ARTIFICIAL INTELLIGENCE AND SPATIAL MODELING TO ESTIMATE TRAFFIC VOLUME MEASURES ON LOCAL ROADWAYS

by

Mahmuda Sultana Mimi, B.S.

A thesis submitted to the Graduate Council of

Texas State University in partial fulfillment

of the requirements for the degree of

Master of Science

with a Major in Engineering

June 2025

Committee Members:

Dr. Subasish Das, Chair

Dr. Anandi Dutta

Dr. Yihong Yuan

**COPYRIGHT**

by

Mahmuda Sultana Mimi

2025

**ACKNOWLEDGMENTS**

Praise be to Allah, the Most Gracious, the Most Merciful. As I reflect on my journey over the past two years at Texas State University, I am filled with gratitude for the people and experiences that have shaped this chapter of my life. This thesis is not only the product of academic effort, but also a reflection of the support, mentorship, and kindness I have received along the way. First and foremost, I want to express my heartfelt thanks to Dr. Subasish Das, my advisor and committee chair. His unwavering support, patient guidance, and belief in my potential have had a profound impact on me. His mentorship has helped me grow into an independent thinker and researcher, and his passion for meaningful work continues to inspire me. I will always carry the lessons he taught me. I am also sincerely grateful to Dr. Anandi Dutta, whose thoughtful feedback and gentle encouragement have made a real difference in the development of this work. Her insights challenged me to think deeper and approach problems with clarity and rigor, and I truly value the role she played in shaping this thesis. To Dr. Yihong Yuan, I extend my deepest thanks for her time, generosity, and expertise. Her constructive suggestions and warm support throughout this process meant a great deal to me, and her presence on the committee brought both depth and perspective to my research. I would also like to thank everyone in the Artificial Intelligence in Transportation (AIT) Lab for creating such a welcoming and intellectually stimulating environment. The discussions, collaborations, and encouragement I received there made a lasting impact on my academic journey. Lastly, I extend heartfelt gratitude to my friends and family. Their prayers, love, and quiet strength gave me the courage to keep going, even during the most challenging moments. This accomplishment is as much theirs as it is mine.

**TABLE OF CONTENTS**

Page

[LIST OF TABLES vi](#_Toc201222850)

[LIST OF FIGURES vii](#_Toc201222851)

[LIST OF ABBREVIATIONS 1](#_Toc201222852)

[ABSTRACT 4](#_Toc201222853)

[1. INTRODUCTION 5](#_Toc201222854)

[1.1 Background 5](#_Toc201222855)

[1.2 Research Gap 8](#_Toc201222856)

[1.3 Objectives of the Study and Research Questions 9](#_Toc201222857)

[1.4 Thesis Organization 10](#_Toc201222858)

[2. LITERATURE REVIEW 11](#_Toc201222859)

[2.1 Traditional and Early Statistical Methods 11](#_Toc201222860)

[2.2 Clustering and Classification-Based Techniques 12](#_Toc201222861)

[2.3 Regression Models Using Roadway and Socioeconomic Factors 12](#_Toc201222862)

[2.4 Spatial Modeling Approaches 13](#_Toc201222863)

[2.5 Machine Learning Techniques 13](#_Toc201222864)

[2.6 Hybrid Modeling Approaches 14](#_Toc201222865)

[2.7 Applications of Spatial Random Forest Models 17](#_Toc201222866)

[2.7.1 Methodological Foundations of Spatial Random Forests 17](#_Toc201222867)

[2.7.2 Environmental Exposure and Health Impact Modeling with SRF 18](#_Toc201222868)

[2.8 Summary 19](#_Toc201222869)

[3. DATA COLLECTION AND PREPARATION 21](#_Toc201222870)

[3.1 Dataset Description 21](#_Toc201222872)

[3.2 Exploratory Data Analysis 24](#_Toc201222873)

[3.3 Overall Study Design 27](#_Toc201222874)

[4. NON-SPATIAL AI MODELING 30](#_Toc201222875)

[4.1 Motivation 30](#_Toc201222877)

[4.2 Methodology 31](#_Toc201222878)

[4.2.1 Study Design 31](#_Toc201222879)

[4.2.2 Data Preparation 33](#_Toc201222880)

[4.2.3 Variable Correlation Analysis 38](#_Toc201222881)

[4.2.4 Reducing multicollinearity in the predictors 40](#_Toc201222882)

[4.2.5 Theory/Model 40](#_Toc201222883)

[4.3 Analysis and Results 43](#_Toc201222884)

[4.3.1 Predictive Performance of the Model 43](#_Toc201222885)

[4.3.2 Importance of Factors in Estimating AADT 45](#_Toc201222886)

[4.3.3 Non-linear Impacts of AADT Predictors 46](#_Toc201222887)

[4.3.4 Two-Way Interaction Effects on AADT Prediction 48](#_Toc201222888)

[4.4 Key Findings 51](#_Toc201222889)

[4.5 Summary 52](#_Toc201222890)

[5. SPATIAL AI MODELING 54](#_Toc201222891)

[5.1 Motivation 54](#_Toc201222892)

[5.2 Methodology 55](#_Toc201222893)

[5.2.1 Study Design 55](#_Toc201222894)

[5.2.2 Data Preparation 57](#_Toc201222895)

[5.2.3 Spatial Autocorrelation Analysis of Predictors 62](#_Toc201222896)

[5.2.4 Reducing multicollinearity in the predictors 64](#_Toc201222897)

[5.2.5 Theory/Model 64](#_Toc201222898)

[5.3 Analysis and Results 69](#_Toc201222899)

[5.3.1 Predictive Performance of the Model 69](#_Toc201222900)

[5.3.2 Promising Variable Interactions 70](#_Toc201222901)

[5.3.3 Importance of Factors in Estimating AADT 73](#_Toc201222902)

[5.3.4 Non-linear Impacts of AADT Predictors 75](#_Toc201222903)

