustered or dispersed these land uses are, where these clusters of land uses are located, how they may have transformed over time, and formed new clusters or evaporated existing clusters are the goals that can be achieved by analyzing land use change patterns using spatial statistics.

Spatial statistics relies on two statistics of z-score and p-value, and the value being analyzed is the ratio of Activity and Residential land uses in blocks. Z-score measures how many standard deviations a value is away from the mean of that value's distribution. P-value is the probability that the observed pattern is random. The distribution pattern is clustered if the z-score is large and positive (and the p-value is small-close to 0). The distribution pattern is scattered if the z-score is large and negative (and the p-value is small-close to 0). In both cases, we reject the null hypothesis of CSR. Otherwise, if the z-score is zero (and the p-value is large-close to 1), the distribution pattern is random, and we cannot reject the null hypothesis of CSR.

In this chapter, we analyzed the distribution of Activity and Residential land uses from 1990 to 2020 using various analytics methods of spatial statistics at the macro and micro levels. For the macro level, we analyzed the land use change trend in the DFW Urban Area, distance from downtown Dallas and Fort Worth, individual cities, and counties the study area encompasses. At the micro level, we used Global Moran's I, Getis-Ord Gi*, and Anselin Local Moran's I at the block level.

The analysis at the macro level shows that Activity land uses are evacuating central areas in the DFW Metropolitan Area in favor of peripheral areas. On the other hand, Residential land uses, while expanding toward suburbs, tend to fill the vacancies in the CBDs created by Activity land uses. This upward trend in residential land uses in CBDs could indicate millennials' (and other younger generations) willingness to live in downtown areas that appreciate their values of walkability, convenience, and environmental sustainability, which are often associated with a downtown-living lifestyle.

On the other hand, the analysis at the micro level reveals that for Activity land uses, Global Moran’s I is getting smaller (74 in 1990 to 49 in 2020) but still positive. For Residential land uses, Global Moran’s I is shifting from large-positive (36 in 1990) to large-negative (-15 in 2020). This indicates the dispersion of Activity and Residential land uses; therefore, the null hypothesis of CSR is rejected. For Getis-Ord Gi* and Local Moran's I, the z-score value is getting smaller for both Activity and Residential land uses, and p-values are slightly increasing, indicating scatteredness of these land uses and forming new hot spots. The new hot spots for Activity land uses are primarily located in the north, east, and south of the DFW Metropolitan Area; the inner parts (including downtown Dallas, Downtown Fort Worth, and the corridor between them) are losing Activity land uses.

A cross-sectional analysis of Activity and Residential land uses changes between 1990 and 2020 shows that about 24 percent of blocks for Activity land uses and 42 percent for Residential land uses changed from not significant or cold spot to hot spot with various significance levels. These changes indicate decentralization, as Activity and Residential land uses spread across the Metropolitan Area by creating new hot spots far away from the Dallas-Fort Worth corridor.

Overall, 21st-century technologies (telecommunication, internet, social media, online shopping, telework, etc.) have impacted the distribution pattern of Activity and Residential land uses in the DFW Metropolitan Area. Economies of scale (agglomeration economies) and spatial dependency theories of land use distributions are weakening in explaining such a distribution pattern, or, at least, need to be reexamined to incorporate the new reality. As a result, the research hypothesis of this chapter is proven. In this regard, Waldo Tobler updated his original spatial dependency theory: Everything is related to everything else, but near and recent things are more related than distant (both space and time) things.

Chapter 5: Drivers of Land Use Change in DFW Metropolitan Area

5.1 Introduction

In Chapter 4, we evaluated Activity and Residential land use changes in the DFW Metropolitan Area from 1990 to 2020 using spatial statistics methods of Global Moran's I, Getis-Ord Gi*, and Anselin Local Moran's I. As discussed and concluded, the overall pattern of land use change for both Activity and Residential land uses is toward decentralization. While peripheral areas and suburbs of the Metropolitan Area are gaining these uses, traditional CBDs and the Dallas - Fort Worth corridor are losing such land uses, and the evacuation rate has accelerated in recent years. This is proven at any level of analysis, including the study area, county, cities, distance from downtown, census tracts, block groups, and blocks.

By comparing the cluster distribution of Activity and Residential land uses in the DFW Urban Area between 1990 and 2020, z-score (as the measure of cluster, scatter, or random) is getting smaller, and p-values (as the probability of the observed pattern) are getting (slightly) larger, indicating the dispersion of Activity and Residential land uses and formation of new clusters for these two land uses beyond the Dallas-Fort Worth downtowns corridor. As a result of these outcomes of spatial statistics, the null hypothesis of Complete Spatial Randomness (CSR) was rejected; it means Activity and Residential land uses, while transforming toward scatteredness of their clusters, are still clustered.

In this chapter, we explore why these changes are happening by discussing the second aspect of the research purpose, which is the drivers of land use change. We will test theoretical and empirical drivers of land use change for Activity and Residential land uses in the DFW Urban

Area using Spatial Panel Data Regression (SPDR). The result helps us to answer the following research question:

- What are the significant factors (site-specific, physical, proximity, socioeconomic, etc.) of land use change?

In this regard, the research hypothesis is that land use change is autoregressive, meaning the ratio of Activity or Residential land uses in a block group at a previous time (t-1) is a significant driver of land use change. Also, the second law of geography, which says, "the phenomenon external to a geographic area of interest affects what goes on inside" (Tobler, 1999), applies to land use change as a spatial phenomenon.

5.2 Panel Data Regression

Overall, three types of regression models are available: cross-sectional, time series, and panel data regression. Cross-sectional regression uses data from a single point in time but from different individuals, groups, or regions[25]. Time series regression uses data from a single individual, group, or region but over multiple periods[26] (Green, 2008). Panel data regression uses data from multiple individuals, groups, or regions over multiple periods[27] (Baltagi, 2021).

Each type of regression model has implications, advantages, and disadvantages; it depends on the research question and assumptions, the availability and quality of data, and the methods being used. Panel data regression, which combines cross-sectional and time series into a single regression model, has several advantages[28] (Hsiao, 2007). First, it can control the unobserved

[25] For instance, cross-sectional regression could analyze the relationship between income and education level across different countries in 2020.
[26] For instance, time series regression could analyze the relationship between GDP growth and inflation in the United States from 1990 to 2020.
[27] For instance, panel data regression could analyze the relationship between health cost and life expectancy for 50 countries from 2000 to 2020.
[28] Panel data regression has challenges and limitations, such as dealing with missing data, endogeneity, serial correlation, or spatial dependence.

heterogeneity among the entities since it can account for the differences in the entities not captured by the observed variables. Second, it can exploit the temporal variation in the data, which means it can use the changes over time to identify the causal effects of some variables. Third, it can increase the efficiency of the estimation by reducing the standard errors and increasing the precision of the estimates. Finally, it can test more complex hypotheses and models than multiple individual regressions.

As with other regression models, panel data regression comprises several components: regression equation, dependent variable (Y), independent/explanatory variables (X), coefficients (β), P-value (ρ), R-Squired (R^2), and residuals (ε). By providing dependent/explanatory variables, the regression model gives us the rest of the model components, by which we can explore the results and make decisions about the goodness of fit of the regression model. Figure 5-1 shows the overall form of a regression model and its components.

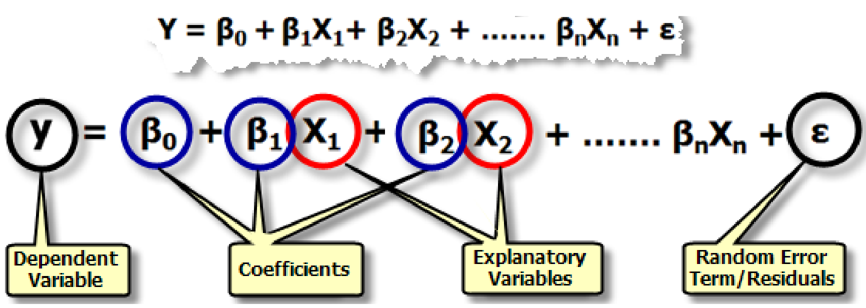


Figure 5-1: General form of a regression model (credit: Esri)

In regression models, even though explanatory variables (Xs) explain changes in the dependent variable (Y), other factors may impact Y, which may be unknown or unmeasurable. If these effects are ignored, it will create omitted variable bias (Burkey, 2018; Kanters, 2022). We use fixed effect and random effect methods to measure the effect of these missing dependent variables in a panel data regression (Elhorst, 2011).

In the fixed-effect model, we assume the effect of missing independent variables (Xs) is probably related to available Xs; that is, the effect of individual, time, or both are correlated with the regression variables. Such an effect can be incorporated into the regression model by adding a dummy variable for each individual, time, or both (Burkey, 2018; Elhorst, 2011).

Three types of fixed effects exist: individual, time, and two-way. In the individual fixed effect model, the values of missing (i.e., unobserved or unmeasurable) variables are related to the individual[29]. In the time-fixed effect, the value of missing variables is related to the periods[30]. Lastly, in a two-way fixed effect, the value of the missing variables is related to both time and individual.[31] The general form of a panel data regression with two-way fixed effects is given by equation 5-1:

$$y_{it} = \alpha_i + \lambda_t + x_{it}\,\beta + \epsilon_{it} \quad (5\text{-}1)$$

Where:

y_{it} is the dependent variable for unit i at time t, α_i is the individual fixed effect, λ_t is the time fixed effect, x_{it} is a vector of covariates, β is a vector of coefficients, and ϵ_{it} is the error term of the regression model.

In the random effect model, however, we assume the effect of missing dependent variables (Xs) is not correlated to other Xs; that is, the unit/time-specific effects are independent of the available regression variables (Baltagi et al., 2007; Burkey, 2018;). In other words, other missing independent variables impact model results unrelated to individual or time. Therefore, we can drop dummy variables related to individual/time effects in the Fixed Effect model and measure it by adding a second parameter to the regression's residual. It is worth noting that the random

[29] For instance, exempting block groups with income lower than the specific threshold impacts land development.
[30] For instance, a major legislature related to a property tax break in 2010 could impact land development in the area.
[31] For instance, a major legislature related to a property tax break in 2010 exempting block groups with income lower than the specific threshold could impact land development in the area.

effect only applies to individuals; there is no random effect for time. The general form of a panel data regression with random effect is given in equation 5-2:

$$y_{it} = \beta_0 + \beta_1 x_{it} + \alpha_i + \epsilon_{it} \quad (5\text{-}2)$$

Where:

y_{it} is the dependent variable for unit i at time t, x_{it} is the independent variable for unit i at time t, β is the vector of coefficients, α_i is the individual-specific effect, and ϵ_{it} is the error term.

As discussed earlier in this chapter, Panel Data Regression has many advantages over cross-sectional and time series regressions. However, in modeling spatial phenomena (like land development and land use change), it suffers from a crucial drawback of not considering interaction effects among geographical units over time (Bera et al., 2020; Brunsdon et al., 1996; Elhorst, 2011; Fotheringham et al., 2002). We will address this issue extensively in the next section while discussing the Spatial Panel Data Regression.

5.3 Spatial Panel Data Regression (SPDR)

As things happen in space, they are more than likely impacted by their neighbors, which is called spatial interaction impacts. The spatial panel data regression model extends the panel data regression model, incorporating these spatial interaction effects among the units. Spatial interaction effects capture the dependence or spillover effects among neighboring or related units[32] (Elhorst, 2012; Elhorst, 2017; Millo et al., 2012; Mínguez, 2020; Tepe, 2023). In these models, the impact of space is measured in three different ways; that is, what happens within

[32] Such as the influence of neighboring states' tax rates or policies on a state's economic outcomes.

individual block groups might be correlated with their neighbors in three different ways (Burkey, 2018):

- The value of dependent variable ***Y*** might impact (or be related to) the ***Y*** amount in a neighboring block group[33], and it is called spatially lagged ***Y*** or Spatial Auto Regression (SAR) because it creates an autoregressive impact. This is related to the economy of scale, as discussed in the previous chapter.
- The value of explanatory/independent variables (***X***) in a block group might affect (or be related to) the ***X*** values in neighboring block groups[34]; this is called Spatially Lagged ***X*** (SLX).
- The residuals (ε), the unexplained value in the model, in the neighboring block groups might impact the residual in the processing block group, called the Spatial Error Model (SEM). It means ε is a function of not only the processing block's ε but also a function of neighboring blocks' ε values.

We may have one, two, or all these impacts in a spatial regression model (including spatial panel regression). Mathematically, equations (5-3) through (5-8) explain these three spatial impacts. The general form of the spatially lagged regression model that includes all three effects (SAR, SLX, SEM) is called the Manski model (Equation 5-3). In this model, *W* is the spatial weight matrix (SWM)[35], which provides the neighboring blocks' information into the model:

$$y = \lambda Wy + X\beta + WX\theta + \mu,\ \mu = \rho W \mu + \varepsilon \quad (5\text{-}3)$$

By dropping Xs ($\theta = 0$) from equation (5-3), we get the Kelejian-Prucha model (5-4), which

[33] Income (as Y) in the neighboring block may impact the Income (Y) in the processing block. It considers Y as an independent variable (X).
[34] Education level (as an X) in the neighboring block may impact the education level in the processing block, which impacts income (Y).
[35] Spatial Weight Matrix (SWM) must be created before running the spatial regression.

includes SAR and SEM:

$$y = \lambda Wy + X\beta + \mu,\ \mu = \rho W\ \mu + \varepsilon \ (5\text{-}4)$$

By dropping μ (ρ = 0) from equation (5-3), we get the Durbin model (5-5), which only includes SAR and SLX:

$$y = \lambda Wy + X\beta + WX\theta + \varepsilon \ (5\text{-}5)$$

We get the SLX model by dropping y (λ = 0) from equation (5-5), which includes SLX and SEM.

$$y = X\beta + WX\theta + \varepsilon \ (5\text{-}6)$$

If θ = 0, then equation (5-5) becomes a SAR model:

$$y = \lambda Wy + X\beta + \varepsilon \ (5\text{-}7)$$

if θ = -ρβ, then equation (5-5) becomes a SEM model:

$$y = X\beta + \mu,\ \mu = \rho W\ \mu + \varepsilon \ (5\text{-}8)$$

Land development models (including land use change) are among the phenomena that can be modeled via spatial regression models (Tepe et al., 2020; Tepe, 2023; Zhou, 2021;) since the status of neighboring parcels and blocks impacts the changes within each parcel and block. This has been extensively discussed in the literature (including in the first[36] and second[37] law of geography by Tobler) and proved via other LULC change models, especially CA-based models (Abdae, 2023; Koomen et al., 2011; Pinto, 2015; Pinto et al.; 2021; Singh, 2003; Verburg et al., 2004; Xu, 2022).

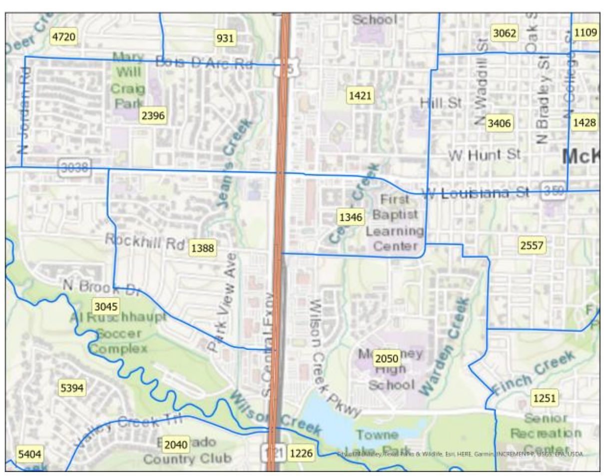
[36] "Everything is related to everything else, but near things are more related than distant things".
[37] What happens in a certain place is not only influenced by the factors within that place but also by the factors outside that place.

Therefore, in this research, we used Spatial Panel Data Regression (SPDR) to model Activity and Residential land uses change in the DFW Metropolitan Area. In the model, we tested the impact of all spatial lags (SAR, SLX, SEM) and presented the performance results of each model. Finally, the model with the best goodness of fit is used for land use change.

5.4 Spatial Panel Data Regression for Land Use Change

This section discusses the Spatial Panel Data Regression (SPDR) model for land use change in the DFW Metropolitan Area. The following actions are taken to prepare and run the SPDR model for Activity and Residential land uses change:

- The Activity, Residential, and Vacant land uses are derived from land use data provided by NCTCOG.
- The Activity and Residential land uses area ratio is calculated and summarized based on census block groups 2020.
- For proximity variables, the average Euclidean distance from each variable (e.g., highway) is calculated and summarized based on block group. (See Map 5-1 as an example).



Map 5-1: Block group distance from the nearest highway

- The US Census Bureau provides socioeconomic variables[38] at the block group level. Due to the mismatch between block group boundaries in different census years, all data is proportionately summarized at the 2020 block group boundaries based on the block group area weighted by the residential area in the block group (Figure 5-2). Then, for each variable, when possible, the ratio is calculated.
- To model the impact of time interval on the model performance, two data sets are created and applied to the model separately: a 10-year interval that includes the years 1990, 2000, 2010, and 2020, and a 5-yeas interval that includes the years 1990, 1995, 2000, 2005, 2010, 2015, and 2020. Since socioeconomic data from 1995 and 2005 is not available at the block group level, the Annual Growth Rate (AGR) between lower and upper values (1990 and 2000 are references for 1995, and 2000 and 2010 are references for 2005) are calculated, then the target years (1995 and 2005) are estimated based on the AGR.

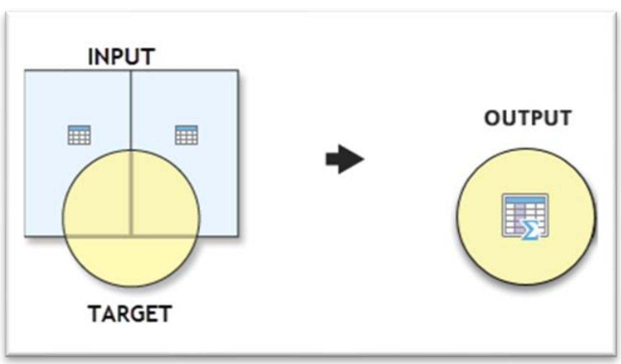


Figure 5-2: Apportion polygon (credit: Esri)

- Pearson correlation is performed to find and remove correlated X variables (collinearity test).
- Various models are tested and evaluated based on these statistics: Adjusted R-squared, Akaike's Information Criterion (AIC), Bayesian Information Criterion (BIC), Variance Inflation Factor (VIF), and Moran's I for residuals.

[38] The data is obtained from Social Explorer and NHGIS.

- Models with the best goodness of fit are tested for the impact of neighborhood type/size and period (i.e., 5-year interval vs. 10-year interval) are tested.
- The final model is applied to evaluate the drivers of land use change for Activity and Residential land uses.

5.5 Regression Variables

This section will discuss the dependent (Y) and independent (X) variables of the Spatial Panel Data Regression (SPDR) model for Activity and Residential land uses. The initial set of variables, selection process, tests, and final variables for the SPDR model will be reviewed and presented.

5.5.1 Dependent Variable

In this research, we are intended to run two separate SPDR models: the "Activity Model" and the "Residential Model." For Activity Model, the dependent variable is the area ratio of Activity land uses in the block group (called "Activity Ratio"). For Residential Model, the dependent variable is the area ratio of Residential land uses in the block group (called "Residential Ratio").

$$Activity\ (Residential)\ Ratio = \frac{Area\ of\ Activity\ (Residential) Land\ uses\ in\ Block\ Group}{Area\ of\ Block\ Group} * 100$$

The area ratio (instead of area) is used to diminish the impact of the size variation of block groups on the dependent variable's value. Because of their size, bigger block groups have more Activity and Residential land uses. On the other hand, the small block groups have fewer Activity and Residential land uses because of their smaller size. Therefore, using the area of land use creates false hot spots and clusters, and results are misleading.