[5.3.5 Two-Way Interaction Effects on AADT Prediction 76](#_Toc201222904)

[5.4 Key Findings 80](#_Toc201222905)

[5.5 Summary 82](#_Toc201222906)

[6. CONCLUSION 83](#_Toc201222907)

[6.1 Key Findings 83](#_Toc201222908)

[6.2 Limitations 84](#_Toc201222909)

[6.3 Future Scope 84](#_Toc201222910)

[REFERENCES 85](#_Toc201222911)

# LIST OF TABLES

Page

[Table 2.1. Summary of Methods Used for AADT Estimation. 14](#_Toc201223053)

[Table 3.1. Definition of Variables. 23](#_Toc201223054)

[Table 3.2.  Descriptive Statistics of the Dataset. 26](#_Toc201223055)

[Table 4.1. Comparison of Model Performance Metrics. 45](#_Toc201223056)

[Table 5.1. Comparison of Model Performance Metrics. 70](#_Toc201223057)

[Table 5.2. Promising Variable Interactions Identified for AADT Estimation Using Spatial RF. 73](#_Toc201223058)

[Table 5.3. Variable Importance from the Spatial RF Model. 74](#_Toc201223059)

# LIST OF FIGURES

Page

[Figure 3.1. Organization of the next chapters. 29](#_Toc201223508)

[Figure 4.1. Study design flowchart. 33](#_Toc201223509)

[Figure 4.2. Distribution of AADT in original and sample datasets. 34](#_Toc201223510)

[Figure 4.3. Spatial distributions of the variables. 36](#_Toc201223511)

[Figure 4.4. Relationship between AADT and predictor variables. 38](#_Toc201223512)

[Figure 4.5. Correlation analysis of predictor variables for AADT estimation. 40](#_Toc201223513)

[Figure 4.6.Important variables from the Non-spatial RF model. 46](#_Toc201223514)

[Figure 4.7.Response curves of the top predictor variables. 48](#_Toc201223515)

[Figure 4.8.Two-way interaction effects of varying urban and transportation variables on AADT prediction. 51](#_Toc201223516)

[Figure 5.1.Study design flowchart. 56](#_Toc201223517)

[Figure 5.2.Distribution of AADT in original and sample datasets. 58](#_Toc201223518)

[Figure 5.3.Spatial distributions of the variables. 60](#_Toc201223519)

[Figure 5.4.Relationship between AADT and predictor variables. 62](#_Toc201223520)

[Figure 5.5 Moran’s I heatmap of spatial autocorrelation across predictors at varying distance thresholds. 63](#_Toc201223521)

[Figure 5.6. Top Variable interactions and their impact on AADT prediction performance in the Spatial RF model. 72](#_Toc201223522)

[Figure 5.7. Important variables from the Spatial RF model. 74](#_Toc201223523)

[Figure 5.8. Response curves of the top predictor variables. 76](#_Toc201223524)

[Figure 5.9. Impact of regional accessibility and medium-income workforce on the AADT. 77](#_Toc201223525)

[Figure 5.10. Combined influence of centrality and greenhouse gas emissions on traffic volume. 78](#_Toc201223526)

[Figure 5.11. Interaction between the low-wage workforce share and emissions in predicting AADT. 79](#_Toc201223527)

[Figure 5.12. Combined effect of socioeconomic status and centrality on traffic intensity. 80](#_Toc201223528)

# LIST OF ABBREVIATIONS

|  |  |
| --- | --- |
| **Abbreviation** | **Full Form** |
| AADT | Annual Average Daily Traffic |
| ACS | American Community Survey |
| AIC | Akaike Information Criterion |
| AASHTO | American Association of State Highway and Transportation Officials |
| AI | Artificial Intelligence |
| ANN | Artificial Neural Network |
| BP Test | Breusch–Pagan Test |
| CART | Classification and Regression Tree |
| CBSA | Core-Based Statistical Area |
| CBG | Census Block Group |
| D2A\_EPHHM | Employment and Household Entropy |
| D2R\_WRKEMP | Workers Per Job Ratio (Deviation from Regional Average) |
| D3A | Total Road Network Density |
| D5CEI | Regional Centrality Index – Auto |
| DOT | Department of Transportation |
| E\_PctLowWa | Percentage of Low-Wage Workers at Work Location |
| EPA | Environmental Protection Agency |
| EV | Electric Vehicle |
| FIPS | Federal Information Processing Standards |
| FHWA | Federal Highway Administration |
| GHG | Greenhouse Gas |
| GIS | Geographic Information System |
| GWR | Geographically Weighted Regression |
| HPMS | Highway Performance Monitoring System |
| LVR | Low-Volume Road |
| MGWR | Multiscale Geographically Weighted Regression |
| ML | Machine Learning |
| MSPE | Mean Squared Prediction Error |
| MTC | Metropolitan Transportation Commission |
| MUTCD | Manual on Uniform Traffic Control Devices |
| NatWalkInd | National Walkability Index |
| NG | Number of Groups |
| OLS | Ordinary Least Squares |
| PA | Pennsylvania |
| P\_WrkAge | Percent of Working-Age Population |
| Pct\_AO1 | Percent of One-Car Households |
| PennDOT | Pennsylvania Department of Transportation |
| QGIS | Quantum Geographic Information System |
| R\_PCTLOWWA | Percent of Low-Wage Workers at Home Location |
| RF | Random Forest |
| RSE | Residual Standard Error |
| RMSE | Root Mean Square Error |
| SLD | Smart Location Database |
| SLC\_score | Smart Location Choice Score |
| SLR | Simple Linear Regression |
| SVR | Support Vector Regression |
| TMG | Traffic Monitoring Guide |
| UPTpercap | Unlinked Public Transit Trips per Capita |
| VMT | Vehicle Miles Traveled |
| W\_P\_Highwa | Percent of High-Wage Workers at Workplace |
| W\_P\_Medwag | Percent of Medium-Wage Workers at Workplace |

# ABSTRACT

This study explores the integration of artificial intelligence (AI) and spatial modeling techniques to estimate Annual Average Daily Traffic (AADT) on local roadways, which are often data-scarce yet crucial for transportation planning and infrastructure development. Traditional traffic monitoring methods, such as permanent traffic count stations and short-term manual counts, are cost-prohibitive and fail to capture the variability and complexity of traffic flow on low-volume roads. To address this gap, the research develops and compares two modeling frameworks: a non-spatial Random Forest (RF) model and an enhanced spatial RF model. Using the comprehensive Smart Location Database (SLD) dataset from Texas that incorporates socioeconomic, land use, environmental, and transportation accessibility variables, the study applies advanced machine learning methods to capture nonlinear relationships and interaction effects. The spatial RF model, augmented with geospatial diagnostics and cross-validation, demonstrates superior predictive performance over both the non-spatial RF and conventional Geographically Weighted Regression (GWR) models. Key predictors influencing traffic volume include regional centrality, road network density, transit ridership, and employment-residential balance. The results reveal complex, context-dependent relationships, emphasizing the importance of spatial heterogeneity and urban form in shaping traffic demand. The findings contribute valuable insights to data-driven traffic estimation on local roads, with practical implications for sustainable transportation planning, emission control, and equitable infrastructure investments. The study concludes by identifying model limitations and proposing future directions for improving the integration of dynamic temporal data and enhancing the interpretability of AI-based traffic models.

# INTRODUCTION

**I**

## Background

Annual Average Daily Traffic (AADT) is a critical metric in transportation engineering, widely used for infrastructure planning, traffic operations, and safety assessments (Baffoe-Twum et al., 2022; Jessberger et al., 2016; Tsapakis et al., 2020). The Federal Highway Administration (FHWA) Traffic Monitoring Guide (TMG) defines AADT as the total volume of vehicle traffic on a roadway divided by 365 days, serving as a fundamental input for decision-making processes in transportation agencies (FHWA, 2022). AADT data is essential for traffic forecasting, roadway capacity analysis, and environmental impact assessments. State Departments of Transportation (DOTs) are required to report AADT annually to the Highway Performance Monitoring System (HPMS) for all Federal-aid roadways, ensuring comprehensive traffic data availability for national transportation planning (FHWA, 2016). The Manual on Uniform Traffic Control Devices (MUTCD) defines low-volume roads as those located outside the developed regions of cities, towns, and communities. These roads typically have an AADT of around 400 vehicles daily (Apronti et al., 2016). Meanwhile, the Pennsylvania Department of Transportation (PennDOT) provides a slightly broader classification, identifying low-volume roads as those carrying up to 500 vehicles per day (PA Act 89 of 2013) (PennDOT Traffic, 2014). The 2019 American Association of State Highway and Transportation Officials (AASHTO) guidelines on the geometric design of low-volume roads provide a detailed framework for their classification. These roads primarily facilitate transportation and economic activity in areas designated as ‘local’ within regional transit networks (AASHTO, 2019). At the federal, state, and local levels, transportation engineers recognize the significance of the AADT dataset in transportation planning and management (Sharma et al., 2000). AADT data serves as a key input for assessing roadway usage, prioritizing infrastructure improvements, and informing decisions on new road construction projects (Baffoe-Twum et al., 2022). Additionally, this dataset is integral to designing and implementing traffic control systems, which help mitigate congestion and enhance road safety. Engineers and planners use AADT estimates in pavement and roadway geometry design to analyze road section characteristics and track land development trends in affected regions. Moreover, AADT data contributes to air quality assessments, supports travel mode validation, and plays a critical role in data-driven decision-making processes (Shamo et al., 2015; Sun and Das, 2015; Zhao and Chung, 2001a). Due to the substantial financial and human resources required for traffic data collection, transportation agencies allocate significant budgets to data acquisition programs. However, obtaining complete and highly detailed traffic datasets remains infeasible due to cost constraints (Zhao and Chung, 2001a). Despite these challenges, precise estimation and modeling of traffic volumes remain essential for effective transportation planning and policy-making (Sun and Das, 2019, 2015).

Accurate estimation of AADT is essential for designing roadways, evaluating environmental impacts, and allocating resources effectively (Jin et al. 2008; Zhao and Park 2004). Traditional approaches to AADT estimation have relied on permanent traffic count stations, short-term manual counts, and expansion factor methods that adjust traffic data based on seasonal and day-of-week variations (Drusch, 1966a; Hartgen and Lemmerman, 1983; Xia et al., 1999). However, these conventional methods are often costly, time-consuming, and inadequate for capturing variations in traffic flow, especially on local roads where data collection infrastructure is sparse (Pan, 2008; Sharma and Leng, 1994). To overcome these challenges, researchers have explored a range of statistical and computational approaches to improve the accuracy and efficiency of AADT estimation. Regression-based models incorporating roadway characteristics, land use variables, and socioeconomic factors have been widely applied to predict traffic volumes in unmonitored locations (Mohamad et al., 1998; Zhao and Chung, 2001). Studies have shown that such models can provide reasonable estimates of AADT, yet the assumption of linear relationships between predictors and traffic volume often constrains their effectiveness (Xia et al., 1999). More recent efforts have utilized machine learning techniques, which allow for more complex and nonlinear relationships to be captured, leading to improved predictive performance (Castro-Neto et al., 2009; Sharma et al., 1999). Among these techniques, artificial neural networks (ANNs), random forests, and support vector regression (SVR) have shown significant promise in traffic volume prediction (Jin and Fricker, 2008; Sharma et al., 2001). For instance, studies have demonstrated the effectiveness of Gradient Boosting models and deep learning approaches in estimating AADT more accurately than traditional regression-based methods (Jiang et al., 2007; Zhao et al., 2004). Beyond machine learning, spatial modeling techniques have also been employed to enhance traffic volume estimation by accounting for spatial dependencies among road networks. Methods such as Geographically Weighted Regression (GWR) and kriging have been utilized to model the influence of nearby road segments on AADT predictions, improving estimation accuracy for locations where direct traffic counts are unavailable (Zhao and Park, 2004; Selby and Kockelman, 2011). Additionally, Bayesian methods have been introduced to integrate satellite imagery and remotely sensed data with ground-based traffic counts, further refining AADT estimates (Goel et al., 2006). However, these models often require extensive calibration despite their benefits and may not generalize well about diverse traffic conditions (Eom et al., 2006).