5.5.2 Independent Variables

The independent variables are categorized into proximity, site-specific, socioeconomic, and natural factors. As discussed in the literature review, the explanatory variables impacting land use change might differ in time, space, and model type (land use vs. land cover change). Considering these determining factors in selecting exploratory variables, a set of initial factors that could theoretically and empirically impact land use change are selected and described in Table 5-1. The variable, category, name (as used in the regression model), and a description of each variable are provided for each variable.

Table 5-1: Explanatory variables of land use change for the regression model

Categories	Variables	Variable Name	Measure
Socioeconomic factors	Population Density	Population_Density	Population density in block group
	Minority Population	Minority_Population	Ratio of minority poulation in block group
	Less Than High School Education	Less_HighSchool	Ratio of population older than 25 with no high school degree
	More Than High School Education	More_HighSchool	Ratio of population older than 25 with a high school degree ore more
	School Enrollment	School_Enrollment	Ratio of population older than three years enrolled in school
	Employment	Employment	Ratio of employed population
	Jobs	Jobs	Number of jobs in block group
	Gross Rent	Gross_Rent	Ratio of Average household income paid for rent
	Mortgage	Mortgage	Ratio of housing units with a mortgage in block group
	House Value	House_Value	Median house value in block group
	Land Value	Land_Value	Median land value in block group
	Poverty	Poverty	Ratio of families with income below the poverty level
	Personal Car	Personal_Car	Ratio of workers who use personal car to commute
	Public Transit	Public_Transit	Ratio of workers who use public transit to commute
	Work From Home	Work_Home	Ratio of workers who worked from home
	Phone/Internet Service	Phone	Ratio of housing units with access to phone/internet service
	Income	Income	Median household income
	Travel Time	Travel_Time	Average of maximum travel time to work in block group
Site-Specific Factors	Vacant Land	TL_Vacant_Ratio	Ratio of vacant land in previous time (t-1)
Proximity factors	Airport Proximity	Dist2Airports	Distance from airports
	Downtown Proximity	Dist2Downtowns	Distance from downtowns Dallas and Fort Worth.
	Highway Proximity	Dist2Highways	Distance from highways
	School Proximity	Dist2Schools	Distance from schools
	Lake Proximity	Dist2Lakes	Distance from lakes
	Parks Proximity	Dist2Parks	Distance from major parks
	River Proximity	Dist2Streams	Distance from main rivers/streams
	Non-Passenger Rail Proximity	Dist2NonPassRailNet	Distance from non-passenger rail network
	Passenger Rail Proximity	Dist2PassRailNet	Distance from passenger rail networks (public transit rail)
	Passenger Rail Station Proximity	Dist2PassRailStation	Distance from passenger rail stations (public transit stations)
Natural/physical factors	Slope	Slope_Percent	Average slope percentage in block group
	Elevation	Elevation	Average elevation in block group
	Precipitation	AVG_Precipitation	Average precipitation in block group

Due to the considerable number of variables (32) that could impact land use change, it is necessary to evaluate these variables to ensure they are not highly (positively or negatively) correlated and, therefore, are not causing misleading results. In other words, we need to use in-place guardrails to ensure the integrity and validity of inputs and the credibility of the results by testing the model variables (Burnham et al., 2002). There are numerous statistical tests depending on the nature of the study and the practicality of the tests. In this research, we tested SPDR variables for omitted explanatory variables, nonlinear relationships, data outliers, and multicollinearity. Figure 5-3 shows the schematic variable selection process.

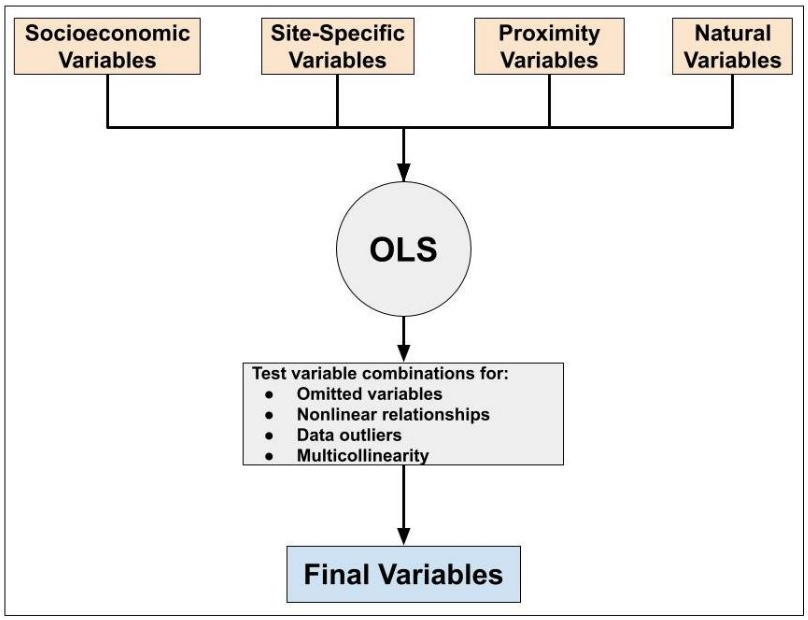


Figure 5-3: Testing process for regression variables

Omitted explanatory variables - If some relevant independent variables are omitted from the model, the estimated coefficients and significance levels may be biased and misleading. Even though we cannot claim that there are no omitted variables, as mentioned earlier, panel data regression can reduce or eliminate the impact of omitted variables via fixed or random effect methods.

Nonlinear relationships - A linear model (such as OLS) assumes the relationship between the dependent variable and each explanatory variable is linear. However, this assumption may not be held when the actual relationship is nonlinear. In such situations, a linear model will not be able to capture the complexity of the data and will result in poor performance. As illustrated by the scatterplot matrix in Figure 5-4 (lower triangle), all variables' relationships are linear; therefore, all variables can be used in the model.

Collinearity - Collinearity refers to the situation where two or more predictors in a regression model have a high degree of correlation. This can affect the model in various ways, such as increasing the standard errors, altering the coefficients, and lowering the model fit. Collinearity can also lead to biased estimates of the effects of the predictor variables, as they may share some of the variance that should be attributed to only one of them. Moreover, Collinearity can make the model unstable and unreliable, as small changes in the data or the model specification can result in large changes in the coefficients and their significance levels. We applied the Pearson Correlation (PC) to test variables for collinearity, and the result of this process is shown in Figure 5-4 (upper triangle). Highly positively or negatively correlated variables are shown in the chart. As it shows, the majority of correlated variables are positively correlated. There are several variables highly correlated (PC >= 0.51), which include Dist2Parks, School_Enrolment, Minority_Population, Poverty, Dist2PassRailNet, Income, Employment, and Personal_Car.

Data outliers - Outliers are data points that differ markedly from the majority of the data, which could indicate errors, anomalies, or special cases that need further investigation. Outliers can cause the regression line to be drawn closer to them and away from the actual relationship between the variables, leading to biased estimates of the regression coefficients and inaccurate predictions of the outcome variable. Even though there is no outlier within tested variables, since spatial regression models use the average values of neighboring block groups, it minimizes the impact of outliers.

Figure 5-4: Scatter plot matrix (lower triangle) and Pearson correlation (upper triangle) between variables

The tests mentioned above resulted in an initial removal of a set of explanatory variables, as described below; most of the omitted variables are due to collinearity (PC >= 0.51) between two or more variables:

- Dist2PassRailNet and Dist2PassRailStation are collinear → Dist2PassRailStation is preserved; rail stations play an important role in developing surrounding areas (Alquhtani, 2017) than rail lines.
- House_Value and Land_Value → Land_Value preserved; house value is also correlated with Income; therefore, it is omitted.
- House_Value and Income → Income preserved; house value is also correlated with land value; therefore, it is omitted.
- Employment and Personal_Car → Personal_Car preserved; employment (*X*) is correlated with the area ratio of Activity Land uses (*Y*).

- Employment and More_HighSchool → Both variables were omitted due to high correlation with many other variables.
- Poverty and Less_HighSchool → Both variables were omitted due to high correlation with many other variables.
- Poverty and Minority_Population → Both variables were omitted due to high correlation with many other variables.
- Minority_Population and Less_HighSchool → Both variables were omitted due to high correlation with many other variables.
- Minority_Population and School_Enrollment → Both variables were omitted due to high correlation with many other variables.
- School_Enrollment and Less_HighSchool → Both variables were omitted due to high correlation with many other variables.
- Dist2Parks and Dist2Schools → Dist2Schools preserved. Even though there is no definitive reason, as buyers of different properties (homes or businesses) may have different preferences and priorities, we believe the distance to school would play a more important role in land use change than the distance to parks.

As the result of the variable selection process, eight independent variables are omitted from further consideration; the remaining independent variables can be categorized into two categories:

- **Time-variant variables** - Variables whose value changes over block group (i) and time (t). These variables are: Vacant_Ratio, Population_Density, Income, Jobs, Land_Value, Mortgage, Gross_Rent, Phone, Travel_Time, Personal_Car, Public_Transit, Worked_Home, Dist2Highways, and Dist2PassRailStation.

- **Time-invariant variables** - Variables whose value changes over block group (i) but is fixed over time (t). These variables are Dist2Downtowns, Dist2NonPassRailNet, Dist2Airports, Dist2Parks, Dist2Schools, Dist2Lakes, Dist2Streams, Elevation, Slope_Percent, and AVG_Precipitation.

The regression models are tested on time-variant and time-invariant variables as part of the model selection process (described in the next section). We tested the regression model once with two categories of variables combined (i.e., 24 variables) and once with only a time-variant category (i.e., 14 variables).

5.6 Regression Model Selection

Regression model selection is choosing the most appropriate model that best fits the data among a set of candidate models, called goodness of fit. It involves determining which independent variables to include and which to exclude from a regression equation. We want a model that is simple but accurate and avoids overfitting and underfitting. Overfitting means the model is too complex and does not generalize well to new data. Underfitting means the model is too simple and misses important patterns in the data (Zellner, 2001).

Since we can use spatial and non-spatial regression models for land use change, we tested both models with different parameters to compare the results and select a model with the best performance. The model selection process happened in two phases. First, we tested various models based on several statistics to find a model with a better goodness of fit. Then, selected models (one model for Activity land uses and one for Residential land uses) were further examined for the impact of spatiotemporal elements on the model performance.

5.6.1 Testing Models for Goodness of Fit

As mentioned earlier, we tested various models based on several statistics to find a model with a better goodness of fit. To do so, the following statistics are tested for each model (separately for Activity and Residential land uses), and results are presented in Table 5-2.

- **Spatial Lag Y**: The value of the dependent variable (Y) in the neighboring block groups at the current time (t) might impact the value of Y in the processing block group. It is calculated based on the weighted average of block groups in the neighborhood of the processing block group.
- **Spatial Lag X**: The value of the independent variable (X) in the neighboring block groups at the current time (t) might impact the value of X in the processing block group. It is calculated based on the weighted average of block groups in the neighborhood of the processing block group.
- **Temporal Lag T**: The value of Y in the previous time (t-1) is a driver for the current (t) value of the dependent variable (Y). Therefore, it is included as one explanatory variable. In this case, the number of periods in the data becomes t-1. In this case, we will have six periods rather than seven.
- **P-value**: A p-value is a numerical measure that helps us evaluate how likely it is that the observed association between two variables in a sample reflects the true association in the population. It assumes that there is no association between the two variables, which is called the null hypothesis. A small p-value (usually less than 0.05) indicates that we can reject the null hypothesis and claim that there is a statistically meaningful association. Conversely, a large p-value (usually greater than 0.05) indicates that we cannot reject the null hypothesis, and there is insufficient evidence to support the association.

- **AIC/BIC**: Akaike Information Criterion (AIC) estimates the goodness of fit of a model by maximizing the likelihood function. Bayesian Information Criterion (BIC), on the other hand, minimizes the Bayesian information loss function. AIC tends to favor more complex models than BIC, especially when the sample size is small. Therefore, AIC might be more suitable for finding the optimal model for prediction purposes, while BIC might be more reliable for choosing the true model. However, both criteria are asymptotically consistent, meaning that they will converge to the same model as the sample size increases. The smaller (i.e., closer to zero) the AIC/BIC is, the better the model performs.
- **Adjusted R-Squared:** The adjusted R-squared is a measure that penalizes the model for adding predictors that do not improve the fit. The adjusted R-squared indicates whether adding more predictors to the model is worthwhile or not. A regression model with more predictors may have a higher R-squared, but it does not necessarily mean that the model is better. The higher the value of Adjusted R-Squired, the better performing the model.
- **Moran's I for Residuals:** Moran's I is a measure of spatial autocorrelation used to test for spatial dependence in regression residuals (as discussed extensively in Chapter 4). It is a statistic that measures the degree of clustering of similar values in space. A positive Moran's I value indicates that similar values are clustered, while a negative value indicates that dissimilar values are clustered. A zero value indicates no spatial autocorrelation, which is preferred for a model residual.
- **Number of Insignificant Variables:** Another indication of a good-performing model is how many variables are/are not significant. Overall, regression models with many insignificant variables are not preferred because having too many insignificant variables can increase the

standard errors and the risk of overfitting and multicollinearity, undermining the model's credibility and interpretability (Kalnins, 2022).

Table 5-2 shows the goodness of fit for various models tested for Activity and Residential land uses. For each model, two sets of independent variables (X) are tested, and the goodness of fit results are presented; one set for all variables, including time-invariable variables (Models 1 - 8 for Activity land uses and 17 – 24 for Residential land uses), and another set for time-variable variables (Models 9 – 16 for Activity land uses and Models 25 – 32 for Residential land uses). The result of analysis and testing various models based on the statistics and criteria mentioned above shows:

- Adding more variables (especially time-invariant variables) does not improve the model's performance. Therefore, Models 1 – 8 for Activity land uses, and Models 17 – 24 for Residential land uses are omitted from further consideration. These models have slightly higher AIC/BIC values than those with fewer but time-variant variables. As mentioned above, AIC/BIC is applicable when comparing models with different numbers of parameters, and the model with the lowest AIC/BIC value is a better fit.
- The models that include Temporal Lag T have higher Adjusted R-squared values than those without such variable. However, they are underperforming regarding AIC/BIC.
- The Spatial Lag X models performed slightly better than those without Spatial Lag X. While for Activity land uses (Models 12 and 16), the number of insignificant variables is smaller for models with Spatial Lag X, for Residential land uses, models without Spatial Lag X have a smaller number of insignificant variables (Models 27 and 31). Also, for these models, the Adjusted R-squared is slightly higher.

Table 5-2: Goodness of fit for various models tested for Activity and Residential land uses

Y	Time-Invariable Xs Included	Model #	Spatial lag Y	Spatial Lag X	Temporal Lag T	Adj R^2	P-Value	AIC	BIC	Moran's I	Total Xs (#)	Insignificant Xs (#)
Activities Area Ratio	Yes	1	No	No	No	0.1898	2.20E-16	-20051	-19859	0.080	23	3
		2	No	Yes	No	0.1176	2.20E-16	-18679	-18487	0.051	23	5
		3	No	No	Yes	0.7521	2.20E-16	-29160	-28968	0.051	24	9
		4	No	Yes	Yes	0.7465	2.20E-16	-28892	-28700	0.038	24	11
		5	Yes	No	No	0.2606	2.20E-16	-21518	-21318	-0.0022	23	10
		6	Yes	Yes	No	0.1631	2.20E-16	-19530	-19330	-0.0174	23	19
		7	Yes	No	Yes	0.7543	2.20E-16	-29267	-29067	0.037	25	9
		8	Yes	Yes	Yes	0.7488	2.20E-16	-29003	-28803	0.023	25	11
	No	9	No	No	No	0.1558	2.20E-16	-19399	-19276	0.093	14	2
		10	No	Yes	No	0.1051	2.20E-16	-18463	-18340	0.063	14	5
		11	No	No	Yes	0.7484	2.20E-16	-28990	-28865	0.056	15	6
		12	No	Yes	Yes	0.7453	2.20E-16	-28845	-28719	0.041	15	4
		13	Yes	No	No	0.2413	2.20E-16	-21114	-20984	-0.0002	14	5
		14	Yes	Yes	No	0.1629	2.20E-16	-19534	-19404	-0.0180	14	11
		15	Yes	No	Yes	0.7501	2.20E-16	-29075	-28942	0.042	16	7
		16	Yes	Yes	Yes	0.7481	2.20E-16	-28979	-28846	0.024	16	3
Residential Area Ratio	Yes	17	No	No	No	0.4817	2.20E-16	-11295	-11103	0.181	23	5
		18	No	Yes	No	0.2469	2.20E-16	-5292	-5100	0.086	23	3
		19	No	No	Yes	0.8064	2.20E-16	-21758	-21566	0.111	24	5
		20	No	Yes	Yes	0.7881	2.20E-16	-20671	-20479	0.105	24	10
		21	Yes	No	No	0.5543	2.20E-16	-13717	-13517	0.051	23	5
		22	Yes	Yes	No	0.3284	2.20E-16	-7131	-6932	-0.0215	23	15
		23	Yes	No	Yes	0.8102	2.20E-16	-21996	-21796	0.089	25	6
		24	Yes	Yes	Yes	0.7924	2.20E-16	-20914	-20714	0.079	25	7
	No	25	No	No	No	0.427	2.20E-16	-9693	-9570	0.200	14	6
		26	No	Yes	No	0.218	2.20E-16	-4696	-4573	0.111	14	1
		27	No	No	Yes	0.8009	2.20E-16	-21429	-21303	0.118	15	1
		28	No	Yes	Yes	0.7849	2.20E-16	-20500	-20374	0.110	15	5
		29	Yes	No	No	0.5399	2.20E-16	-13217	-13086	0.048	14	4
		30	Yes	Yes	No	0.327	2.20E-16	-7107	-6977	-0.0201	14	11
		31	Yes	Yes	Yes	0.8055	2.20E-16	-21710	-21577	0.092	16	2
		32	Yes	No	Yes	0.7902	2.20E-16	-20795	-20662	0.080	16	7

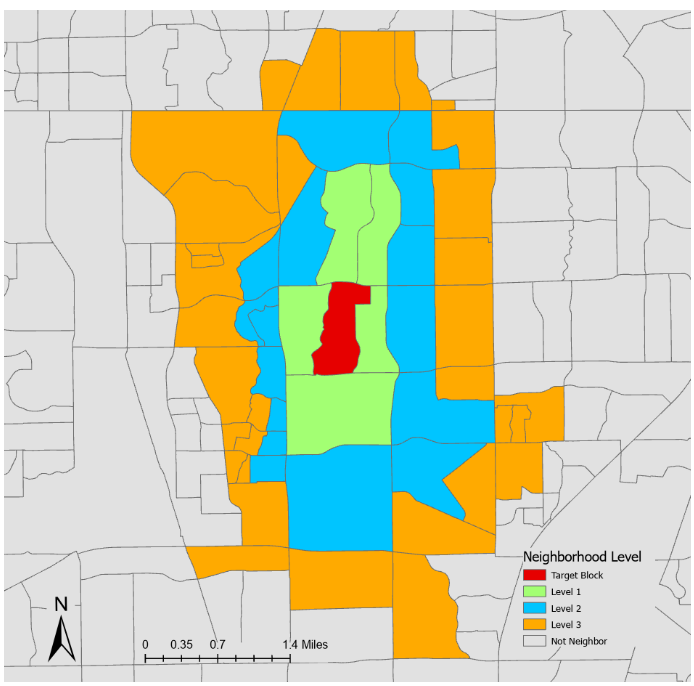
Considering the above observations on the tested models, Model 16 for Activity land uses, and Model 31 for Residential land uses (highlighted in red in Table 5-2) are tested further for the impacts of the scale of analysis, neighborhood type and size, and variables' time interval.

5.6.2 Testing Models for Spatiotemporal Impacts

To test the impacts of scale of analysis, neighborhood type and size, and variables' time-interval on models 16 and 31 (which were selected in the previous step of regression model selection) performance, two more steps are taken: calculating the neighborhood type and size, and variables' time-interval.