Despite the crucial role of accurate traffic volume estimation in transportation planning and traffic management, current AADT estimation methods for local roadways suffer from significant limitations, including data scarcity, substantial estimation errors, and insufficient management of spatial and temporal variations. Most prior studies have focused on highways and arterial roads, where data availability is relatively high, while local roads, often characterized by complex traffic patterns and limited historical data, have received less attention. The novelty of this study lies in the development of an integrated artificial intelligence and spatial modeling framework that improves AADT estimation on local roadways. By combining machine learning algorithms with geospatial variables, this research aims to develop a scalable, data-driven approach that can adapt to varying traffic conditions while reducing dependence on traditional traffic count data. The findings of this study have practical implications for policymakers, urban planners, and transportation agencies seeking to optimize traffic monitoring and infrastructure planning in areas with limited traffic count data. By addressing the limitations of existing approaches and exploring the integration of machine learning with spatial models, this research offers a more accurate, scalable, and interpretable solution for traffic volume estimation on local roadways.

## Research Gap

Despite significant advancements in AADT estimation methodologies, current approaches remain inadequate for local roadways due to challenges in data availability, spatial variability, and computational efficiency. Traditional factor-based and regression models have been widely applied, but they often rely on limited traffic count stations and assume linear relationships between predictors and traffic volume, which restricts their applicability to low-volume roads with irregular patterns. While machine learning techniques such as neural networks, support vector regression, and gradient boosting models have improved predictive accuracy, they still struggle to incorporate spatial dependencies effectively. Similarly, spatial modeling approaches like GWR and kriging account for spatial relationships but require extensive calibration and perform poorly in data-scarce environments. Most existing studies focus on highways and arterial roads, where continuous traffic monitoring is available, leaving local road networks largely underrepresented. This gap highlights the need for an integrated AI and spatial modeling framework that combines machine learning with geospatial analysis to enhance AADT estimation for local roads. Using AI-driven predictive modeling and spatial dependency structures, this research aims to develop a scalable, data-driven approach that addresses data scarcity, improves estimation accuracy, and reduces dependency on traditional traffic count data collection methods.

## Objectives of the Study and Research Questions

This research contributes to traffic volume estimation by developing a spatial AI modeling framework to improve the accuracy of AADT estimation on local roadways. While traditional methods rely on permanent traffic count stations, short-term manual counts, and factor-based estimation techniques, these approaches often fail to capture the complexity of traffic flow on low-volume and unmonitored roads due to data limitations. Most studies have focused on highways and arterial roads, with higher data availability. This study will utilize the Non-spatial and Spatial Random Forest model to improve AADT estimation for local roads. Through investigating key influencing factors such as roadway characteristics, socioeconomic variables, and geospatial dependencies, this research identifies how these factors can improve traffic volume predictions. Additionally, this study provides insights into the relative effectiveness of various predictive approaches. The proposed methodology addresses current limitations in data scarcity and estimation errors, making it valuable to data-driven traffic management, safety planning, and infrastructure development. This research will address the following research questions:

1. How can artificial intelligence and spatial modeling be integrated to improve the accuracy of AADT estimation on local roadways where traffic count data is sparse?
2. What are the key roadways, socioeconomic, and geospatial factors that influence AADT estimation?
3. What are the limitations and challenges of using AI-driven spatial modeling techniques for AADT estimation, particularly for roads with highly variable or low-volume traffic patterns?

## Thesis Organization

This thesis is organized into six chapters, each addressing a distinct component of the research process.

**Chapter One** introduces the study by outlining the background, significance, and motivation behind estimating Annual Average Daily Traffic (AADT) using machine learning models. It also defines the research objectives, scope, and the specific questions the thesis aims to address.

**Chapter Two** presents a concise literature review, discussing existing approaches to AADT estimation, including traditional regression methods and recent advancements in machine learning and spatial modeling. The chapter also identifies gaps in current research that this study aims to fill.

**Chapter Three** outlines the data and its preparation steps, along with a brief description of the upcoming chapters’ structure.

**Chapter Four** presents the development of the Non-spatial Random Forest (RF) model, including model tuning, evaluation metrics, and performance validation.

**Chapter Fiv**e focuses on the Spatial RF modeling approach. It introduces spatial enhancements such as spatial cross-validation, variable interaction analysis, and spatial permutation importance, and presents the results of the spatial RF model in comparison to GWR.

**Chapter Six** concludes the thesis by summarizing the overall contributions, outlining the limitations of the study, and suggesting potential future research directions to improve AADT modeling and its applicability in transportation planning

# LITERATURE REVIEW

**II.**

Accurate estimation of AADT plays a crucial role in transportation planning, infrastructure design, and policy formulation. Over the years, a wide array of methodologies has been proposed to improve the precision and reliability of AADT estimation, reflecting the evolving availability of data and analytical tools. This section reviews the key developments in the field, beginning with traditional factor-based and regression models, followed by advancements in clustering, spatial analysis, and machine learning techniques. The review also highlights hybrid approaches that integrate multiple methods to enhance estimation accuracy, especially in data-sparse environments. By tracing the progression of these techniques, the following subsections aim to contextualize the current study within the broader research landscape and identify existing gaps and opportunities for future work.

## Traditional and Early Statistical Methods

Over the years, various methodologies have been developed to enhance the accuracy of AADT estimation, including traditional statistical models, machine learning techniques, and spatial modeling approaches. Early approaches to AADT estimation relied heavily on short-term manual traffic counts and factor-based adjustment methods. Drusch (1966) proposed a method for grouping Continuous Count Stations (CCSs) based on monthly adjustment factors, which reduced the number of necessary seasonal counts and provided cost savings. Likewise, Hartgen and Lemmerman (1983) optimized AADT estimation by refining factoring procedures and enhancing count scheduling, which led to a 30% decrease in the number of traffic counts over five years. Ritchie (1986) introduced a statistical regression framework that adjusted short-term axle counts using group-specific regression coefficients, though its applicability was constrained by data availability. Other studies, such as, Erhunmwunsee (1991) investigated how count duration impacts estimation accuracy, concluding that eight-hour counts performed comparably to longer counts.

## Clustering and Classification-Based Techniques

Several researchers explored classification and clustering techniques to refine the assignment of traffic count data. Garber and Bayat-Mokhtari (1986) applied k-means clustering to classify road segments based on AADT and roadway characteristics, improving group homogeneity over traditional functional classification methods. Similarly, Sharma and Allipuram (1993) used hierarchical clustering to develop systematic assignment frameworks while Flaherty (1993) validated the accuracy of AADT estimates derived from cluster-based seasonal factors. These approaches demonstrated that proper classification of count sites significantly impacts AADT estimation accuracy.

## Regression Models Using Roadway and Socioeconomic Factors

Regression models have played a significant role in refining AADT estimation by incorporating roadway characteristics, socioeconomic factors, and land use variables. Mohamad et al., (1998) developed a multiple linear regression model that integrated demographic and economic variables to estimate AADT, achieving a reasonable level of accuracy. Zhao and Chung (2001) expanded on this approach by identifying key predictors such as roadway classification, lane count, and accessibility to regional employment centers. Similarly, Xia et al. (1999) applied regression techniques to estimate AADT on non-state roadways, leveraging GIS-based data aggregation for improved prediction.

## Spatial Modeling Approaches

Spatial modeling techniques have emerged as powerful tools for enhancing traffic volume estimation. Zhao and Park (2004) developed a GWR model that incorporated spatial dependencies among road segments, outperforming traditional linear regression methods. Selby and Kockelman (2011)explored kriging-based interpolation techniques, demonstrating the effectiveness of spatial autocorrelation in predicting AADT at unmonitored locations. Eom et al. (2006)further highlighted the importance of spatial correlation, showing that regression models incorporating spatial dependencies improved AADT estimation accuracy on local roads.

## Machine Learning Techniques

Recent advancements in machine learning have provided new opportunities for improving AADT estimation. Sharma et al. (1999) compared artificial neural network (ANN) models with traditional factor-based approaches, showing that ANNs could capture complex traffic patterns but require large datasets for training. Castro-Neto et al. (2009) employed Support Vector Regression (SVR) with data-dependent parameters, demonstrating superior predictive accuracy compared to regression-based models. Jin and Fricker (2008) explored k-nearest neighbor (KNN) algorithms, finding that weighted KNN models outperformed conventional factor-based methods. Das and Tsapakis (2020) proposed an interpretable machine learning framework to estimate AADT on low-volume roads in Vermont, demonstrating that ML models outperformed traditional methods, with population and work area densities emerging as key predictors. In addition, Das (2021) developed a machine learning framework using Minnesota traffic data to estimate AADT on low-volume roads, with the Cubist model outperforming other approaches and offering interpretable rule-based predictions for transportation planning.

## Hybrid Modeling Approaches

Hybrid approaches that combine machine learning with spatial modeling have also gained popularity. Wu and Zhang (2009) applied collaborative filtering techniques to group ATRs, significantly improving AADT prediction accuracy. More recently, Yang et al. (2011)proposed a novel variable selection procedure using the Smoothly Clipped Absolute Deviation (SCAD) method, enhancing regression model performance for local road AADT estimation. Zhong et al. (2012) introduced a Bayesian framework to assign seasonal factors to short-term counts, reducing estimation errors compared to traditional classification approaches. Table 2.1 represents the list of methods that have been used for AADT estimation over the years.

Table .. Summary of Methods Used for AADT Estimation.

|  |  |  |  |
| --- | --- | --- | --- |
| **Category** | **Methods Used** | **Findings** | **Sources** |
| Factor Grouping | Grouping sites based on monthly adjustment factors, k-means clustering, hierarchical clustering, fuzzy decision trees, Bayesian methods | Grouping improves the homogeneity of factor groups; factors such as functional classification and geographical location play a significant role | Drusch (1966), Garber and Bayat-Mokhtari (1986), Sharma and Allipuram (1993), Davis and Guan, (1996), Rossi et al., (2012), Gecchele et al. (2011), Lu et al. (2013), Schneider and Tsapakis (2009) |
| Regression Models | Linear regression, spatial regression, GWR, multiple linear regression, CART | Socio-economic variables, roadway characteristics, and spatial trends significantly impact AADT estimation accuracy. | Ritchie (1986), Mohamad et al., (1998), Xia et al. (1999), Zhao and Park (2004), Pan (2008), Eom et al. (2006), Lowry (2014) |
| Neural Networks | ANN, fuzzy basis function networks (FBFN) | Neural networks can perform comparably or better than traditional methods when sufficient input data and counts are available. | Sharma et al. (1999),  Sharma et al. (2001), Jin et al. (2008) |
| Clustering Techniques | Hierarchical clustering, k-means, fuzzy c-means, partitioning clustering | Clustering improves grouping accuracy, but results can vary based on clustering algorithms and weighting schemes | Garber and Bayat-Mokhtari (1986), Flaherty (1993), Rossi et al. (2012), Gecchele et al. (2011) |
| Use of Satellite/Imagery | Kriging, image-based density estimation | Satellite-based AADT estimates provide reasonable accuracy, particularly when complemented with ground-based counts | McCord et al. (2003), McCord and Goel, (2009), Jiang et al. (2007) |
| Temporal/Seasonal Factors | Monthly/day-of-week adjustment factors, seasonal adjustment factors | Seasonal factors reduce estimation errors; duration and timing of counts are critical | Granato (1998), Zhao et al. (2004), Davis and Guan, (1996), Gastaldi et al. (2013), Desai et al. (2014) |
| Innovative Techniques | ML models, including Support Vector Regression (SVR), coefficient of variation (COV), collaborative filtering | Data-driven techniques such as SVR and COV improve assignment accuracy and AADT estimation | Castro-Neto et al. (2009), Zhong et al. (2012), Wu and Zhang (2009), Das and Tsapakis (2020), Das (2021) |
| GIS-Based Tools | Geostatistics, stress centrality, connectivity importance index | GIS-based tools help spatially interpolate AADT in uncounted segments and account for spatial dependencies | Wang and Kockelman (2009)**,**  Selby and Kockelman (2011), Lowry and Dixon (2012) |
| Short-Term Count Duration | Varying count durations (24, 48, 72 hours) | Longer count durations improve AADT accuracy; counts during specific periods yield better results | Erhunmwunsee (1991), Sharma and Leng (1994), Sharma et al. (2001), Gastaldi et al. (2013) |
| Assignment Techniques | Discriminant analysis (DA), Bayesian assignment, fuzzy decision trees | Advanced assignment methods outperform traditional functional classification methods | Tsapakis et al. (2011), Lu et al. (2013), Li et al. (2006) |