In spatiotemporal models, the value being tested (e.g., Activity or Residential Ratio) in each block group (i) is calculated based on the weighted average of neighboring block groups. There are several methods of defining neighborhood (e.g., Queen, Rook, KNN, IDW) and, therefore, calculating weights (SWM); the type of the neighborhood (which defines the size and number of neighbors) is a critical factor that could impact the model results. Therefore, we tested several neighborhood types and sizes. The Queen neighborhood provides the best results and is more appropriate for modeling land use change with a discrete value at the block group level because it guarantees that the impact of all the block groups that touch the border of the processing block group is considered in the neighborhood calculation.

Map 5-2 shows the Queen neighborhood method with three different levels of the neighborhood: the higher the level, the bigger the size of the neighborhood. Results of testing three levels of the Queen neighborhood show that for Activity land uses, Level 2 provides better results; for Residential land uses, however, Level 3 provides a slightly better result. It reveals that the second law of geography impacts Activity and Residential Land uses differently. Therefore, these two levels (Level 2 for Activity land uses and Level 3 for Residential land uses) are selected.



Map 5-2: Queen neighborhoods concept with three levels of neighborhood.

Lastly, since the Socioeconomic data is unavailable for 1995 and 2005, the data for these two times are estimated based on the annual growth rate (AGR). However, the regression model is tested on both 5-year and 10-year intervals to observe the goodness of fit of the regression model for different time intervals. Models based on 5-year intervals provide higher Adjusted R-squared and higher AIC/BIC[39]; therefore, the final model is based on 5-year interval data. The test results are presented in Table 5-3, and the final model for each independent variable is highlighted; models 5-16-3 for Activity land uses, and 5-31-2 for Residential Land uses are selected as final models for Activity land uses change, and Residential land uses change in the DFW Metropolitan Area.

Table 5-3: Selected models tested for the impact of time interval and neighborhood

Y	Intervals (Y)	Model #	Neighborhood	lambda	Lamda (p-value)	Adj R^2	AIC	BIC	Moran I	Moran I (p-value)	Total Xs (#)	Insignificant Xs (#)
Activities Ratio	5	5-16-1	Queen-Level 1	0.096	2.20E-16	0.844	-71515.71	-71386.27	0.117	3.07E-19	15	8
		5-16-2	Queen-Level 2	0.098	2.20E-16	0.844	-71344.19	-71214.75	0.070	2.31E-20	15	8
		5-16-3	**Queen-Level 3**	**0.106**	**2.20E-16**	**0.842**	**-71280.89**	**-71151.45**	**0.048**	**7.08E-19**	**15**	**8**
	10	10-16-1	Queen-Level 1	0.152	2.20E-16	0.724	-29289.53	-29171.18	0.138	1.54E-40	15	7
		10-16-2	Queen-Level 2	0.154	2.20E-16	0.724	-29141.46	-29023.11	0.086	4.14E-49	15	7
		10-16-3	Queen-Level 3	0.150	2.20E-16	0.726	-29072.95	-28954.6	0.058	6.43E-56	15	7
Residential Ratio	5	5-31-1	Queen-Level 1	0.160	2.20E-16	0.807	-55887.22	-55757.78	0.255	1.06E-53	15	5
		5-31-2	**Queen-Level 2**	**0.150**	**2.20E-16**	**0.820**	**-55349.76**	**-55220.33**	**0.160**	**1.40E-49**	**15**	**3**
		5-31-3	Queen-Level 3	0.152	2.20E-16	0.819	-55146.07	-55016.64	0.111	1.00E-43	15	3
	10	10-31-1	Queen-Level 1	0.229	2.20E-16	0.631	-22285.98	-22167.64	0.279	4.92E-122	15	2
		10-31-2	Queen-Level 2	0.206	2.20E-16	0.670	-21858.59	-21740.24	0.170	2.95E-121	15	2
		10-31-3	Queen-Level 3	0.199	2.20E-16	0.679	-21703.17	-21584.82	0.119	6.22E-106	15	2

5.6.3 Fixed/Random Effect Test

Panel Data Regression (including SPDR) must be tested for pooled, fixed, or random effects. Several tests can be run to decide between these tests; we ran the Hausman and Baltagi-Song-Koh SLM tests (Millo et al., 2012).

In the Hausman test, the null hypothesis (H_0) is that the preferred model is random effect; the alternate hypothesis (H_a) is that the model is fixed effect. The result of the Hausman test on both the Activity and Residential tests shows we are dealing with a fixed effect model. However, the

[39] AIC/BIC is used to compare models with different numbers of parameters. It measures how well the model will fit new data, not existing ones. Adjusted R-squared is used to compare models with the same number of parameters. Since we run different models with the same number of variables, the Adjusted R-squared is the reference.

model results are unreliable; for instance, it reverses the impact of proximity to highways (the sign of the coefficient for highways is negative for Residential and positive for Activity, while we expect the reverse).

In the second test, the Baltagi-Song-Koh SLM marginal test, the null hypothesis (H_0) is that the preferred model is fixed effects; the alternate hypothesis (H_a) is that the model is random effects. The result of the Baltagi-Song-Koh test on both the Activity and Residential tests shows we are dealing with a Random Effect model; however, the random effect coefficient is not statistically significant. Due to these contradictory results between the two tests, we ran the land use change model as a pooled SPDR model.

Hausman test results:
Chisq = 11,064
df = 15
p-value < 2.2e-16
Alternative hypothesis (H_0): Fixed Effect

Baltagi-Song-Koh SLM test results:
LM1 = -9.2078
p-value = 2
Alternative hypothesis (H_0): Random Effects

5.7 SPDR Model Results and Discussions

The final form of the Spatial Panel Data Regression (SPDR) model for Activity and Residential land uses change in the DFW Metropolitan Area is the Durbin Model (Equation 5-5), which is programmed within R and R Studio using PLM, SPDEP, and SPLM packages.

$$Y_{it} = \lambda W_{it} Y_{i,t-1} + \beta X_{it} + \theta W_{it} X_{it} + \varepsilon$$

Where:

$\mathbf{Y_{it}}$: Ratio of the Area of Activity/Residential Land Uses in Block Group i at time t.

$\mathbf{W_{it}}$: Spatial Weighted Matrix for Block Group i at time t

$\mathbf{Y_{i,t-1}}$: Ratio of the Area of Activity/Residential Land Uses in Block Group i at time t-1.

$\mathbf{X_{it}}$: Explanatory Variable for Block Group i at time t

λ: Coefficient for autoregression impact

β: Coefficient for independent variable

θ: Coefficient for significance of neighborhood impact

ε: Error term

We will run this regression model separately for Activity and Residential land uses.

5.7.1 Activity Land Uses Model

We ran the Spatial Panel Data Regression (SPDR) model for Activity land uses, with 15 independent variables (14 variables from section 5.5.2 plus Lag of Activities Ratio in t-1, as an independent variable for current time (t)). The results of the SPDR for Activity land use are presented in Table 5-4, which indicates:

Table 5-4: SPDR model results for Activity land uses

```
Model 5-16-3-------------------------------------------------------------

Residuals:
    Min.  1st Qu.   Median     Mean  3rd Qu.     Max.
-0.80796 -0.00445  0.00736  0.01395  0.02391  0.91333

Spatial autoregressive coefficient (Y: Activities Ratio in Block Group):
        Estimate Std. Error t-value  Pr(>|t|)
lambda 0.1055927  0.0087258  12.101 < 2.2e-16 ***

Coefficients:
                        Estimate  Std. Error  t-value  Pr(>|t|)
(Intercept)           6.8355e-03  2.9957e-03   2.2818 0.0225022 *
Lag_Activities_Ratio  9.2549e-01  2.9632e-03 312.3251 < 2.2e-16 ***
TL_Vacant_Ratio       3.1748e-02  1.8806e-03  16.8815 < 2.2e-16 ***
Population_Density   -4.8762e-07  8.0791e-08  -6.0356 1.584e-09 ***
Income                6.3233e-09  1.4490e-08   0.4364 0.6625493
Jobs                  2.9301e-06  8.3807e-07   3.4962 0.0004719 ***
Land_Value            5.3617e-10  1.1076e-09   0.4841 0.6283347
Mortgage              9.6040e-05  2.0634e-03   0.0465 0.9628767
Gross_Rent            1.6831e-03  1.2606e-03   1.3351 0.1818283
Phone                -5.0308e-06  3.6382e-05  -0.1383 0.8900223
Travel_Time          -1.1694e-04  6.4722e-05  -1.8068 0.0707938 .
Personal_Car         -1.0914e-03  1.9614e-03  -0.5565 0.5778892
Public_Transit       -5.3753e-03  9.7985e-03  -0.5486 0.5832920
Worked_Home          -1.2200e-02  7.5952e-03  -1.6063 0.1082160
Dist2Highways        -5.8707e-03  5.6978e-04 -10.3035 < 2.2e-16 ***
Dist2PassRailStation  2.7067e-04  8.0943e-05   3.3439 0.0008260 ***

AIC: -71280.89
BIS: -71151.45
Adj. R-Squared: 0.8424587
Moran's I: 0.1599785 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

- Lag of Activity land uses, the ratio of the area of Activity land uses at the previous time (t-1) as an independent variable for the current time (t), is the most significant driver of Activity land uses change. It proves the impact of the economy of scale and spatial dependency.
- Lambda (λ) is statistically significant; the Activity land uses ratio in the neighboring block groups positively impacts the Activity land uses change (autoregression).
- Distance to highways is a significant driver of land use change and negatively impacts Activity land uses change. The activity ratio in block groups closer to highways is higher than in blocks far away from highways, indicating a close tie between transportation infrastructure and Activity land uses.
- Vacant land ratio in previous time (t-1) positively impacted the Activity land uses change. It indicates that ample vacant land is a driver of land use change and urban sprawl in the DFW Metropolitan Area.
- The relationship between population density and Activity land uses is negative; the higher the population density, the lower the density of Activity land uses (and vice versa).
- Travel time is barely significant (p-value < 0.1) but negatively impacts the Activity land uses. This is an indication of Residential and Activity mixed-use. In other words, Activity and Residential land uses are moving closer; it is another indication of the decentralization in the DFW Metropolitan Area.
- The number of jobs in each block group positively impacts the Activity land uses change, even though the number of jobs and the area of Activity land uses in each block group are not correlated (Figure 5-5). One explanation for the lack of correlation between these two is that the building density is not reflected in calculating the ratio of Activity (and Residential) land uses due to lack of historical data.

- Distance from passenger rail stations, as public transit facilities, which is expected to provide opportunities for mixed-use developments (Alquhtani, 2017; Meyer, 2000), positively impacts the Activity land uses change. It means block groups far distant (to an extent) from rail stations have a higher ratio of Activity land uses. One explanation is that passenger rail stations are mainly located within the center of the DFW Metropolitan Area (Map 5-3). This is another indication that Activity land uses are moving away from the center (especially from the Dallas-Fort Worth corridor) to the periphery of the DFW Metropolitan Area, insofar as public transit facilities are not providing enough attraction in bringing them to the central part of the DFW Metropolitan Area.

Map 5-3: Public transit rail system in the DFW Metropolitan Area

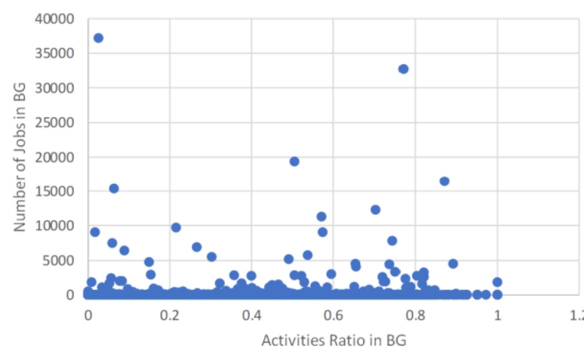


Figure 5-5: Ratio of Activity land uses and number of jobs in block groups

5.7.2 Residential Land Uses Model

We ran the Spatial Panel Data Regression (SPDR) model for Residential land uses, with 15 independent variables (14 variables from section 5.5.2 plus Lag of Residential Ratio in t-1, as an independent variable for current time (t)). The results of the SPDR for Residential land uses are presented in Table 5-5, which indicates:

Table 5-5: SPDR model results for Residential land uses

```
Model 5-31-2-----------------------------------------------------------
Residuals:
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
-0.6687  0.0185  0.0554  0.0559  0.0855  0.9004

Spatial autoregressive coefficient (Y: Residential Ratio in Block Group):
        Estimate Std. Error t-value  Pr(>|t|)
lambda 0.1499109  0.0054126  27.697 < 2.2e-16 ***

Coefficients:
                         Estimate  Std. Error  t-value  Pr(>|t|)
(Intercept)            -7.4447e-02  3.9053e-03 -19.0630 < 2.2e-16 ***
Lag_Residential_Ratio   8.4585e-01  3.4329e-03 246.3951 < 2.2e-16 ***
TL_Vacant_Ratio         8.1412e-02  2.9875e-03  27.2511 < 2.2e-16 ***
Population_Density      4.1169e-06  1.2492e-07  32.9559 < 2.2e-16 ***
Income                  2.3455e-07  2.0055e-08  11.6955 < 2.2e-16 ***
Jobs                   -1.5564e-06  1.1659e-06  -1.3349 0.1819151
Land_Value             -2.5813e-10  1.5423e-09  -0.1674 0.8670839
Mortgage                3.6257e-02  2.9020e-03  12.4939 < 2.2e-16 ***
Gross_Rent              6.6194e-03  1.7527e-03   3.7767 0.0001589
Phone                  -2.8667e-05  5.0634e-05  -0.5662 0.5712906
Travel_Time             1.6051e-04  8.9447e-05   1.7945 0.0727396 .
Personal_Car            2.0864e-02  2.7339e-03   7.6315 2.320e-14 ***
Public_Transit         -1.1361e-01  1.3648e-02  -8.3241 < 2.2e-16 ***
Worked_Home             1.7788e-02  1.0577e-02   1.6818 0.0926160 .
Dist2Highways           5.6992e-03  7.9012e-04   7.2131 5.469e-13 ***
Dist2PassRailStation    6.6319e-04  1.1273e-04   5.8831 4.026e-09 ***

AIC: -55349.76
BIC: -55220.33
Adj. R-Squared: 0.8196399
Moran's I: 0.04794126 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

- Lag of Residential land uses; that is, the ratio of Residential land uses in the previous time (t-1) as an independent variable for the current time (t) is the most significant driver of Residential land uses change. It proves the impact of the economy of scale and spatial dependency.
- Lambda (λ) is statistically significant; the Residential land uses ratio in the neighboring block groups positively impacts the Residential land uses change (autoregression).
- Vacant land ratio in previous time (t-1) positively impacted the Residential land uses change. It indicates that ample vacant land is a driver of land use change and urban sprawl in the DFW Metropolitan Area.

- Travel time and working from home are barely significant ($p\text{-value} < 0.1$); however, they positively impact the Residential ratio. This is an indication of Residential and Activity mixed-use. In other words, Activity and Residential land uses are moving closer. It is another indication of the decentralization in the DFW Metropolitan Area.
- Public Transit (the number of workers using public transit to commute) negatively impacts Residential land uses change, which is corroborated by a positive impact of both distance from rail stations and the personal car. It shows that Residential land uses prefer locating away from public transit facilities, indicating urban sprawl in the DFW Metropolitan Area because people are willing to travel longer distances with their cars. This contradicts the notion that access considerations are essential in residential location decisions, and commuting costs affect its willingness to bid for land with good access to central locations (Hoover et al., 1984).
- The positive impact of distance from highways indicates that Residential land uses prefer being far from highways, even though the highway system in the region provides a good level of accessibility. It shows how urban infrastructure can play a double-edged sword; while providing better accessibility, it can lead to urban sprawl (Barnes et al., 2012; Doxiadis, n.d./1998).
- Distance from passenger rail stations, as public transit facilities, which are expected to provide opportunities for mixed-use developments (Alquhtani, 2017; Meyer, 2000), positively impacts the Residential land uses change (as for Activity land uses). It means block groups far distant (to an extent) from rail stations have a higher ratio of Residential land uses. One explanation is that passenger rail stations are mainly located within the center of the DFW Metropolitan Area (Map 5-3). This is another indication that Residential land

uses are moving away from the center (especially from the Dallas-Fort Worth corridor) to the periphery of the DFW Metropolitan Area, insofar as public transit facilities are not providing enough attraction in bringing them to the central part of the DFW Metropolitan Area.

- The relationship between population density and Residential ratio is positive; the higher the population density, the higher the Residential land uses ratio. Also, income and the number of housing units with mortgages positively impact the Residential land uses change.

5.8 Chapter Summary and Conclusion

The location decision of a firm is complex and depends on various factors. These factors may include the availability of resources, demand for the product or service, cost of production, competition, government policies, and socioeconomic and environmental impacts. Depending on the situation, each factor may have a different weight and influence on the location decision. Thus, it is difficult to determine which factor is the most important or dominant in each case (McCann, 2013).

In this research, we tested the impact of proximity, site-specific, socioeconomic, and natural factors on Activity and Residential land uses change in the DFW Metropolitan Area. The goal was to find the impact of each factor on land use changes by running the Spatial Panel Data Regression (SPDR) model for Activity land uses and Residential land uses.

The dependent variables are the ratio of Activity land uses in block groups, and the ratio of Residential land uses in block groups for the Activity land uses model and Residential land uses model, respectively. Theoretically and empirically, there are 32 possible influential independent variables of land use change. Before executing the models, we tested the variables for omitted explanatory variables, nonlinear relationships, data outliers, and multicollinearity to ensure the integrity and validity of inputs and the credibility of the results.

We tested various spatial and non-spatial panel regression models to select the model with better performance and goodness of fit. The model selection test is done in two phases: First, we tested based on several statistics to find a model with a better goodness of fit. In this regard, for each model, we tested the impact of Lag Y (autoregression), Lag X (impact of neighboring block groups), Lag T (temporal impact), p-value, AIC/BIC, Adjusted R-squared, Moran's I for residuals, and the number of insignificant variables. Then, selected models in phase one (one for Activity land uses and one for Residential land uses) were further examined for the impact of spatiotemporal characteristics on the model performance. These characteristics are neighborhood type, neighborhood size, and period (for data). Also, we tested the final models for pooling, fixed, and random effects using Hausman and Baltagi-Song-Koh SLM tests.

The final SPDR model selected for Activity and Residential land uses change is called the Durbin Model (equation 5-5), which includes spatial lag Y and spatial lag X. In this model, the value of Y in the previous time (t-1) is a driver of land use change at the current time (t); also, the value of both Xs and Ys in the neighboring block groups are impacting land use change in the processing block group. Key findings of the model are:

- Spatial panel data regression models outperform non-spatial panel data regression models due to the inclusion of the spatial elements (i.e., economy of scale and spatial dependency) into the modeling process.
- The results of the Hausman test show SPDR is a Fixed Effect model. However, the fixed effect model's results are unreliable. The Baltagi-Song-Koh SLM marginal test results indicate a random effect model. However, the random effect coefficient is not statistically

significant. Due to these conflicting results, we ran the model without any effects, as pooled spatial panel data regression.

- Results of testing three levels of the Queen neighborhood (Map 5-2) show that neighborhood size plays an essential and distinguishing role in the model. While the level 3 neighborhood provides a better result for the Activity land uses model, the Residential land uses model, the level 2 neighborhood provides a slightly better result. This indicates the implication of the second law of geography. Also, it reveals that the influence of neighboring block groups on the Activity ratio is broader than the Residential Ratio.
- The socioeconomic data for 1995 and 2005 is unavailable, and we estimated the factors for those years based on upper and lower centennial data. To test the impact of time intervals, we rand the model for 10-year and 5-year intervals. The models based on 5-year interval data provide higher Adjusted R-squared and better goodness of fit. It shows that having data in smaller intervals results in a better result.
- The Vacant land ratio in the previous time (t-1) positively impacted both Activity and Residential land uses change. It indicates that ample vacant land is a driver of land use change and urban sprawl in the DFW Metropolitan Area.
- Natural and environmental factors like precipitation, slope, and proximity to major rivers (e.g., Trinity River) are not involved in the model because they are time-invariant; that is, their value does not change over time, at least for short periods – like 30 years of this research.
- Regardless of the data time interval, number of variables, neighborhood type, and effect type (random effect vs. fixed effect), the land use change is highly autoregressive. It means not only does the land use type (and area ratio) in the neighboring block groups impact the land

use change, but also, and more significantly, the ratio of the area of Activity and Residential land uses in the previous time (t-1) impacts land use change. Therefore, the research hypothesis of this chapter is proven. The land use change is autoregressive, meaning the ratio of Activity or Residential land uses in a block group at a previous time (t-1) is a significant driver of land use change. Also, the second law of geography, which says the phenomenon external to a geographic area of interest affects what goes on, applies to land use change.

Table 5-6 summarizes the results of SPDR for Activity and Residential land uses models, in which we can compare the model performance and estimated value, sign, and significance of independent variables for Activity and Residential land uses. The model parameters are provided in the upper part of the table; in the second part, the regression variables and their significance are provided.

Table 5-6: Summary of SPDR results for Activity and Residential land uses models

Variable	Estimate (Activity)	Estimate (Residential)
P-Value	*2.2e-16 ****	*2.2e-16 ****
Adj. R-Squared	*0.8424587*	*0.8196399*
Lambda (λ)	*0.1055927 ****	*0.1499109 ****
Intercept (α)	*6.8355e-03 ***	*-7.4447e-02 ****
Lag_Activity/Residential_Ratio	*9.2549e-01 ****	*8.4585e-01 ****
Vacant_Land_Ratio	*3.1748e-02 ****	*8.1412e-02 ****
Population_Density	*-4.8762e-07 ****	*4.1169e-06 ****
Income	*6.3233e-09*	*2.3455e-07 ****
Jobs	*2.9301e-06 ****	*-1.5564e-06*
Land_Value	*5.3617e-10*	*-2.5813e-10*
Mortgage	*9.6040e-05*	*3.6257e-02 ****
Gross_Rent	*1.6831e-03*	*6.6194e-03*
Phone	*-5.0308e-06*	*-2.8667e-05*
Travel_Time	*-1.1694e-04 **	*1.6051e-04 **
Personal_Car	*-1.0914e-03*	*2.0864e-02 ****
Public_Transit	*-5.3753e-03*	*-1.1361e-01 ****
Worked_Home	*-1.2200e-02*	*1.7788e-02 **
Dist2Highways	*-5.8707e-03 ****	*5.6992e-03 ****
Dist2PassRailStation	*2.7067e-04 ****	*6.6319e-04 ****

Significance codes: '***' 0.01 '**' 0.05 '*' 0.1
Note: Significant code signs are slightly different from those outputted from R to simplify and standardize the code signs.

Chapter 6: Impacts of Land Use Change in DFW Metropolitan Area

6.1 Introduction

As we mentioned in the introduction, the objective of this research is threefold: dynamics, drivers, and impacts of land use change in the DFW Metropolitan Area. We discussed the dynamics and drivers of land use change in Chapters 4 and 5. In this chapter, we will deliberate the impacts of land use change on urban sprawl, traffic flow, and, ultimately, the urban structure of the DFW Metropolitan Area by answering the following research question:

- How did land use change impact the urban structure of the DFW Metropolitan Area?
- What are the impacts on the pattern of the Activity and Residential land uses in terms of loss and gain and intensity changes?
- What are the impacts on the transportation, particularly traffic flow?

In this regard, the research hypothesis is that multiple Activity and Residential land uses hot spots have changed, and new clusters of Activity and Residential land uses have emerged beyond the Dallas-Fort Worth corridor because of land use change.

Studying all impacts of land use change is beyond the scope of this research due to the extent and complexity of the relationships between various formation elements of a Metropolitan Area. However, as one of the research questions, we examined the impact of land use change on physical growth, traffic interactions, and, eventually, the urban structure of the DFW Metropolitan Area. These structural changes could impact further land use change and other aspects (social, economic, etc.) of the Metropolitan Area.

Conceptually, Figure 6-1 depicts that land use change has a cyclical impact on urban structure, urban growth, transportation interactions, and socioeconomic changes. Understanding the causes and effects of land use changes helps local and regional legislators, policymakers, and planners to see the impacts of such development policies and regulations in the Metropolitan Area. Furthermore, via urban and regional policies[40] (McCann, 2013), they can promote policies to improve the attractiveness of urban or suburban areas as locations for investment and intervene with policies to mitigate the negative consequences of these changes. Since the terms Downtown Dallas, Downtown Fort Worth, and Dallas – Fort Worth core corridor are used frequently in this chapter, they are defined intuitively as depicted in Map 6-1.

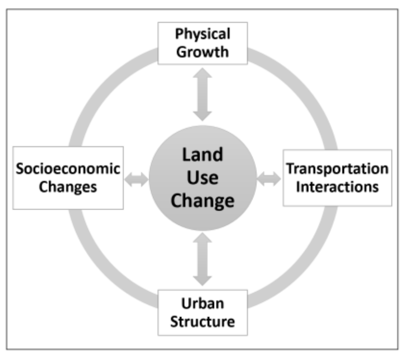
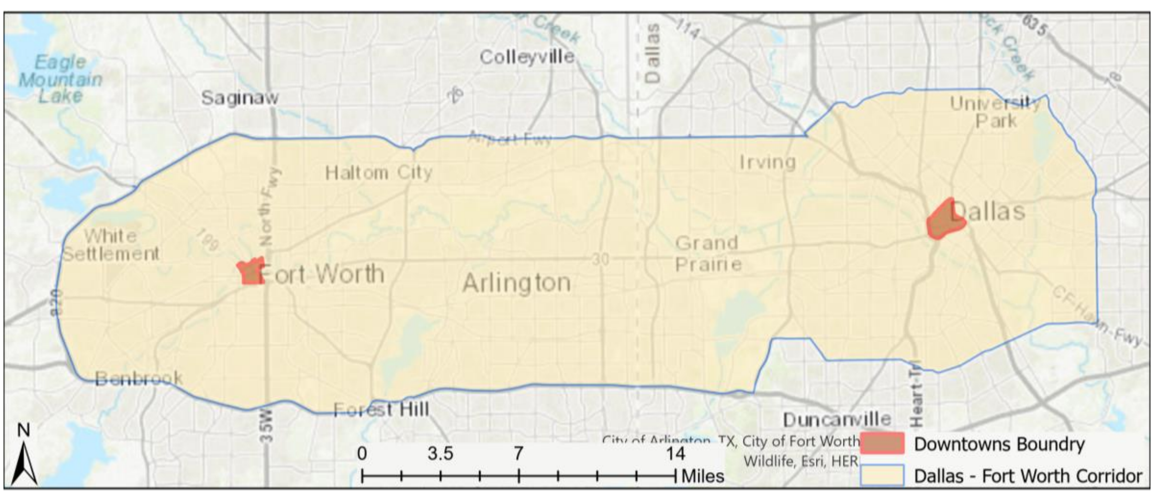


Figure 6-1: Cyclical impact of land use change on urban structure, transportation, population, and economy



Map 6-1: Downtowns Dallas (right) and Fort Worth (left) and the Dallas-Fort Worth corridor

[40] Urban and regional policies differ in scope and scale of coverage, nature of policies, involved institutions, and analytical approach.

6.2 Impact of Land Use Change on Physical Growth (Urban Sprawl)

The debate over urban development is complex and multifaceted. Some people oppose the idea of compact cities, which aim to reduce urban sprawl and increase density (G&R, 1997). They argue that compact cities negatively affect the environment by encroaching on valuable farmland, increasing energy consumption, and reducing transit options. They also claim that compact cities do not match the preferences of most residents, who prefer more spacious and private living conditions.

On the other hand, some people support the idea of compact cities (Ewing, 1997) by suggests active planning as an answer to sprawl. They believe compact cities can enhance urban efficiency and livability. They contend that compact cities can save energy resources, promote TODs, suburbanization costs, revitalize downtown areas, foster social equity, and improve the competitiveness of cities in the global market. They also assert that compact cities can adapt to the changing needs and preferences of residents, especially in the age of advanced telecommunications.

As mentioned in Chapter 1, the analysis of land use distribution in the DFW Urban Area from 1990 to 2020 revealed that the developed area has grown by 77 percent, Activity land uses increased by 83 percent, and Residential land uses increased by 80 percent. This increase in the Activity and Residential land uses is not the same all over the Metropolitan Area regarding the direction of growth and intensity.

In Chapter 4 of this research, we analyzed the dynamics of land use change in the DFW Urban Area. We concluded that the Activity and Residential land uses are still clustered in space, and the null hypothesis of Complete Spatial Randomness (CSR) was rejected. However, the clusters and hot spots of Activity and Residential land uses are dispersed across the region beyond the

Dallas-Fort Worth core corridor. Pairing this finding with the SPDR model results from Chapter 5 and all significant variables (e.g., proximity to rail stations, proximity to highways, personal car usage, public transit usage, travel time, and working from home) indicates decentralization of the DFW Metropolitan Area.

The analysis of Activity and Residential land uses distribution patterns shows that between 1990 and 2020, about 50 percent[41] of new Activity land uses hot spots were formed within cities in Ring 1 and Ring 2 (as defined in Chapter 4, Map 4-5), which includes cities and places located in the NE-N-NW belt. This result is supported by Al-Shammari (2007) findings about employment sub-center locations in the DFW Metropolitan Area.

Also, about 34 percent[42] of Activity land uses hot spots within 2 miles of downtown Dallas and Fort Worth changed to cold spots or not significant (primarily within downtown Dallas). It indicates that the Activity land uses hot spots that are forming in the peripheral areas of the DFW Metropolitan Area are not all new Activity land uses centers but also because of leaving existing Activity land uses from CBDs.

Given the widespread distribution of economic activity in urban regions of the United States, it is not surprising that the central core of a Metropolitan Area has a slower growth rate (based on indicators such as employment, business sales, and daytime population) than the surrounding areas (Hoover et al., 1984).

For Residential land uses, new hot spots mainly formed in the NW-E-S belt, 27 percent of which formed within cities in Rings 1 and 2. On the other hand, the area of hot spots within downtown Dallas and Fort Worth increased by 1.5%, indicating people's willingness to return to downtown

[41] Calculated as the ratio of new hot spots to the area of all hot spots.
[42] Calculated as the area of Activity land uses cold spots divided by the total area of Activity land uses.

for living. It could be related to the walkability and livability of the downtowns and the willingness of Millennials (aka, Generation Y) and Generation Z to live near their workplace in the CBDs[43]. Working from home is a factor that is slightly significant for the Residential land use change model. The trend shows that starting in 2015, work from home gained momentum, and in 2020, the number of people working from home doubled, which is directly related to the Covid-19 outbreak starting in early 2020[44] (Figure 6-2).

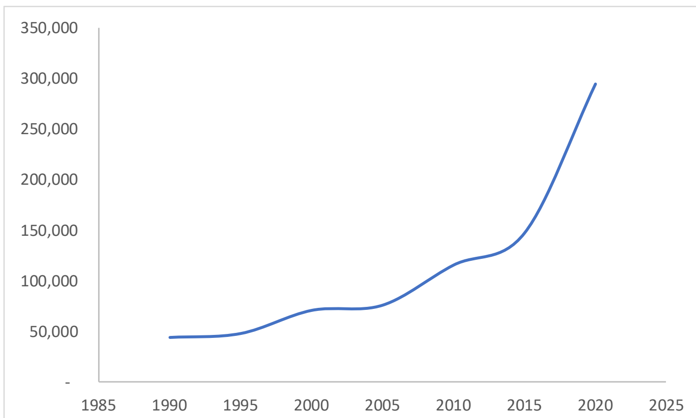
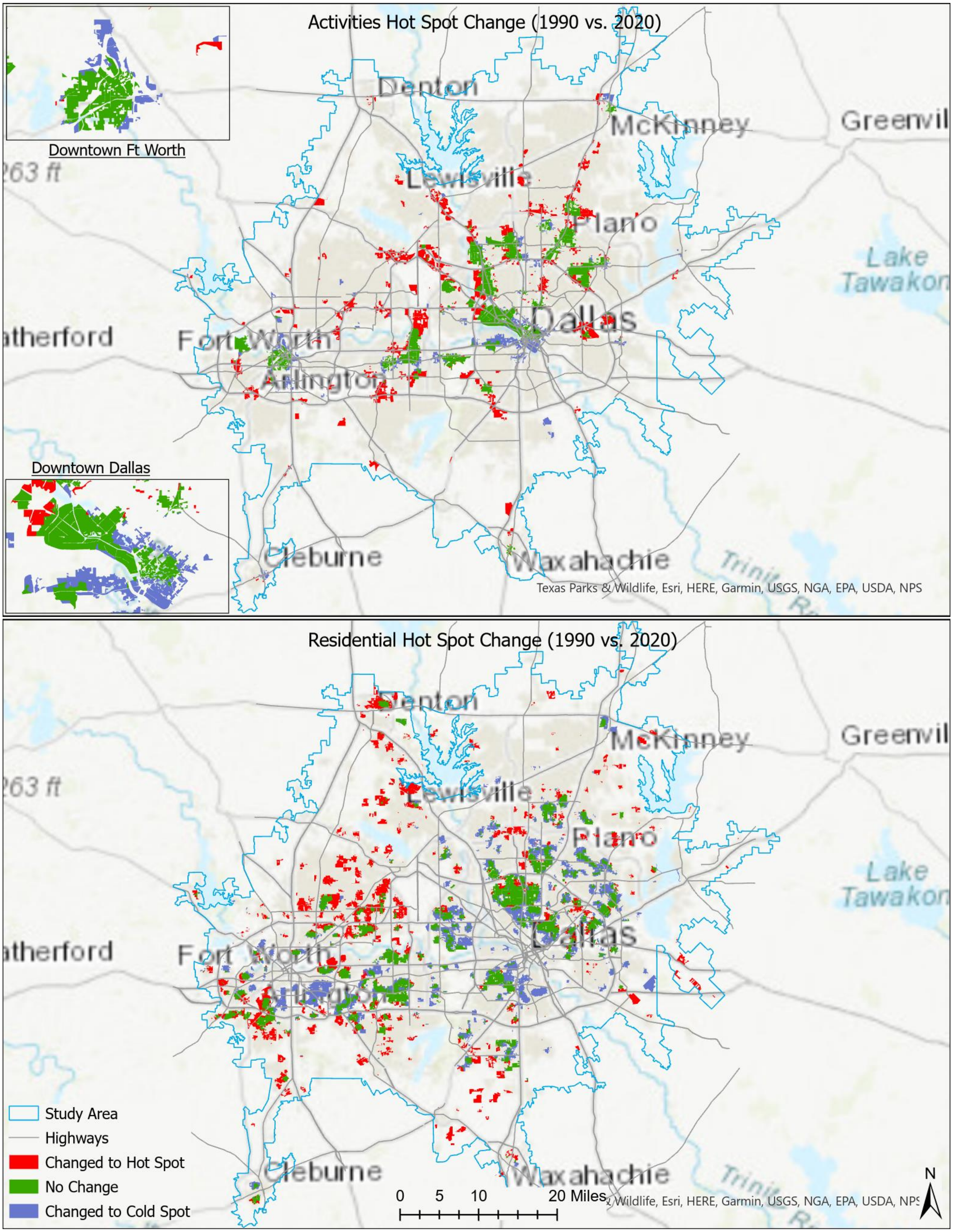


Figure 6-2: Working from home growth between 1990 to 2020 in the study area.

Map 6-2 shows changes in hot spots for Activity and Residential land uses between 1990 and 2020. As highlighted, inner parts (including downtown Dallas, Downtown Fort Worth, and the corridor between them) are losing Activity and Residential land uses since their Activity and Residential density did not change or change to cold spots from 1990 to 2020. New hot spots of Activity land uses are primarily formed perpendicular to the Dallas-Fort Worth corridor in the north (mostly), east, and south of the Metropolitan Area. While expanding toward suburbs, Residential land uses tend to fill the vacancies created in the CBDs (around downtown Dallas and Fort Worth) by Activity land uses.

[43] A separate study is needed to prove and explain why this is happening.
[44] Since the data is from early 2020, the total number of people working from home (350,000) may not reflect the actual number; however, the increased trend reflects the impact of the Covid-19 outbreak on working from home.



Map 6-2: Hot spots change for Activity land uses (top) and Residential land uses (bottom) between 1990 and 2020.

6.3 Impact of Land Use Change on Traffic Flow

The mobility of a community resident is influenced by various indicators that measure the performance of the transportation system. These indicators include, but are not limited to, vehicular congestion, which reflects the amount of traffic on the road network; road level of service, which evaluates the quality of traffic flow based on speed, density, and travel time; and delay hours, which quantify the extra time travelers spend due to congestion (Berke et al., 2006; Meyer et al., 2000). These indicators reflect how easily and quickly a resident can travel from one place to another.

To measure the impact of land use change on traffic flow and congestion in the DFW Urban Area, we analyzed a vehicular congestion indicator called Annual Average Daily Traffic (AADT). AADT is the average daily traffic volume at a given location over a year; higher AADT values indicate denser traffic flow. The Texas Department of Transportation (TxDOT) has collected these data yearly since early 2000s.

The distribution of AADT between 2000 and 2020 shows a significant shift in traffic flow in the DFW Urban Area. As summarized in Table 6-1 and depicted in Figure 6-3, the traffic statistics (e.g., Range, STD, Kurtosis, etc.) indicate the spatial clustered spread of traffic in the region. For example, while the range of AADT is smaller in 2020 (≈252,000) than in 2000 (≈280,000), the Standard Deviation is slightly bigger in 2020. This distribution patter of AADT matches the distribution pattern of Activity and Residential land uses, which discussed in Chapter 4 as scattered hot spots/clusters.

Table 6-1: AADT statistics for 2000 and 2020

Statistics	Min	Max	Range	Mean	Std	Skewness	Kurtosis	Coefficient of Variation
2000	60	280,000	279,940	21,225	36,235	3.5	17.6	1.7
2020	76	252,190	252,114	25,796	37,494	2.8	11.8	1.5

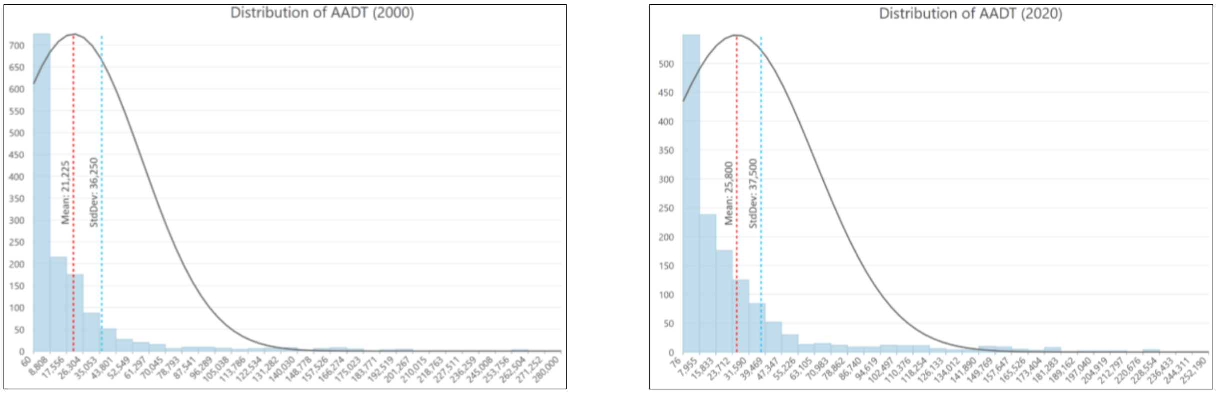


Figure 6-3: Change in the distribution of traffic flow (AADT) from 2000 to 2020

To measure and analyze the impact of land use change on Traffic flow, we used QQ Plot, which plots the quantiles of one numeric variable against the quantiles of a second numeric variable. If the distributions of the two variables are the same, the points on the plot will lie on a diagonal line (straight 45-degree line). The deviation from this line indicates how different the distributions of the two variables are; the higher the deviation, the farther apart the distribution of the two variables is.