## Applications of Spatial Random Forest Models

The Spatial RF has emerged as a robust modeling framework capable of addressing spatial autocorrelation and heterogeneity in geospatial data. Building upon the standard RF algorithm, spatial RF introduces spatial components either through spatial predictors (e.g., coordinates, buffer distances), model disaggregation, or explicit incorporation of spatial autocorrelation. This review synthesizes foundational and applied research employing spatial RF across diverse fields such as geosciences, environmental health, forestry, and urban modeling.

### Methodological Foundations of Spatial Random Forests

The SRF framework was first formalized by Hengl et al. (2018), who proposed the RF for Spatial Prediction (RFsp) method using buffer distances as spatial covariates to improve spatial predictions. Their comparative analysis showed that RFsp models performed comparably to kriging while offering greater flexibility. This was further extended by Talebi et al. (2022), who developed a truly spatial RF algorithm based on higher-order spatial statistics. Their model captured complex spatial patterns and was shown to outperform both standard RF and kriging in geoscientific applications. Georganos et al. (2021) also proposed the Geographical Random Forest (GRF), which disaggregates RF into localized models that better capture spatial heterogeneity, reducing residual autocorrelation and improving interpretability. These developments align with broader efforts to spatially adapt machine learning models to overcome the limitations of non-spatial learners.

A considerable challenge in spatial modeling lies in the precise adherence to observed values during the process of spatial prediction. Fouedjio (2020) introduced a novel SRF method that combines regression RF with a Bayes-linear-Gauss conditioning framework to perfectly match response values at sampled locations. This innovation is particularly relevant in geosciences, where prediction fidelity at sampled sites is critical (e.g., mineral mapping, ore body modeling). Similarly, Hengl et al. (2018) and Talebi et al. (2022) acknowledged that non-spatial RF often underestimates spatial variability, advocating for the inclusion of spatial structure either through engineered features or enhanced model structure.

### Environmental Exposure and Health Impact Modeling with SRF

SRF models have been increasingly used to explore the relationship between environmental exposures and public health outcomes. For instance, Labib (2024) utilized SRF to assess the combined effects of greenness, air pollution, and temperature on premature mortality and morbidity in Greater Manchester. By incorporating spatial autocorrelation, the model significantly improved prediction accuracy over non-spatial RF, revealing nonlinear and synergistic exposure effects. Similarly, SRF has been used in air pollution exposure modeling and land cover classification tasks, where accounting for spatial heterogeneity improved predictive power. Hengl et al. (2018) applied RFsp to predict environmental variables such as PM10 concentrations, demonstrating its applicability for spatio-temporal air quality assessment.

## Summary

Accurate estimation of AADT has long been a central concern in transportation planning, evolving through a wide range of methodologies, from traditional statistical techniques to modern machine learning and spatial models. Early methods relied heavily on short-term traffic counts and adjustment factors, which were cost-effective but often lacked precision and failed to account for spatial and temporal variability. Clustering and classification techniques improved the grouping of road segments, but their effectiveness was sensitive to the choice of clustering algorithms and required expert knowledge for interpretation. Regression models that incorporated roadway, socioeconomic, and land use variables enhanced prediction accuracy, especially for local roads, yet often assumed linearity, limiting their performance in complex real-world scenarios. In addition, they also rely on the assumption of homoscedasticity, constant variance of errors, despite the heteroskedastic behavior often present in traffic data. Moreover, multicollinearity among correlated predictors like land use and socioeconomic indicators can undermine the reliability of coefficient estimates. Traditional regression approaches typically ignore spatial autocorrelation, treating observations as independent, which is problematic in geographically structured data where neighboring areas influence each other. Additionally, these models lack flexibility in capturing interactions unless manually specified, are sensitive to outliers, and generally perform poorly when extrapolating beyond the observed data range, especially in data-sparse or unmonitored regions.

Spatial models like GWR addressed spatial autocorrelation but were prone to overfitting, particularly when dealing with high-dimensional data or small bandwidths. Machine learning techniques such as neural networks, support vector regression, and RF offered greater flexibility in modeling nonlinear relationships, but traditional implementations typically ignored spatial dependencies, potentially reducing their effectiveness in spatially structured data. Hybrid approaches sought to overcome these challenges by combining spatial and non-spatial insights, yet they often required extensive tuning and lacked generalizability. Despite these advancements, no prior study has applied Spatial RF models for AADT estimation, leaving a gap in leveraging spatially aware ensemble methods that can simultaneously handle nonlinearity, high-dimensional inputs, and spatial heterogeneity. This study addresses that gap by exploring the application of Spatial RF for estimating AADT on local roadways, offering a more holistic approach to capture both spatial structure and complex variable interactions.