We already had derived Z-score (which shows how clustered or dispersed Activity and Residential land uses are), we calculated the average AADT value in each block to match the Z-score units. Then, we cross-examined them with the Z-score of each block for Activity land uses and Residential land uses using the QQ Plot. The results are presented in Figure 6-4 and Figure 6-5 for Activity land uses and Residential land uses, respectively.

For Activity land uses, in 2000, the number of blocks[45] that deviated from the 45-degree line (which falls outside the green area in Figure 6-4) was 12 percent; in 2020, this number is 5 percent, even though the total number of blocks in the hot spots has increased in this period. It indicates that the relationship between Activity land uses and traffic congestion is becoming

[45] The total number of blocks being analyzed (which means they have a value for AADT or Z-score for Activity or Residential land uses) is 23,600.

more similar. Activity land uses hot spots have a higher positive correlation with traffic density in 2020 than in 2000. Also, the straight-line area (green area in Figure 6-4) is smaller in 2020 than in 2000. This implies more intensity for clusters of Activity land uses and traffic distributed in the DFW Metropolitan Areas since more blocks have closer AADT and Z-score values in 2020 (95%) than in 2000 (88%).

All this also implies that the distribution of the Activity land uses and AADT in the hot spot areas has become more picked, meaning their intensity has intensified further. It fulfills the derived demand axiom that easing the ability to reach a destination is an objective of transportation planning (Meyer et al., 2000), in which having well-distributed hot spots of Activity land uses has increased accessibility[46] across the DFW Metropolitan Area.

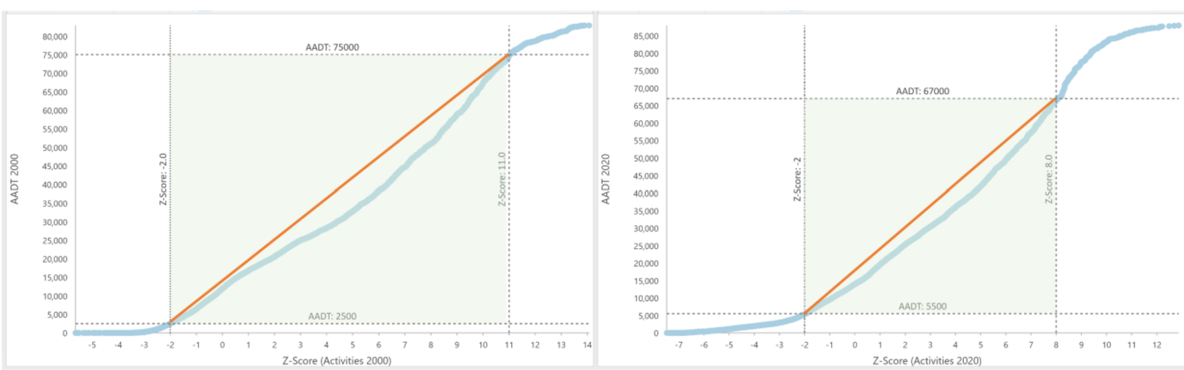


Figure 6-4: Similarity between the distribution of AADT and Z-score for Activity land uses (2000 vs. 2020)

A similar pattern (but at a different level) is observed for the Residential land uses and AADT distribution pattern. While the number of blocks that deviated from the 45-degree line (which falls outside the green area in Figure 6-5) was 34 percent in 2000, in 2020, it is 28 percent, suggesting that the relationship between the distribution pattern of Residential land uses and traffic congestion is getting similar.

---

[46] The difference between relieving traffic congestion and improving the ease of reaching destinations is often portrayed by mobility and accessibility. Whereas mobility connotes movement, fluidity, and one's ability to move through space, accessibility is the ease of getting to destinations (Altshuler et al., 1977).

Also, the straight-line area (green area in Figure 6-5 is smaller in 2020 than in 2000, indicating more clusters of Residential and traffic distributed in the DFW Urban Areas since a larger number of blocks have similar values of AADT and Z-score (72% in 2020 vs. 66% in 2000). A similar intensification discussed earlier for the Activity land uses and AADT can also be derived for the Residential land uses and related AADT here.

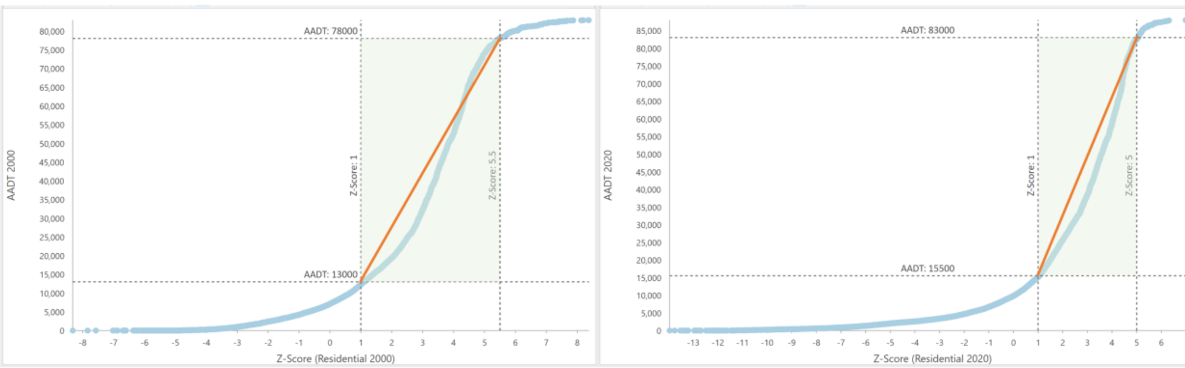
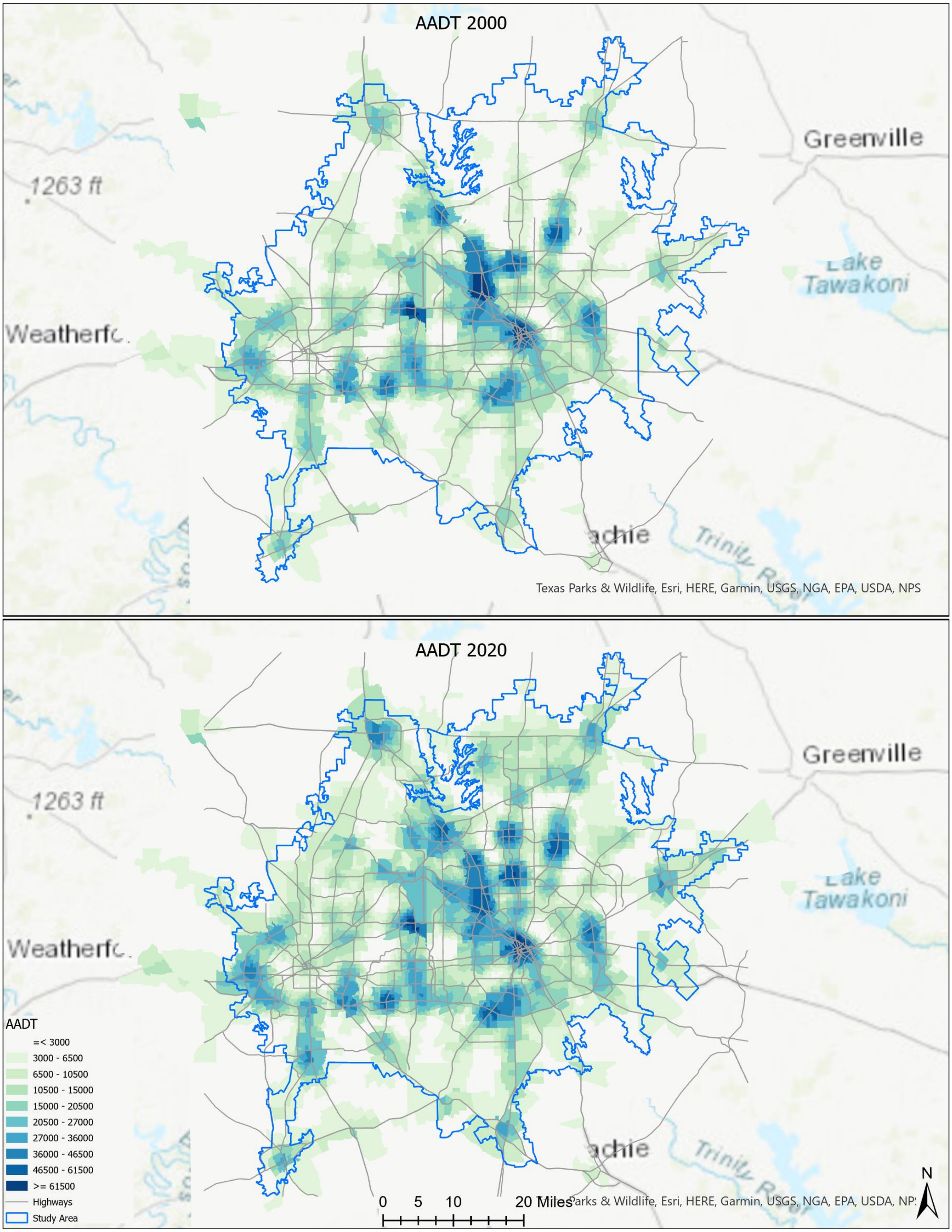


Figure 6-5: Similarity between the distribution of AADT and Z-score for Residential land uses (2000 vs. 2020)

The analysis of AADT and Z-scores distribution for Activity and Residential land uses reveals that the change in distribution pattern of hot spots of Activity and Residential land uses resulted in traffic flow change. As new Activity and Residential hot spots have formed in the peripherals of the Metropolitan Area, the new traffic clusters are forming perpendicularly (north-south), even though the intensity of the traffic flow in the Dallas-Fort Worth corridor (east-west) remains high.

As depicted in Map 6-3, the main new corridors with the highest level of congestion are Dallas-McKinney (along US-75), Dallas-Frisco (along Dallas North Tollway), Dallas-Denton (along I-35E) in the north, and Fort Wort–Burleson-Cleburne (along I-35W) and Dallas-Waxahachie (along I-35E) in the south. Also, there are existing corridors whose level of congestion intensified since 2000. They include Lewisville Lake - Joe Pool Lake (along Highway 360) and Mesquite - Fort Worth (along I-20).



Map 6-3: Average Annual Daily Traffic (AADT) growth between 2000 and 2020

6.4 Impact of Land Use Change on Urban Structure

Urban structure (or urban form) captures the physical features of a city. It includes the dimensions, contours, and arrangements of urban spaces or their components (Živković, 2019). Urban structure can reveal how a city has developed over time, how it functions, and how it relates to its environment. Population density, economic activities, transportation, and environmental conditions influence the spatial distribution of land use in Metropolitan Areas. Among these influential factors, there is a fundamental relationship between land use change and transportation (since one cannot exist without another) in changing the form of a Metropolitan Area.

In such an intertwined relationship, the urban form of a Metropolitan Area is primarily determined by its land development pattern (that is, the land development pattern is the starting point), which affects how the transportation system can function. However, the transportation system can also shape the urban form over time by providing new infrastructure and enhancing accessibility. Therefore, the land development pattern is the initial factor, but the transportation system is the dynamic factor in the urban form of a Metropolitan Area (Meyer et al., 2000).

As has been shown in numerous studies[47], this intertwined relationship between land use change and transportation eventually changes the urban structure of a Metropolitan Area in various shapes and forms, impacting other aspects of urban life in a Metropolitan Area. Some of the highlighted impacts relevant to this research include employment growth, change in land or property value, transfer of land development from one region to another, urban decentralization, suburbs creation, and downtown reinforcement.

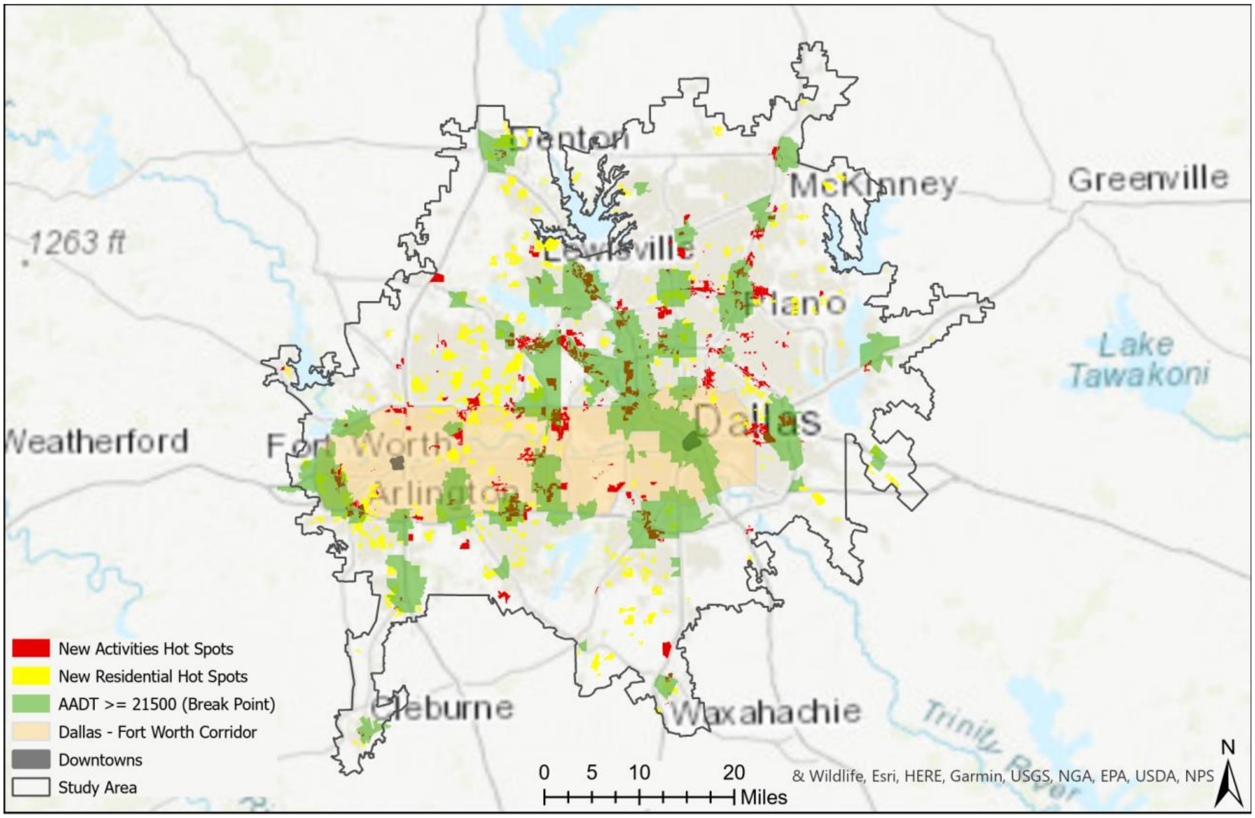
[47] Meyer and Miller provided a summary of studies of land use and transportation impacts (Meyer et al., 2000, pp. 132-134).

As discussed in Chapter 5 (refer to Table 5-6), transit-related variables (Travel_Time, Personal_Car, Public_Transit, Worked_Home, Dist2Highways, Dist2PassRailStation) and site-specific variables (Lag_Activity_Ratio, Lag_Residential_Ratio, Lag_Vacant_Land_Ratio) have significant impact on land use change in the DFW Metropolitan Area. It signifies the close mutual relationship between land use change and transportation.

The central core region of the DFW Metropolitan Area, that is, the Dallas-Fort Worth corridor, is losing Activity and Residential land uses over time. New hot spots of such Activity and Residential land uses are forming in the peripheral areas of the Metropolitan Area toward the north of the Metropolitan Area. Following this development pattern, the traffic flow is intensified toward the north, perpendicular to the Dallas-Fort Worth corridor.

Pairing the distribution pattern of Activity land uses and Residential land uses with traffic data as two pillars (land use and transportation) of the urban form and considering the definition of sprawl, the DFW Metropolitan Area is a sprawled multi-centric Metropolitan Area. It is a region of dynamism and growth, where several Activity and Residential centers transformed, and new Activity and Residential centers emerged. It proves the research hypothesis of the impacts of land uses change, which says multiple hot spots of Activity and Residential land uses have changed, and new clusters of Activity and Residential land uses have emerged beyond the Dallas-Fort Worth corridor because of land use change.

Map 6-4 shows the current urban structure of the DFW Metropolitan Area, resulted from land use change in the past 30 years. The map shows the location of new hot spots for Activity land uses, Residential land uses, and traffic congestion.



Map 6-4: Urban structure of the DFW Metropolitan Area

6.5 Chapter Summary and Conclusion

In this chapter, we combined the results of dynamics of land use change (Chapter 4) and drivers of land use change (Chapter 5) to analyze the impact of land use change on the urban structure of the DFW Metropolitan Area. Key findings are:

- Land use change resulted in urban sprawl and traffic congestion, which eventually changed the urban structure of the DFW Metropolitan Area toward.
- Most of the new hot spots of Activity land uses (and Residential land uses to some degree) are towards the north of the Dallas-Fort Worth corridor. As a result of these new developments, traffic flow spread across the region as new clusters of Activity and Residential land uses formed, mostly perpendicular to the Dallas-Fort Worth corridor.

- Comparing all significant variables resulting from SPDR for the Activity land uses change model and Residential land uses change model indicates decentralization of Activity and Residential land uses in the DFW Metropolitan Area.
- Transit-related and site-specific factors significantly impact land use change; it signifies the close tie between land use change and transportation.
- Land use changes resulted in urban sprawl and traffic congestion. Therefore, the DFW Metropolitan Area has evolved into a sprawled multi-centric Metropolitan Area. It is a region of dynamism and growth, where several Activity and Residential centers transformed, and new Activity and Residential centers emerged. It proves the research hypothesis of the impacts of land uses change.

Chapter 7: Findings, Discussion, and Implications

7.1 Introduction

Land use is a crucial element of urban and regional planning and policymaking; it encompasses more than just the basic categories of land use, and it is reflective of all social, economic, and environmental changes; therefore, it has been the subject of many types of research in planning and related fields. One aspect of land use that has always interested urban spatial modelers, planners, and policymakers is land use change and its distribution pattern because of its potential impacts on other areas in planning.

In this research, we used spatial statistics and Spatial Panel Data Regression (SPDR) to explore Activity and Residential land uses changes in the DFW Metropolitan Area, and we explored three aspects of land use change: dynamics, drivers, and impacts of land use change in the DFW Metropolitan Area.

7.2 Findings and Discussion

As mentioned previously, we studied three aspects of land use change (i.e., dynamics, drivers, and impacts) in the DFW Metropolitan Area, along with one hypothesis related to each aspect. Overall, we concluded that DFW Metropolitan Area is a sprawled multi-centric Metropolitan Area, and all three research hypotheses are proven and accepted:

- Economies of scale (agglomeration economies) and spatial dependency theories of land use distribution are weakening in explaining such a distribution pattern.
- Land use change is autoregressive, meaning the ratio of Activity and Residential land uses in

a block group at a previous time (t-1) is a significant driver of land use change, and the second law of geography applies to the land use change.

- Several hot spots of Activity and Residential land uses have changed, and new clusters of Activity and Residential land uses have emerged beyond the Dallas-Fort Worth corridor because of land use change.

In the following sections, we will provide a summary and discuss the key findings of the research according to each aspect of the research purpose.

7.2.1 Dynamics of Land Use Change

For the dynamics of land use change, we analyzed the distribution of Activity and Residential land uses from 1990 to 2020 at the macro and micro levels. For the macro level, we analyzed the land use change trend in the DFW Urban Area, distance from downtown Dallas and Fort Worth, individual cities, and counties. The results show that Activity land uses evacuate central areas in favor of peripheral areas. On the other hand, Residential land uses, while expanding toward suburbs, tend to fill the vacancies in the CBDs created by Activity land uses.

At the micro level, on the other hand, we used three spatial statistical methods called Global Moran's I (for spatial autocorrelation), (Getis-Ord Gi* (for hot spot analysis), and (Anselin Local Moran's I (for cluster/outlier analysis) at the block level. The results reveal that Global Moran’s I is getting smaller for Activity land uses but is still positive. For Residential land uses, Global Moran’s I is shifting from large-positive to large-negative. This indicates the dispersion of Activity and Residential land uses; therefore, the null hypothesis of Complete Spatial Randomness (CSR) is rejected. For Getis-Ord Gi* and Local Moran's I, the z-score value is getting smaller for both Activity and Residential land uses, and p-values are slightly increasing, indicating scatteredness of these land uses and forming new hot spots for Activity and

Residential land uses. The new hot spots for Activity land uses are primarily located in the north, east, and south of the DFW Metropolitan Area; the inner parts (including downtown Dallas, downtown Fort Worth, and the corridor between the two) are losing Activity land use.

A cross-sectional analysis of hot spots of Activity and Residential land uses between 1990 and 2020 shows that about 24 percent of blocks for Activity land uses and 42 percent for Residential land uses changed from not significant or cold spots to hot spots with various significance levels. These changes indicate decentralization, as new hot spots of Activity and Residential land uses spread across the Metropolitan Area far away from the Dallas-Fort Worth corridor.

7.2.2 Drivers of Land Use Change

To model the drivers of land use change, we used Spatial Panel Data Regression (SPDR) using proximity, site-specific, socioeconomic, and natural factors. The dependent variables are the ratio of Activity land uses and ratio of Residential land uses in block groups for Activity land uses, and Residential land uses models, respectively. The independent variables are 32 theoretical and empirical variables of land use change. We tested variables for omitted explanatory variables, nonlinear relationships, data outliers, and multicollinearity to ensure the integrity and validity of inputs and the credibility of the results.

Also, we tested various spatial panel and non-spatial panel regression models to select the model with better performance and goodness of fit. The model selection test is done in two phases: First, we tested models based on several statistics to find a model with a better goodness of fit. Then, selected models in phase one were further examined for the impact of spatiotemporal characteristics on the model performance. Also, we tested the final models for pooling, fixed, and random effects. The final SPDR model that is selected for Activity and Residential land uses change is a pooled model, which includes spatial lag Y and spatial lag X. Key findings are:

- Spatial panel data regression models outperform non-spatial panel data regression models due to the inclusion of spatial elements into the modeling process.
- Results of testing various neighborhood levels show that the influence of neighboring block groups on the Activity ratio is broader than the Residential ratio. Also, it proves the implication of the second law of geography.
- The Vacant land ratio in the previous time (t-1) positively impacted both Activity and Residential land uses change. It indicates that ample vacant land is a driver of land use change and urban sprawl in the DFW Metropolitan Area.
- Natural and environmental factors like precipitation, slope, and proximity to major rivers (e.g., Trinity River) are not included in the model because they are time-invariant.
- Land use change is autoregressive. It means not only does the land use type (and area ratio) in the neighboring block groups impact the land use change, but more importantly, the ratio of the area of Activity and Residential land uses in the previous time (t-1) impacts land use change.

7.2.3 Impacts of Land Use Change

Lastly, by combining the results of the dynamics and drivers of land use change, we analyzed how land use change impacted the urban structure of the DFW Metropolitan Area. To do so, we cross-examined the Z-score (which shows how clustered or dispersed Activity and Residential land uses are) and distribution of AADT using QQ Plot. Key findings are:

- Most of the new Activity land uses (and Residential land uses to some degree) are towards the north of the Dallas-Fort Worth corridor. As a result of these new developments, traffic flow spread across the DFW Metropolitan Area as new clusters of Activity and Residential land uses formed, mostly perpendicular (north-south) to the Dallas-Fort Worth corridor.

- Comparing all significant variables resulting from SPDR for the Activity land uses change model and Residential land uses change models indicates decentralization of Activity and Residential land uses in the DFW Metropolitan Area. Transit-related and site-specific factors significantly impact land use change; it signifies the close tie between land use change and transportation.
- Land use change resulted in urban sprawl and the spread of traffic congestion across the DFW Metropolitan Area. Therefore, the DFW Metropolitan Area has evolved into a sprawled multi-centric Metropolitan Area; it is a region of dynamism and growth, where several Activity and Residential centers have transformed, and new Activity and Residential centers have emerged.

7.3 Implications of Land Use Change Model Results in DFW Metropolitan Area

A comprehensive model of land use change should be able to address where (location), when (time), why (drivers), and what (impacts) regarding land use change (Sleeter, 2012). From a planner's perspective, land use change is affected by three general factors: the actions of developers who respond to the real estate market demand; the community values and interests that seek to maintain and enhance the quality of life; and plans, policies, decisions, capital investments, and regulations of the government that aim to manage the development of the community (Berke et al., 2006). Therefore, modelers, planners, and policymakers must take land use change seriously because not only does land use change happens due to their policies, decisions, and regulations, but also land use change would impact their policies, decisions, and regulations.

Based on these research findings, which show that hot spots of Activity and Residential land uses have shifted from the Dallas – Fort Worth corridor toward peripheral areas in the DFW Metropolitan Area, we will discuss the policy implications of these land use changes in the following sections.

7.3.1 Modeling Implications

Spatial modelers can model the land use change to understand the growth trends in land use types (e.g., commercial, residential, and industrial), the distribution pattern of land use types, and potential anomalies in a specific land use type. More importantly, they can explain the cause and effect of these growth, distribution patterns, and incongruities.

The modeling approach we used in modeling land use change in the DFW Metropolitan Area (which combines the power of spatial analysis and statistics) delivers all expectations of a land use change model even though, to the best of my knowledge and based on the literature being reviewed, it has not been used before in similar situations. However, there are some challenges and limitations as well, as we will discuss later in this chapter. This modeling approach provides a road map for future similar land use change modeling by overcoming and addressing these challenges and limitations.

7.3.2 Development Policy Implications

In the big picture of land development policies in the DFW Metropolitan Area, planners, policymakers, decisionmakers, legislatures, and other stakeholders should consider changes in land use patterns and the causes and effects of these changes in implementing future development policies. These policies could include specific incentives (e.g., tax exemption) in the areas that have not attracted much Activity or Residential land uses in the past 30 years (e.g., the south side of the Metropolitan Area). Also, via regulations or imposing higher taxes, they can limit further developments in areas that are environmentally or socioeconomically sensitive but have constantly attracted Activity or Residential land uses. These policies could impact the development in desired areas and eventually balance the distribution of Activity and Residential land uses, which subsequently impacts the environmental and socioeconomic aspect of the DFW Metropolitan Area.

7.3.3 Land Use Policy Implications

The DFW Metropolitan Area is becoming a sprawled Metropolitan Area, and the amount of vacant land (which is more than 25 percent of the area of the study area) is playing a key factor; it is the second most significant factor of land use change. Also, the SPDR model results showed that travel time impacts Activity land uses negatively, but Residential land uses positively indicate people's willingness to live close to work and the development of more mixed uses. To avoid further sprawl and mitigate the negative impacts of such development in the DFW Metropolitan Area, we recommend:

- Limiting further urban expansion beyond the current borders of the DFW Metropolitan Area. Otherwise, it would convert more farmland and forests to the built area, which consequently causes more sprawl and environmental issues and increases travel time by creating long-distance trips.
- Promoting more mixed uses in the region by using the vacant lands available within the current borders of the study area. Public transit stations provide an ideal location for such developments (via TODs) because not only do they provide accessibility, but they also promote the use of public transit in the Metropolitan Area.
- Promoting high-density (high rises) developments beyond CBDs (downtown Dallas and Fort Worth). These high-density areas can have more mixed uses, which could promote

walkability and social interactions. The new hot spots of Activity and Residential land uses that resulted from this research are good candidates for such developments.

7.3.4 Transportation Policy Implications

One of the most urgent challenges facing cities today is how to make transportation more sustainable and resilient to the impacts of climate change and, consequently, land use change. Transportation is a major source of greenhouse gas emissions, air pollution, noise, and congestion, affecting urban residents' health, well-being, and productivity. To address this challenge, cities need to adopt a holistic and integrated approach that involves rebuilding mass transit systems and boosting multimodality, electrifying transportation modes and infrastructure, enabling and promoting walking and cycling as active and low-carbon mobility options, building infrastructure that can withstand the effects of extreme weather events and sea level rise, and investing in new technologies that reduce emissions and enhance efficiency. By doing so, cities can reduce their environmental footprint and improve their livability, accessibility, equity, and competitiveness (Bayan et al., 2022).

One of the findings of this research is that not all transportation-related factors have similar impacts on land use change. While the transportation networks (highways) impact is significant, public transportation has not impacted land use change much. One reason for such different impacts is that public transit infrastructure is not well distributed across the DFW Metropolitan Area; therefore, accessibility and coverage of public transit (especially the rail network) is low in the DFW Metropolitan Area.

Three major public transit agencies exist in the DFW Metropolitan Area, which include DART, Trinity Metro, and DCTA. Map 7-1 shows the public transit rail system and the distribution of rail stations (as one of the drivers of land use change in the SPDR model) in the DFW

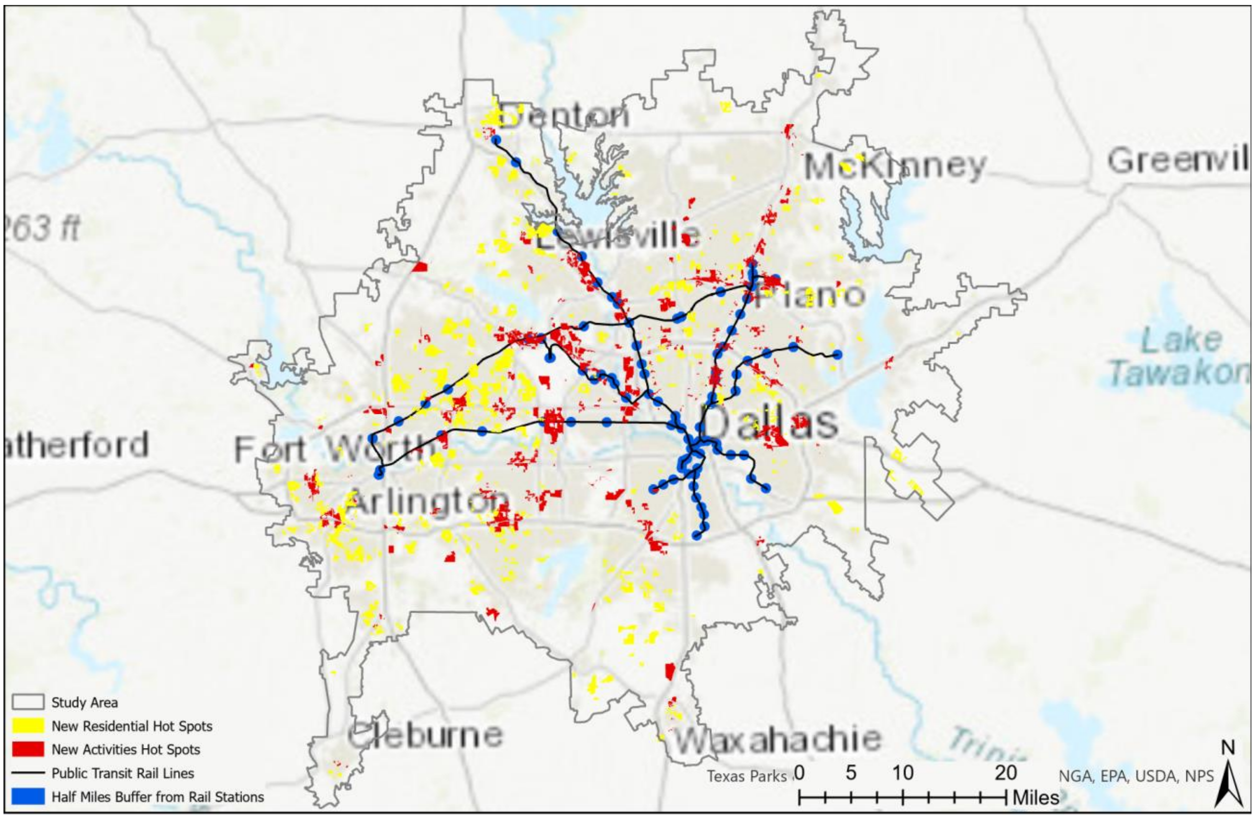
Metropolitan Area along with ½ miles distance (standard walking distance referenced in public transit) from rail stations[48]. As it shows, the public transit rail system is concentrated in the central part of the Metropolitan Area; as a result, it is not playing a substantial role in the distribution pattern of Activity and Residential land uses. We estimated[49] that the population in ½ miles from rail stations[50], based on 2020 census data, is about 10 percent of the population in the study area. Due to the significant impact of transportation on land use change, we recommend:

- Expand public transit infrastructure in the DFW Metropolitan Area via innovative ideas, like micro transit and on-demand services. For instance, DART has expanded its service by implementing GoLink; using a variety of vehicles and providers, it provides curb-to-curb service within a designated zone. Each GoLink zone services a rail station or transit center for connections to other DART services. GoLink has become very popular in the DART service area; its ridership increased by about 425% between 2019 and 2023, while DART bus and rail system ridership plummeted during the same period[51].
- Promote and encourage usage of existing public transit system (bus and rail) by increasing safety, cleanness, on-time performance, improved real-time info through digital technology, better and reliable service frequency, and incentives (free or discounted passes, employer-provided subsidies, reimbursements, partial payments, or pre-tax payroll reductions)[52].
- Decrease the level of investment in the construction of new highways (which encourages more sprawl in the region); instead, increase investment in the public transit infrastructure and local sociable streets.

[48] It includes DART Silver Line, which is under construction, and not operational yet.
[49] Population of block groups that are withing ½ miles of rail stations.
[50] We did not include other modes of transit (e.g., bus) since they are not part of our model.
[51] Even though a big part of this drop is due to Covid-19, however, the overall trend is showing that the ridership is down for rail and bus modes. There is a declining trend in transit ridership in the 2010s (Lee et. al, 2022). Also see https://crsreports.congress.gov/product/pdf/IN/IN11913.
[52] Expand Public Transportation Systems and Offer Incentives | US Department of Transportation

- Empower cities financially and legally to invest in walkable and sociable streets to foster social interaction; it could have social, economic, environmental, and political benefits (Steuteville, 2021).

Implementing these policy recommendations could increase the density of the Metropolitan Area by constructing more mixed-use and transit-oriented developments (TODs). Also, they could increase social interaction.



Map 7-1: Public transit rail system (DART, Trinity Metro, DCTA) in the DFW Metropolitan Area

7.3.5 Socioeconomic Policy Implications

Land use change has various socioeconomic impacts; most of the impacts mentioned in the literature result from land cover change. However, various impacts have resulted from land use change at the metropolitan scale. Wu (2008) mentions loss in community identity, intensification of income segregation, economic disparity among communities, hindering market function, and raising housing prices.

Planners and policymakers can relate the Residential land uses change pattern in the DFW Metropolitan Area to their housing policies for disadvantaged groups to evaluate the effectiveness of their policies in the past three decades (which is the duration of this research). If the results are not as they have been planning for or promoting, they may re-evaluate their policies for a possible shift in their strategies.

For instance, if the focus has been promoting Activity or Residential land uses south of the DFW Metropolitan Area but most of the new hot spots of Activity and Residential land uses formed in the north (which is proven by the findings of this research), they may need to reconsider their policies. On the other hand, if it coincides with their development policies, it proves the effectiveness of such plans and policies, and they might promote more of these policies in other parts within the Metropolitan Area.

Also, cross-examining spatial distribution of income and housing price trends and the location of new hot spots of Activity and Residential land uses gives planners and policymakers ideas on why such a relationship exists and how they can mitigate negative impacts or promote positive impacts.

7.3.6 Environmental Policy Implications

We could not use two key environmental phenomena in the region (flood zones and proximity to Trinity River) as explanatory variables in the SPDR model because they are time-invariant. Therefore, to analyze the impacts of land use change, we overlayed the new hot spots of Activity and Residential land uses by flood zones and distance from the Trinity River.

By exploring the results shown on Figure 7-1 and Map 7-2, it does not seem the proximity to Trinity River is a factor in location preference of new hot spots of Activity and Residential land uses. Overall, most blocks with hot spots of Activity and Residential land uses are located far from a reasonable walking distance from Trinity River. Also, they are located far from hazardous flood zones (zones that are categorized as “Floodway” by FEMA).

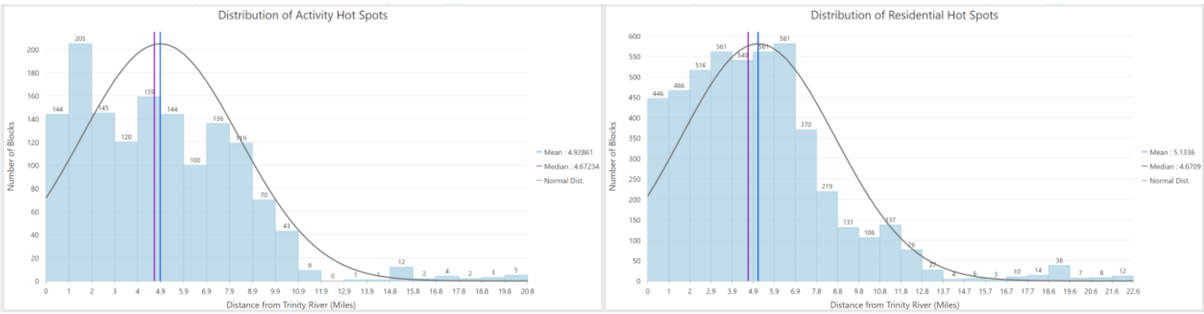
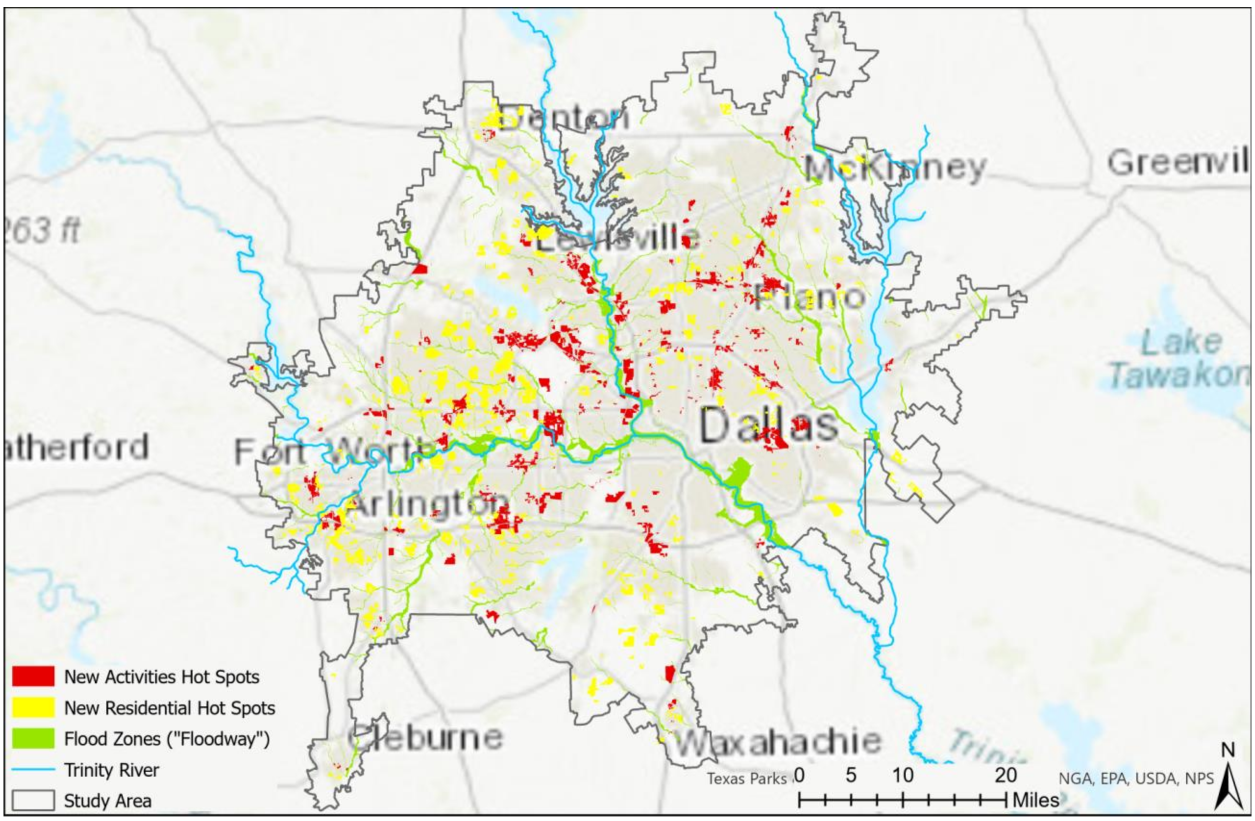


Figure 7-1: Distribution of hot spots of Activity and Residential land uses by distance from Trinity River



Map 7-2: Spatial relationship between hot spots of Activity and Residential land uses, floodway, and Trinity River

The average distance between blocks with hot spots of Activity land uses, and Trinity River is 4.9 miles, and for Residential land uses, the average distance is 5.1 miles, far from a reasonable walking distance. The percentage of blocks with hot spots of Activity and Residential land uses within a 1-mile distance from Trinity River are 10 percent (144/1428) and 9.6 percent (464/4839), respectively.