# DATA COLLECTION AND PREPARATION

**III.**

This chapter presents a comprehensive overview of the dataset used for AADT estimation, including detailed definitions of all variables, their spatial and statistical characteristics, and results from an exploratory data analysis. It aims to uncover underlying patterns, relationships, and data quality issues that inform the subsequent modeling process.



## Dataset Description

The dataset utilized in this study is collected from the Systematic Longitudinal Database (SLD). It is specifically curated to support the estimation of AADT on local roadways. Understanding the distribution and characteristics of the key variables is crucial in analyzing AADT and its influencing factors. The dataset contains variables related to socioeconomic indicators, transportation accessibility, greenhouse gas emissions, land use characteristics etc. These variables are essential for developing a spatial RF model to estimate AADT.

Annual greenhouse gas emission (Annual\_GHG) is an important environmental indicator, representing the estimated emissions generated by workers in a census block group (CBG) over a 260-day work period. This variable is derived from Urban Data for Health (UD4H) and helps assess the environmental impact of commuter behavior. Another key predictor, employment and household entropy (D2A\_EPHHM), measures land use diversity within a CBG by capturing the balance between residential housing and employment types. Higher entropy values indicate mixed-use development, which has been linked to increased pedestrian activity and reduced vehicular dependence. The entropy measure is computed using the Shannon entropy equation (shown in equation 1), where the proportion of each activity type households and different employment categories contributes to the overall diversity score.

|  |  |
| --- | --- |
|  | (1) |

where Total Activities (TotAct) represents the sum of total employment and household units, and N refers to the number of distinct activity categories with nonzero values. To capture the balance between workers and employment opportunities, the study includes the Household Workers per Job Index (D2R\_WRKEMP), which quantifies the deviation of the ratio of household workers to jobs within a CBG from the regional average. This metric computation is shown in equation 2.

|  |  |
| --- | --- |
|  | (2) |

where *b* represents the regional workforce-to-employment ratio at the Core-Based Statistical Area (CBSA) level. A value closer to 1 suggests a well-balanced job-to-worker ratio, whereas deviations indicate spatial mismatches in employment accessibility. Other spatial and socioeconomic predictors include the road network density (D3A), which represents the total road length per unit area and serves as a proxy for regional connectivity and infrastructure development. The regional centrality index (D5CEI) measures the accessibility of CBG to the working-age population within the CBSA, capturing its relative importance within the regional road network. Additionally, the percentage of low-wage workers in a workplace (E\_PctLowWa) provides insight into economic disparities and their influence on commuting behavior. The study also incorporates measures of urban form and transportation accessibility, such as the National Walkability Index (NatWalkInd). This composite index evaluates the walkability of a location by integrating factors such as employment and household entropy, intersection density, and proximity to public transit. The index is derived using a weighted sum of standardized rankings, where higher values indicate greater pedestrian accessibility.

|  |  |
| --- | --- |
|  | (3) |

Where 𝑊, 𝑋, 𝑌, and 𝑍 represent ranked scores for intersection density, transit proximity, employment mix, and household-employment mix, respectively. In addition to land use and accessibility measures, workforce characteristics such as the percentage of the working-age population (P\_WrkAge), the percentage of one-car households (Pct\_AO1), and the percentage of low-wage workers residing in a CBG (R\_PCTLOWWA) are considered. These variables reflect the demographic and economic conditions that influence travel demand. The Smart Location Choice Score (SLC\_score), a composite index derived from the EPA’s Smart Location Database, evaluates the sustainability of transportation and land use decisions. Higher values indicate greater accessibility to jobs, services, and transit, thereby promoting multimodal travel behavior.

Table .. Definition of Variables.

|  |  |
| --- | --- |
| **Variables** | **Definition** |
| AADT | Annual Average Daily Traffic |
| Annual\_GHG | Annual Greenhouse Gas Emission |
| D2A\_EPHHM | Employment and household entropy |
| D2R\_WRKEMP | Household Workers per Job, as compared to the region: Deviation of Census Block Group (CBG) ratio of household workers/job from regional average ratio of household workers/job |
| D3A | Total road network density |
| D5CEI | Regional Centrality Index – Auto: CBG [D5ce] score relative to max CBSA [D5ce] score |
| E\_PctLowWa | % LowWageWk of total #workers in a CBG (work location), 2017 |
| NatWalkInd | Walkability index |
| P\_WrkAge | Percent of population that is working aged 18 to 64 years, 2018 |
| Pct\_AO1 | Percent of one-car households in CBG, 2018 |
| R\_PCTLOWWA | Percent of low-wage workers in a CBG (home location), 2017 |
| SLC\_score | A composite measure that evaluates how well a location supports sustainable transportation and land use. |
| W\_P\_Highwa | Percent of high-wage workers (workplace) |
| W\_P\_Medwag | Percent of medium wage workers (workplace) |
| UPTpercap | Unlinked passenger trips per capita for the CBSA |

## Exploratory Data Analysis

Descriptive statistics in Table 3.2 provide insights into the central tendency, dispersion, and shape of data distributions, enabling researchers to identify potential patterns and anomalies. The mean AADT in the dataset is 1344.21 vehicles per day, with a standard deviation of 1085.39, indicating significant variation in traffic volume across locations. The minimum observed AADT is two vehicles, while the maximum reaches 4000 vehicles per day. The skewness of 0.727 suggests a moderately right-skewed distribution, meaning some locations experience substantially higher traffic volumes than others. Annual greenhouse gas (GHG) emissions generated by workers in a given location have a mean value of 5796.93 metric tons, with a high standard deviation of 1685.27, indicating considerable variation in emissions. The positive skewness (1.176) and high kurtosis (3.085) suggest that certain locations exhibit exceptionally high emissions, likely due to increased commuter activity and transportation intensity.