According to EPA, land development (by creating impervious surfaces through the construction of roads, parking lots, and other structures) and agricultural impacts (by affecting the quality of water and watersheds) are two primary areas of the impacts of land use change, which could have a wide variety of other potential impacts (EPA, 2023). Therefore, the findings of this research, which shows where the locational preferences are for new development via new hot spots for Activity and Residential land uses, can be a road map for mitigating the negative impacts of development on the environment.

Planners and spatial modelers can superimpose the hot spots of Activity and Residential land uses in environmentally sensitive areas in the DFW Metropolitan Area and propose plans and zoning restrictions to limit further developments in areas that are environmentally sensitive but have constantly attracted Activity or Residential land uses in the past 30 years. Policymakers and legislatures, on the other hand, can make incentive-based policies and legislation to enforce recommendations by planners and modelers. These incentivized policies may include development impact fees, development rights purchase, property taxation, and direct payments (Wu, 2008).

7.4 Challenges and Limitations

This research, like any other similar research, is not without challenges. Moreover, some limitations constrain us from extending our research beyond what we have presented here. Below, we discuss some of the challenges and limitations of this research and possible solutions (if applicable) to overcome and address them.

- **Land use data**: This research's historical land use data is at 5-year intervals at the block level. Since the modeling approach we used can model the land use change at the parcel level, having land use data at the parcel level and a smaller time interval (1 year) would make the model results more accurate. Therefore, the result of such a model can be used for short-term planning more reliably.
- **Lack of historical elevation data:** Land use change is not happening only on the ground; it happens vertically too, especially in CBD and downtown areas packed with high rises. The core factor in our model is the ratio of Activity and Residential land uses in each block, which is calculated by the area of the footprint of the building (not the actual area, which is calculated according to the elevation). Therefore, this research lacks elevation (building height or number of floors) inclusion in the model due to missing elevation factors in the land use data. It is critical to keep records of elevation change for inclusion in future land use change models.
- **Missing policy factors:** On top of the socioeconomic, proximity, site-specific, and natural factors discussed in this research, land use change happens because of land development policies (Berke et al., 2006). In a Metropolitan Area like DFW, which includes several big and small cities, development policies are decided for each city independently while impacting the land use change and its distribution pattern in the whole Metropolitan Area. In

this research, we did not include these factors because obtaining them over time is difficult, if not impossible.

- **Computation limitation:** Spatial statistics (for Z-score calculation) and Spatial regressions (including SPDR, which we used in this research) need powerful computer resources (CPU, GPU, and RAM) since they run a regression model for each unit (e.g., block group) based on the impact of neighboring units. Increasing the number of units (e.g., at the parcel level, which could be millions of records for the DFW Metropolitan Area) cannot be done via personal computers; supercomputer is needed.
- **Spatial Panel Data Regression (SPDR) limitations:** SPDR is a powerful regression model for modeling land use change because it considers the historical changes, includes the impact of neighboring units, and considers the impact of omitted variables (via random or fixed effects). However, the current package developed in R (i.e., splm package) is limited, and some land use change assumptions and requirements cannot be considered. First, time-invariant variables, in which the value of the variable does not change over time but could impact land use change (e.g., slope, elevation, proximity to the river, to name a few), cannot be included in the model. Second, SPDR is limited to linear variables, while many factors that could impact land use change are nonlinear. Third, pre-regression variables and model selection tests (e.g., goodness of fit) must be done separately, which is challenging. Fourth, despite having numerous spatial analysis software and packages (e.g., ArcGIS, GeoDa, TerrSet, to name a few), none of them can run SPDR, even though some of them have non-panel spatial regression models (for instance, both ArcGIS pro and GeoDa have such capabilities). It is recommended that these features be included in future releases of these commonly used GIS software. Lastly, policy factors, which usually are considered dummy

variables (i.e., 1 for existence and 0 for nonexistence), cannot be part of the modeling due to the time-invariant nature of such variables.

- **Simultaneous regression:** The land use model can be run as a regression to solve common variables simultaneously. In such models, the dependent variable (e.g., Ration of Activity land uses) in one equation can be the dependent variable in another equation. The current SPDR model cannot run simultaneous regression due to the complexity of modeling and relevant tests to make the results dependable.

Addressing these challenges and limitations helps researchers develop a robust land use change (not land cover change) model that explains the status quo (as our model did) and uses the model for predictions (as planners need it), in which planners and policymakers can anticipate and respond to new developments rather than react to them once they happen.

7.5 Future Research Possibilities

In this research, we tried to shed light on the dynamics, drivers, and impacts of land use change in the DFW Metropolitan Area. As mentioned, several factors impact land use change; some may not be measurable, or more data may need to be used. Therefore, we cannot claim that we covered all aspects of the dynamics of land use change, all drivers of land use change, or all impacts of land use change. However, as a new approach in land use change modeling, our modeling approach provides a road map for further land use change modeling by overcoming the challenges and limitations that we discussed earlier. Here are some relevant areas for future studies:

7.5.1 Impacts of Smart Cities on Land Use Change

A smart city is a dynamic and interconnected system that leverages digital technologies to enhance the quality of life, efficiency, and resilience of its residents, infrastructure, services, and environment. According to Microsoft (n.d.) Key elements of smart cities that could impact land use change in the Metropolitan Area are cloud computing, artificial intelligence, Internet of Things (IoT), edge computing, blockchain technology, and Augmented Reality (AR).

Each of these elements of smart cities could have a different impact on land use change; however, one obvious and common impact is that smart cities could reduce (even eliminate) the need to travel. It would impact locational preferences for various land uses, including Activity and Residential land uses, which are the main focus of this research. For example, using AR, a planner can provide expertise and support while working remotely, or an engineer can inspect the project's site remotely. Online shopping is another example that reduces the need for travel unless a person wants to leave his/her house to socialize while shopping.

Therefore, what we used in this research are (mainly) classical land use change factors. With data availability in a much smaller timeframe, further research is needed to study the impact of smart cities on land use change by incorporating relevant factors. Some possible factors are internet access rate, online shopping rate, online banking rate, working from home beyond the Covid-19 era, number of connected devices (IoT), and electric car ownership (its ownership reduces travel to gas pump since people can recharge their car at home).

7.5.2 Remote Work Impacts

We all know that Covid-19 has had a devastating impact on many aspects of life across the world. One of the most significant impacts of the Covid-19 outbreak is (thanks to internet) that it allows people (white-collar workers, mainly) to work remotely, which was uncommon. Thanks to smart cities solutions discussed previously, most of these workers can work remotely, which reduces daily commute (and other work-related trips); consequently, it changes the locational preference for various land uses.

Covid-19 is not the only outbreak that has forced people to lock down; it will not be the last one either (Marani et al., 2021). One of the findings of this research is that in 2020, there are signs of an increase in Residential land uses in CBDs, and we concluded that it could be related to the willingness of people to live in downtowns as the result of a lack of travel need because of Covid-19 lock-down. Therefore, studying the impacts of the Covid-19 outbreak on how it might impact land use change is another area for further research. The result might also be a helpful lesson learned for better preparation for a probable future outbreak.

7.5.3 Livability of CBDs

New research suggests that "Generation Xers and millennials were more likely to net migrate into central locations and less aversive to high density at their young ages than late boomers were in the 1980s" (Lee et al., 2019, p. 1). In other words, the trend of young Americans living in cities rather than suburbs or rural areas is not a temporary phenomenon but a lasting shift in preferences (Florida, 2019).

One of the patterns that we observed in this research is that the Dallas-Fort Worth corridor (including downtown Dallas and Fort Worth) has been attracting residential land uses after 2015. As discussed in Chapter 4 (Figure 4-3), Residential land uses increased by 1.5 percent between 2015 and 2020 while it had a declining trend before 2015. This increase, aside from the impact of the Covid-19 lockdown, could be related to willingness of Generation Y (Millennials) and Generation Z to live near their workplace in the CBDs, which appreciate their values of walkability, convenience, and environmental sustainability.

Therefore, a deep dive study of the relationship between population age cohorts (e.g., Gen X, Gen Y, and Gen Z) and Residential land uses change would be another area of research to see if such a return-to-downtown trend applies to the DFW Metropolitan Area as well or what is observed is just a temporary impact.

Definitions

Since we use several terms in this research, their definitions are provided here for clarity. General terms are defined here; terms for specific topics are defined in the relevant sections of the research.

- **DFW Urban Area:** The "DFW Urban Area" is the study area based on the census-defined urban area of the Dallas-Fort Worth Metropolitan Area (DFW). It encompasses physically connected cities around the Dallas-Arlington-Fort Worth corridor, modified to fit the census 2020 boundaries. The boundary of the DFW Urban Area is defined according to Map 0-1. In the text, we used "DFW Urban Area" when referring to the analysis results in the study area; otherwise, we used "DFW Metropolitan Area," which refers to the DFW Metropolitan Area in its public notion about the region.
- **Developed land area:** The developed land area is all land use categories within the study area, excluding streets, cemeteries, farmland, flood control, landfill, parks/recreation, ranch land, timberland, vacant lands, and water bodies.
- **Downtown Dallas and Fort Worth:** The boundaries of downtown Dallas and Fort Worth are defined according to Map 0-1.
- **Dallas - Fort Worth corridor:** The boundary of the Dallas - Fort Worth corridor is defined according to Map 0-1.
- **Activity land uses:** Activity land uses (in plural form) include industrial, retail, office, mixed-use, institutional, /semi-public, education, and hotel/motel. These categories of land uses are based on land use categories provided by NCTCOG. In the text we used this term in *plural* form because it refers to a group of activity-related land uses.

- **Residential land uses:** Residential land uses (in plural form) include single-family, multi-family, mobile homes, residential acreage, and group quarters. These categories of land uses are based on land use categories provided by NCTCOG. In the text we used this term in *plural* form because it refers to a group of residential land uses.
- **First Law of Geography:** "Everything is related to everything else, but near things are more related (similar) than distant things" (Tobler, 1970).
- **Second Law of Geography:** "The phenomenon external to a geographic area of interest affects what goes on inside" (Tobler, 1999).

Map 8-1:DFW Urban Area (Study Area), downtown Dallas, downtown Fort Worth, and Dallas-Fort Worth corridor

References

- ACS 2010, 2015, 2020 (5-Year Estimates). (n.d.). In SocialExplorer.com, Retrieved May 1, 2023, from https://www.socialexplorer.com/explore-tables
- Addae, B., & Dragićević, S. (2023). Modeling global urban land-use change process using spherical cellular automata. GeoJournal 88, 2737–2754 (2023), DOI: https://doi.org/10.1007/s10708-022-10776-4
- Alig, R. J., & Healy, R. G. (1987). Urban and Built-Up Land Area Changes in the United States: An Empirical Investigation of Determinants. Land Economics, 63(3), 215-226
- Alonso, W. (1964). Location and Land Use: Toward a General Theory of Land Rent. Cambridge, MA: Harvard University Press
- Alquhtani, S. M. (2017). Socioeconomic and spatial impacts of rail transit stations on the surrounding areas in Dallas-Fort Worth region: changes in housing values, racial make-up, and population density. [Unpublished Doctoral dissertation]. University of Texas at Arlington.
- Al-Shammari, M. S. (2007). Growth/decline Of Employment Subcenter in Polycentric Regions: The Case of The Dallas-Fort Worth Metropolitan Area. [Unpublished Doctoral dissertation]. University of Texas at Arlington.
- Amgalan, A., Mujica-Parodi, L. & Skiena, S. S. (2022). Fast spatial autocorrelation. Knowl Inf Syst 64, 919–941 (2022). DOI: https://doi.org/10.1007/s10115-021-01640-x
- Anderson, J. R., Hardy, E. E., Roach, J. T., & Witmer R. E. (1976). A Land Use and Land Cover Classification System for Use with Remote Sensor Data. U.S. Department of Interior

- Anjomani, A. (2021). An integrated land-use/transportation forecasting and planning model: A metropolitan planning support system. Journal of Transport and Land Use, 14(1), 65–86. https://www.jstor.org/stable/48646177
- Anselin, L. (1988). Spatial econometrics: methods and models. Berlin, Germany: Springer Netherlands, 4. DOI: 10.1007/978-94-015-7799-1
- Anselin, L. (1995). Local Indicators of Spatial Association-LISA, Geographical Analysis 27(2): 93–115, 1995.
- APA. (1960). Cluster Subdivisions. APA-PSA Report 135, Retrieved August 20, 2023, from https://planning-org-uploaded-media.s3.amazonaws.com/document/PAS-Report-135.pdf
- Baker, W. L. (1989). A review of models of landscape change. Landscape Ecology, 2(2), 111-133.
- Baltagi, B. H. (2021). Econometric Analysis of Panel Data (6th edition). Springer.
- Baltagi, B. H., Egger, P., & Pfaffermayr, M. (2007). A generalized spatial panel data model with random effects, Retrieved August 20, 2023, from http://ideas.repec.org/p/max/cprwps/113.html
- Banzhaf, E., Reyes-Paecke, S., Müller, A., & Kindler, A. (2013). Do demographic and land-use changes contrast urban and suburban dynamics? A sophisticated reflection on Santiago de Chile. Habitat International, 39, 179-191.
- Barmelgy, M. M., Shalaby, A. M., Nassar, U. A. & Ali, S. M. (2014). Economic Land Use Theory and Land Value in Value Model, International Journal of Economics and Statistics, Volume 2, 2014.
- Barnes K. B., Morgan J. M., Roberge, M. C., & Lowe, S. (2012). Sprawl Development: Its Patterns, Consequences, and Measurement, Geospatial Research and Education Laboratory,

Department of Geography and Environmental Planning, Towson University. Baltimore, Maryland. Retrieved September 20, 2023, from https://tigerweb.towson.edu/morgan/files/sprawl_development.pdf

- Bayan, A., Thibault, G. (2022). Here are five policies to make transport more sustainable in cities. World Economic Forum. Retrieved December 12, 2023, from https://www.weforum.org/agenda/2022/03/five-transit-policies-cities-should-prioritize-to-become-more-sustainable
- Bera, A. K., Doğan, O., & Taşpınar, S. (2020). Specification tests for spatial panel data models. J Spat Econometrics 1, 3. DOI: https://doi.org/10.1007/s43071-020-00003-y
- Berke, P. R., & Godschalk, D. R. (2006). Urban Land Use Planning (Fifth Edition). Chicago, IL: University of Illinois Press.
- Bhatta, B. (2009). Analysis of urban growth pattern using remote sensing and GIS: a case study of Kolkata, India. International Journal of Remote Sensing, 30(18), 4733–4746.
- Bockstael, N. E., & Irwin, E.G. (2000). Economics and the Land Use-Environment Link. In Yearbook of Environmental Economics (eds). Folmer and Titienberg.
- Bonye, S. Z., Aasoglenang, T. A. & Yiridomoh, G. Y. (2021). Urbanization, agricultural land use change and livelihood adaptation strategies in peri-urban Wa, Ghana. SN Soc Sci 1, 9. DOI: https://doi.org/10.1007/s43545-020-00017-1
- Briassoulis, H. (2000). Analysis of Land Use Change: Theoretical and Modeling Approaches. The Web Book of Regional Science. Regional Science Institute, West Virginia University. https://researchrepository.wvu.edu/cgi/viewcontent.cgi?article=1000&context=rri-web-book

- Britannica, T. Editors of Encyclopedia. (2023, November 24). Metropolitan Area. Encyclopedia Britannica. Retrieved December 5, 2023, from https://www.britannica.com/topic/metropolitan-area
- Brueckner, J. K., & Fansler, D. A. (1983). The Economics of Urban Sprawl: Theory and Evidence on The Spatial Sizes of Cities. The Review of Economics and Statistics, 65, 479-482.
- Brunsdon, C., Fotheringham, A. S., & Charlton, M. E. (1996). Geographically weighted regression: a method for exploring spatial nonstationarity. Geographical analysis, 28(4), 281-298.
- Burkey, M. (2018). [BurkeyAcademy]. YouTube. Retrieved July 20, 2023, from https://www.youtube.com/@BurkeyAcademy
- Burnham, K. P., & Anderson, D. R. (2002). Model Selection and Multimodal Inference: A Practical Information-Theoretic Approach (2nd Edition). New York: Springer. Section 1.5.
- Camacho, C., & Pérez-Barahona, A. (2015). Land use dynamics and the environment, Journal of Economic Dynamics and Control, Volume 52, 2015, Pages 96-118.
- Census 1990, 2000, 2010, 2020. (n.d.). In SocialExplorer.com. Retrieved May 1, 2023, from https://www.socialexplorer.com/explore-tables
- Daniels, T. (2019). McHarg's theory and practice of regional ecological planning: retrospect and prospect. Socio Ecol Pract Res 1, 197–208. DOI: https://doi.org/10.1007/s42532-019-00024-4
- Daxpadis, C. A. (1998). Eagle of the two heads, from the past to the future of the human settlement (I. Etesam, Trans.). Mullin Jam Publication. (Original work published n.d.)

- Deng, X., Huang, J., Rozelle, S., & Uchida, E. (2008). Growth, population and industrialization, and urban land expansion of China. Journal of Urban Economics, 63, 96–115.
- Deng, X., Huang, J., Rozelle, S., & Uchida, E. (2010). Economic growth and the expansion of urban land in China. Urban Studies, 47, 813–843.
- Ebdon, D. (1991). Statistics in Geography: A Practical Approach (2nd edition). Wiley-Blackwell.
- Elhorst, P. (2012). Dynamic spatial panels: models, methods, and inferences, J Geogr Syst (2012) 14:5–28.
- Elhorst, P. (2017). Spatial Panel Data Analysis. In: Shekhar, S., Xiong, H., Zhou, X. (eds) Encyclopedia of GIS. Springer, Cham. DOI: https://doi.org/10.1007/978-3-319-17885-1_1641
- ESRI. (2023). ArcGIS Pro help. https://pro.arcgis.com/en/pro-app/3.1/help/main/welcome-to-the-arcgis-pro-app-help.htm
- Esri. (2023b). What is a z-score? What is a p-value? Retrieved on August 20, 2023, from https://pro.arcgis.com/en/pro-app/latest/tool-reference/spatial-statistics/what-is-a-z-score-what-is-a-p-value.htm
- Esri. (2023c). How Hot Spot Analysis Comparison works. Retrieved on August 20, 2023, from https://pro.arcgis.com/en/pro-app/latest/tool-reference/spatial-statistics/how-hot-spot-analysis-comparison-works.htm
- Ewing, E. (1997). Is Los Angeles-Style Sprawl Desirable? Journal of the American Planning Association, 63:1, 107-126, DOI: 10.1080/01944369708975728

- FAO. (1995). Planning for sustainable use of land resources. FAO Land and Water Bulletin 2, Retrieved June 20, 2023, from http://www.fao.org/3/v8047e/v8047e00.htm#Contents
- Florid, R. (2019). Young People's Love of Cities Isn't a Passing Fad, Retrieved November 15, 2023, from https://www.bloomberg.com/news/articles/2019-05-28/u-s-millennials-really-do-prefer-cities
- Fotheringham, S. A. Brunsdon, C., & Charlton, M. (2002). Geographically Weighted Regression: the analysis of spatially varying relationships. John Wiley & Sons.
- Getis, A., & Ord, J. K. (1992). The Analysis of Spatial Association by Use of Distance Statistics. Geographical Analysis 24, no. 3. 1992.
- Ghisleni, C. (2023). Types of Urban Blocks: Different Ways of Occupying the City. Retrieved September 1, 2023 from https://www.archdaily.com/962819/types-of-urban-blocks-different-ways-of-occupying-the-city
- Ghosh, S., & Chifos, C. (2017). The 1985 siting of a Toyota manufacturing plant in rural Kentucky, USA: The ensuing land use change and implications for planning. Landscape and Urban Planning, 167, 288-301.
- Gnedenko, E. (2020). Land Economics and Policy. An ECI Teaching Module on Social and Economic Issues. Economics in Context Initiative. Global Development Policy Center. Boston University.
- Goodchild, M. F. (1986). Spatial Autocorrelation. Catmog 47, Geo Books.
- Gordon P. & Richardson, H. W. (1997). Are Compact Cities a Desirable Planning Goal? Journal of the American Planning Association, 63:1, 95-106, DOI: 10.1080/01944369708975727

- Gore, T., & Nicholson, D. (1991). Models of the land development process: a critical review. Environmental and Planning A, 23, 705-730.
- Greene, W. H. (2008). Econometric Analysis (6th edition). Englewood Cliffs, NJ: Prentice Hall.
- Griffith, D. (1987). Spatial Autocorrelation: A Primer. Resource Publications in Geography. Association of American Geographers.
- Griffith, D. A., & Paelinck, J. H. P. (2018). The Relative Importance of Spatial and Temporal Autocorrelation. In: Morphisms for Quantitative Spatial Analysis. Advanced Studies in Theoretical and Applied Econometrics, vol 51. Springer, Cham. DOI: https://doi.org/10.1007/978-3-319-72553-6_4
- Hägerstrand, T. (1968). Innovation Diffusion as a Spatial Process. University of Chicago Press, Chicago.
- Hoover, E. M., & Giarratani. F. (1984). An Introduction to Regional Economics. University of Pittsburg. http://d-scholarship.pitt.edu/11165/1/Giarratani/contents.htm
- Hsiao, C. (2007). Panel data analysis-advantages and challenges. TEST 16, 1–22. DOI: https://doi.org/10.1007/s11749-007-0046-x
  https://www.epa.gov/report-environment/land-use
- Irwin E. G., & Geoghegan, G. (2001). Theory, data, methods: developing spatially explicit economic models of land use change, Agriculture, Ecosystems & Environment. Volume 85, Issues 1–3, Pages 7-24, ISSN 0167-8809. DOI: https://doi.org/10.1016/S0167-8809(01)00200-6

- Kaimowitz, D., & Angelsen, A. (1998). Economic Models of Tropical Deforestation: a Review. Centre for International Forestry Research. Jakarta, 139 pp. DOI: https://doi.org/10.17528/cifor/000341
- Kalnins, A. (2022). When does multicollinearity bias coefficients and cause type 1 errors? A reconciliation of Lindner, Puck, and Verbeke (2020) with Kalnins (2018). J Int Bus Stud 53, 1536–1548. DOI: https://doi.org/10.1057/s41267-022-00531-9
- Kanters, S. (2022). Fixed- and Random-Effects Models. In: Evangelou, E., Veroniki, A.A. (eds) Meta-Research. Methods in Molecular Biology, vol 2345. Humana, New York. DOI: https://doi.org/10.1007/978-1-0716-1566-9_3
- Kaplan, D. H., Holloway, S. R., & Wheeler, J. O. (2014). Urban Geography 3E. Wiley.
- Kim T. J., Claus M., Rank J. S., & Xiao Y. (2009). Technology and Cities: Processes of Technology-Land Substitution in the Twentieth Century. Journal of Urban Technology, 16:1, 63-89. DOI: 10.1080/10630730903090305
- Koomen, E. & Beurden, J. B. (2011). Land-Use Modelling in Planning Practice (GeoJournal Library Book 101) 2011th Edition. Springer.
- Lambin, E. F. (1997). Modeling and monitoring land-cover change processes in tropical regions. Prog. Phys. Geogr. 21 (3), 375–393.
- Lambin, E. F., Rounsevell, M. D. A., & Geist H. J. (2000). Are agricultural land-use models able to predict changes in land-use intensity? Agriculture, Ecosystems and Environment, 82, 321–331.
- Lambina, E. F., Turner, B. L, Geista, H. J., Agbolac, S. B., Angelsend, A., Bruce, J. W., & Xu, J. (2001). The causes of land-use and land-cover change: moving beyond the myths. Global Environmental Change, 11, 261–269.

- Lee, Y & Lee, B. (2022). What’s eating public transit in the United States? Reasons for declining transit ridership in the 2010s, Transportation Research Part A: Policy and Practice, Volume 157,2022, Pages 126-143.
- Lee, Y, Lee, B, & Shubho, MTH. (2019). Urban revival by Millennials? Intraurban net migration patterns of young adults, 1980–2010. J Regional Sci. 2019; 59: 538–566. DOI: https://doi.org/10.1111/jors.12445
- Li, M., Zang, S., Zhang, B., Li, S., & Wu, C. (2014). A Review of Remote Sensing Image Classification Techniques: The Role of Spatio-contextual Information. European Journal of Remote Sensing, 47:1, 389-411. DOI: 10.5721/EuJRS20144723
- Lo C. P., & Yang, X. (2002). Drivers of Land Use Land Cover Changes and Dynamic Modeling for the Atlanta, Georgia, Metropolitan Area. Photogrammetric Engineering & Remote Sensing, 68 (10), 1073-1082.
- Logan, J. R., Stults, B. J., & Xu, Z. (2014). Interpolating U.S. Decennial Census Tract Data from as Early as 1970 to 2010: A Longitudinal Tract Database. The Professional Geographer, 66(3), 412–420.
- Maimaitijiang, A., Ghulam, A., Onésimo Sandoval, J. S., & Maimaitiyiming, M. (2015). Drivers of land cover and land use changes in St. Louis Metropolitan Area over the past 40 years characterized by remote sensing and census population data. International Journal of Applied Earth Observation and Geoinformation, 35, 161–174.
- Malczewski, J. (2004). GIS-based land-use suitability analysis: a critical overview. Progress in Planning, 62, 3-65.

- Manson, S., Schroeder, J., Riper, D. V., Knowles, K., Kugler, T., Roberts, F., & Ruggles, S. (2023). IPUMS National Historical Geographic Information System: Version 18.0 [ASC 2009, 2010, 2020]. Minneapolis, MN: IPUMS. DOI: http://doi.org/10.18128/D050.V18.0
- Manson, S., Schroeder, J., Riper, D. V., Knowles, K., Kugler, T., Roberts, F., & Ruggles, S. (2023). IPUMS National Historical Geographic Information System: Version 18.0 [Census 2009, 2010, 2020]. Minneapolis, MN: IPUMS. DOI: http://doi.org/10.18128/D050.V18.0
- Marani, M., Katul, G. G., Pan, W. K., & Parolari, A. J. (2021). Intensity and frequency of extreme novel epidemics. Proceedings of the National Academy of Sciences, 118(35), e2105482118. DOI: https://doi.org/10.1073/pnas.2105482118
- Marsh, W. M. (2010). Landscape Planning: Environmental Applications (5th edition). Wiley.
- McCann, P. (2013). Modern Urban and Regional Economics (2nd edition). Oxford University Press.
- McHarg, I. L. (1995). Design with Nature (25th Anniversary edition). Wiley.
- Mcleod, S. (2023). Qualitative vs. Quantitative Research Methods & Data Analysis. Retrieved November 25, 2023, from https://www.simplypsychology.org/qualitative-quantitative.html
- Mertens, B., Poccard-Chapuis, R., Piketty, M. G., Lacques, A. E., & Venturieri A. (2002). Crossing spatial analyses and livestock economics to understand deforestation processes in the Brazilian Amazon: the case of Sa˜o Fe´lix do Xingu´ in South Para´. Agric Econ 27:269–294.
- Meyer, M., & Miller, E. (2000). Urban Transportation Planning (2nd edition). McGraw-Hill.

- Microsoft. (n.d.). Smart cities: The cities of the future. Retrieved December 10, 2023, from https://www.microsoft.com/en-us/industry/government/resources/smart-cities#SmartCityDefinition
- Miller, E., Kriger, D. S., & Hunt, J. D. (1999). Integrated urban models for simulation of transit and land use policies: Guidelines for implementation and use. Transportation Research Board (TCRP Report 48). https://onlinepubs.trb.org/onlinepubs/tcrp/tcrp_rpt_48.pdf
- Millo, G., & Piras, G. (2012). splm: Spatial Panel Data Models in R. Journal of Statistical Software, 47(1), 1–38. DOI: https://doi.org/10.18637/jss.v047.i01
- Mills, E. S. (1972). Studies in the Structure of the Urban Economy, Baltimore: John Hopkins University Press.
- Mínguez, R., Basile, R., & Durbán, M. (2020). An Alternative Semiparametric Model for Spatial Panel Data. Statistical Methods & Applications 29 (4): 669–708. DOI: https://doi.org/10.1007/s10260-019-00492-8
- Mitchell, A. (2005). The Esri Guide to GIS Analysis, Volume 2: Spatial Measurements and Statistics (Kindle Edition). Redland: Esri Press.
- Mitsuda, Y., & Ito, S. (2011). A review of spatial-explicit factors determining spatial distribution of land use/land-use change. Landscape Ecol Eng, 7, 117–125.
- Moser, S. C. (1996). A Partial Instructional Module on Global and Regional Land Use/Cover Change: Assessing the Data and Searching for General Relationships. Geojournal 39(3): 241-283.
- Muth, R. F. (1969). Cities and Housing, Chicago: University of Chicago Press.
- North Central Texas Council of Governments (NCTCOG). (n.d.). The Regional Data Center. Retrieved, May 15, 2023, from https://data-nctcoggis.opendata.arcgis.com/

- Nuissl, H., & Siedentop, S. (2021), Urbanisation and Land Use Change. In: Weith, T., Barkmann, T., Gaasch, N., Rogga, S., Strauß, C., Zscheischler, J. (eds) Sustainable Land Management in a European Context. Human-Environment Interactions, vol 8. Springer, Cham. DOI: https://doi.org/10.1007/978-3-030-50841-8_5
- Ord, J. K. & Getis, A. (1995). Local Spatial Autocorrelation Statistics: Distributional Issues and an Application. Geographical Analysis, 27: 286-306. DOI: https://doi.org/10.1111/j.1538-4632.1995.tb00912.x
- Overmars, K. P., de Koning, G. H. J., & Veldkamp, A. (2003). Spatial autocorrelation in multi-scale land use models. Ecological Modelling, 164 (2), 257-270.
- Pacione, M. (2005). Urban Geography: A Global Perspective (2nd edition). Routledge.
- Páez, A., Farber, S., & Wheeler, D. (2011). A simulation-based study of geographically weighted regression as a method for investigating spatially varying relationships. Environment and Planning A, 43(12), 2992-3010.
- Pászto, V., Jürgens C., Tominc, P., & Burian, J. (2019). Spationomy: Spatial Exploration of Economic Data and Methods of Interdisciplinary Analytics. Springer.
- Peña, E. N. (2012). Using census data, urban land-cover classification, and dasymetric mapping to measure urban growth of the lower Rio Grande valley, Texas [Unpublished master thesis]. University of Southern California.
- Pinto, N. (2015). Multi-scale integrated cellular modelling for the study of urban change phenomena [Unpublished doctoral dissertation]. Universitat Politècnica de Catalunya.
- Pinto, N., Antunes, P. A., & Roca, J. (2021). A Cellular Automata Model for Integrated Simulation of Land Use and Transport Interactions. ISPRS International Journal of Geo-Information 10, no. 3: 149. DOI: https://doi.org/10.3390/ijgi10030149

- Rafferty, J. P. (2023, November 7). Urban Sprawl. Encyclopedia Britannica. https://www.britannica.com/topic/urban-sprawl
- Rodrigue, J-P. (2020). The Geography of Transport Systems (5th Edition). New York: Routledge, ISBN 978-0-367-36463-2.
- Rosa, M., Knudsen, M. T., & Hermansen, J. E. (2016). A comparison of Land Use Change models: challenges and future developments. Journal of Cleaner Production, 113, 183-193.
- Singh, A. K. (2003). Modelling land use land cover changes using cellular automata in a geo-spatial environment [Unpublished master thesis]. International Institute for Geo-information Science and Earth Observation, Netherlands.
- Sleeter, B. M., Wilson, T. S., & Acevedo, W. (2012). Status and trends of land change in the Western United States—1973 to 2000: U.S. Geological Survey Professional Paper 1794–A, 324 p. Retrieved December 1, 2023, from https://pubs.usgs.gov/pp/1794/a/
- Stacy, C., Davis, C., Freemark, Y. S., Lo, L., MacDonald, G., Zheng, V., & Pendall, R. (2023). Land-use reforms and housing costs: Does allowing for increased density lead to greater affordability? Urban Studies, 60(14), 2919-2940. DOI: https://doi.org/10.1177/00420980231159500
- Steuteville, R. (2021). Ten social benefits of walkable places: We shape our cities and then they shape us. Retrieved October 20, 2023, from https://www.cnu.org/publicsquare/2021/08/12/we-shape-our-cities-and-then-they-shape-us
- Streefkerk, R. (2023). Qualitative vs. Quantitative Research | Differences, Examples & Methods, Retrieved July 25, 2023, from https://www.scribbr.com/methodology/qualitative-quantitative-research/

- Tepe, E. (2023). History, neighborhood, and proximity as factors of land-use change: A dynamic spatial regression model. Environment and Planning B: Urban Analytics and City Science, 0 (0). DOI: https://doi.org/10.1177/23998083231164397
- Tepe, E., & Guldmann, J.-M. (2020). Spatio-temporal multinomial autologistic modeling of land-use change: A parcel-level approach. Environment and Planning B: Urban Analytics and City Science, 47(3), 473-488. DOI: https://doi.org/10.1177/2399808318786511
- Tepe, E., & Guldmann, J-M. (2017). Spatial and temporal modeling of parcel-level land dynamics, computers environment and urban systems, July 2017.
- Texas Department of Transportation (TxDOT). (n.d.). TxDOT open data portal. Retrieved, May 15, 2023, from https://gis-txdot.opendata.arcgis.com/
- Thornton, P. K., & Jones, P.G. (1998). A conceptual approach to dynamic agricultural land-use modeling. Agric. Syst. 57 (4), 505–521.
- TIBCO. (n.d.). What is Data Harmonization? Retrieved August 20, 2023, from https://www.tibco.com/reference-center/what-is-data-harmonization
- Tobler, W. R. (1970). A Computer Movie Simulating Urban Growth in the Detroit Region. Economic Geography, Vol. 46, Supplement: Proceedings. International Geographical Union. Commission on Quantitative Methods, pp. 234-240. DOI: https://doi.org/10.2307/143141
- Tobler, W. R. (1999). Linear pycnophylactic reallocation comment on a paper by D. Martin, International Journal of Geographical Information Science. 13 (1): 85–90.
- Turner, B., Skole, D, Sanderson, S., Fischer, G, Fresco, L. & Leemans, R. (1995). Land-Use and Land-Cover Change Science/Research Plan. IGBP Report No.35 and HDP Report No.7. Retrieved November 1, 2023, from https://lcluc.umd.edu/sites/default/files/lcluc_documents/strategy-igbp-report35_0.pdf

- U.S. Environmental Protection Agency (EPA). (n.d.). Land Use - What are the trends in land use and their effects on human health and the environment? Retrieved August 20, 2023, from
- USGS. (2015). Toward a deeper understanding of land change, Retrieved August 20, 2023 from https://www.usgs.gov/news/featured-story/tracking-causes-and-consequences-land-change
- Vagias, W. M. (2006). Likert-type scale response anchors. Clemson International Institute for Tourism & Research Development, Department of Parks, Recreation and Tourism Management. Clemson University.
- Verburg, P. H., Schot, P., Dijst, M., & Veldkamp, A. (2004). Land use change modeling: current practice and research priorities. GeoJournal 61, 309–324 (2004).
- Wegener, M. (1995). Current and future land use models. Land Use Model Conference organized by the Texas Transportation Institute. Dallas, TX, USA.
- Weintraub, E. R. (2002). Neoclassical Economics. Liberty Fund, Inc. Retrieved March 20, 2023, from https://www.econlib.org/library/Enc1/NeoclassicalEconomics.html
- Wu, J. (2008). Land Use Changes: Economic, Social, and Environmental Impacts. The magazine of food, farm, and resource issues (CHOICES), 4th Quarter, 23(4).
- Xu, L., Liu, X., Tong, D., Liu, Z., Yin, L., & Zheng, W. (2022). Forecasting Urban Land Use Change Based on Cellular Automata and the PLUS Model Land 11, no. 5: 652, DOI: https://doi.org/10.3390/land11050652
- Yang, Y., Tang, X. & Li, Z-H. (2021). Land use suitability analysis for town development planning in Nanjing hilly areas: A case study of Tangshan new town, China. J. Mt. Sci. 18, 528–540 (2021). DOI: https://doi.org/10.1007/s11629-020-6037-z

- Zellner, A. (2001). Keep it sophisticatedly simple. In Keuzenkamp, H. & McAleer, M. Eds. Simplicity, Inference, and Modelling: Keeping it Sophisticatedly Simple. Cambridge University Press, Cambridge.
- Zheng, B., Lin, X., Yin, D., & Qi, X. (2023). Does Tobler's first law of geography apply to internet attention? A case study of the Asian elephant northern migration event. PLoS ONE 18(3): e0282474. DOI: https://doi.org/10.1371/journal.pone.0282474
- Zhou, Q., Wu, X., Zhang, X., & Song, Y. (2021). Investigating the Spatiotemporal Disparity and Influencing Factors of Urban Construction Land Utilization Efficiency: Empirical Evidence from Panel Data of China, Advances in Civil Engineering, vol. 2021, Article ID 1613978. DOI: https://doi.org/10.1155/2021/1613978
- Živković, J. (2019). Urban Form and Function. In: Leal Filho, W., Azeiteiro, U., Azul, A., Brandli, L., Özuyar, P., Wall, T. (eds) Climate Action. Encyclopedia of the UN Sustainable Development Goals. Springer, Cham. DOI: https://doi.org/10.1007/978-3-319-71063-1_78-1

End