D2A\_EPHHM, a measure of employment and household entropy, has a mean of 0.546 and a standard deviation of 0.223, indicating moderate diversity in land use. Its negative skewness (-0.361) suggests that most locations have relatively balanced employment and residential distributions. Similarly, D2R\_WRKEMP, which quantifies the deviation of the local worker-to-job ratio from the regional average, has a mean of 0.462 with minimal skewness, suggesting a relatively even employment distribution across regions. The road network density (D3A) is a crucial factor influencing traffic volume. The mean density is 16.65, with a range from 0.40 to 49.39, highlighting disparities in roadway infrastructure across different locations. This variable exhibits a slight right skew (0.256), indicating that most areas have lower-than-average road density, with a few locations containing significantly higher network density. D5CEI, which represents the regional centrality index for auto travel, has a mean value of 0.566, indicating moderate accessibility to working-age populations. The slightly negative skewness (-0.366) suggests that most locations have relatively central accessibility, with a few outliers exhibiting lower accessibility scores.

Socioeconomic factors also play a critical role in traffic volume estimation. The percentage of low-wage workers in a census block group (E\_PctLowWa) has a mean of 26.82%, with a positive skew (0.873), indicating that most locations have a relatively lower proportion of low-wage workers, with a few areas showing significantly higher values. The percentage of working-age individuals (P\_WrkAge), which reflects the proportion of the population aged 18 to 64, has a mean of 59.91% and exhibits low variability across locations. Walkability and transit accessibility indicators provide further insights into transportation patterns. The National Walkability Index (NatWalkInd) has a mean score of 9.34, ranging from 1.83 to 20, with moderate right skewness (0.544), suggesting that most locations exhibit below-average walkability, while a few regions have well-developed pedestrian infrastructure. Similarly, SLC\_score, a composite measure of sustainable transportation and land use, has a mean of 72.65, indicating moderate accessibility to jobs and services via public transit, biking, and walking.

Public transit accessibility is captured by UPTpercap, which measures unlinked passenger trips per capita. The average value of 9.87 suggests that transit usage varies significantly across locations, with some areas exhibiting higher dependence on public transportation. Wage-based workforce distribution variables, such as W\_P\_Highwa (percentage of high-wage workers in a workplace) and W\_P\_Medwag (percentage of medium-wage workers), exhibit relatively balanced distributions with minimal skewness, indicating that workforce composition remains relatively stable across locations.

Table ..  Descriptive Statistics of the Dataset.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Variables** | **Definition** | **Mean** | **Standard Deviation** | **Minimum** | **Maximum** | **Skewness** | **Kurtosis** | **Standard Error** |
| AADT | Annual Average Daily Traffic | 1344.212 | 1085.388 | 2 | 4000 | 0.727 | -0.587 | 15.350 |
| Annual\_GHG | Total estimated annual GHG generated by a worker (over 260 days) within a block group | 5796.925 | 1685.271 | 1936.721 | 19121.66 | 1.176 | 3.085 | 23.833 |
| D2A\_EPHHM | Employment and household entropy | 0.546 | 0.223 | 0 | 0.993 | -0.361 | -0.703 | 0.003 |
| D2R\_WRKEMP | Household Workers per Job, as compared to the region: Deviation of CBG ratio of household workers/job from regional average ratio of household workers/job | 0.462 | 0.295 | 0 | 1 | 0.180 | -1.212 | 0.004 |
| D3A | Total road network density | 16.647 | 8.894 | 0.405 | 49.388 | 0.256 | -0.523 | 0.126 |
| D5CEI | Regional Centrality Index – Auto: CBG [D5ce] score relative to max CBSA [D5ce] score | 0.566 | 0.262 | 0 | 1 | -0.366 | -0.918 | 0.004 |
| E\_PctLowWa | % LowWageWk of total #workers in a CBG (work location), 2017 | 0.268 | 0.146 | 0 | 1 | 0.873 | 1.180 | 0.002 |
| NatWalkInd | Walkability index | 9.343 | 3.725 | 1.833 | 20 | 0.544 | -0.421 | 0.053 |
| P\_WrkAge | Percent of population that is working aged 18 to 64 years, 2018 | 0.599 | 0.094 | 0 | 0.970 | 0.156 | 3.491 | 0.001 |
| Pct\_AO1 | Percent of one-car households in CBG, 2018 | 0.328 | 0.152 | 0 | 0.901 | 0.340 | -0.219 | 0.002 |
| R\_PCTLOWWA | Percent of low wage workers in a CBG (home location), 2017 | 0.228 | 0.057 | 0.078 | 0.473 | 0.764 | 0.676 | 0.001 |
| SLC\_score | Composite measure that evaluates how well a location supports sustainable transportation and land use. | 72.652 | 16.308 | 0 | 100 | -1.416 | 2.513 | 0.231 |
| W\_P\_Highwa | Percent of high wage workers (workplace) | 0.364 | 0.181 | 0 | 1 | 0.446 | 0.013 | 0.003 |
| W\_P\_Medwag | Percent of medium wage workers (workplace) | 0.367 | 0.114 | 0 | 1 | 0.352 | 2.222 | 0.002 |
| UPTpercap | Unlinked passenger trips per capita for the CBSA | 9.873 | 5.844 | 0 | 18 | -0.625 | -0.855 | 0.083 |

.

## Overall Study Design

Figure 3.1 presents the methodological workflows for Chapter 4 (Non-Spatial AI Modeling) and Chapter 5 (Spatial AI Modeling). Both chapters begin with the same dataset, filtered for local roads in Texas, and follow structured data preparation steps, including correlation and multicollinearity analysis. In Chapter 4, a Non-Spatial RF model is used and benchmarked against an Ordinary Least Squares (OLS) model. In Chapter 5, a Spatial RF model is implemented with interaction analysis and evaluated against a GWR model. The outputs for both approaches include variable importance plots, partial dependence curves, and two-way interaction visualizations.

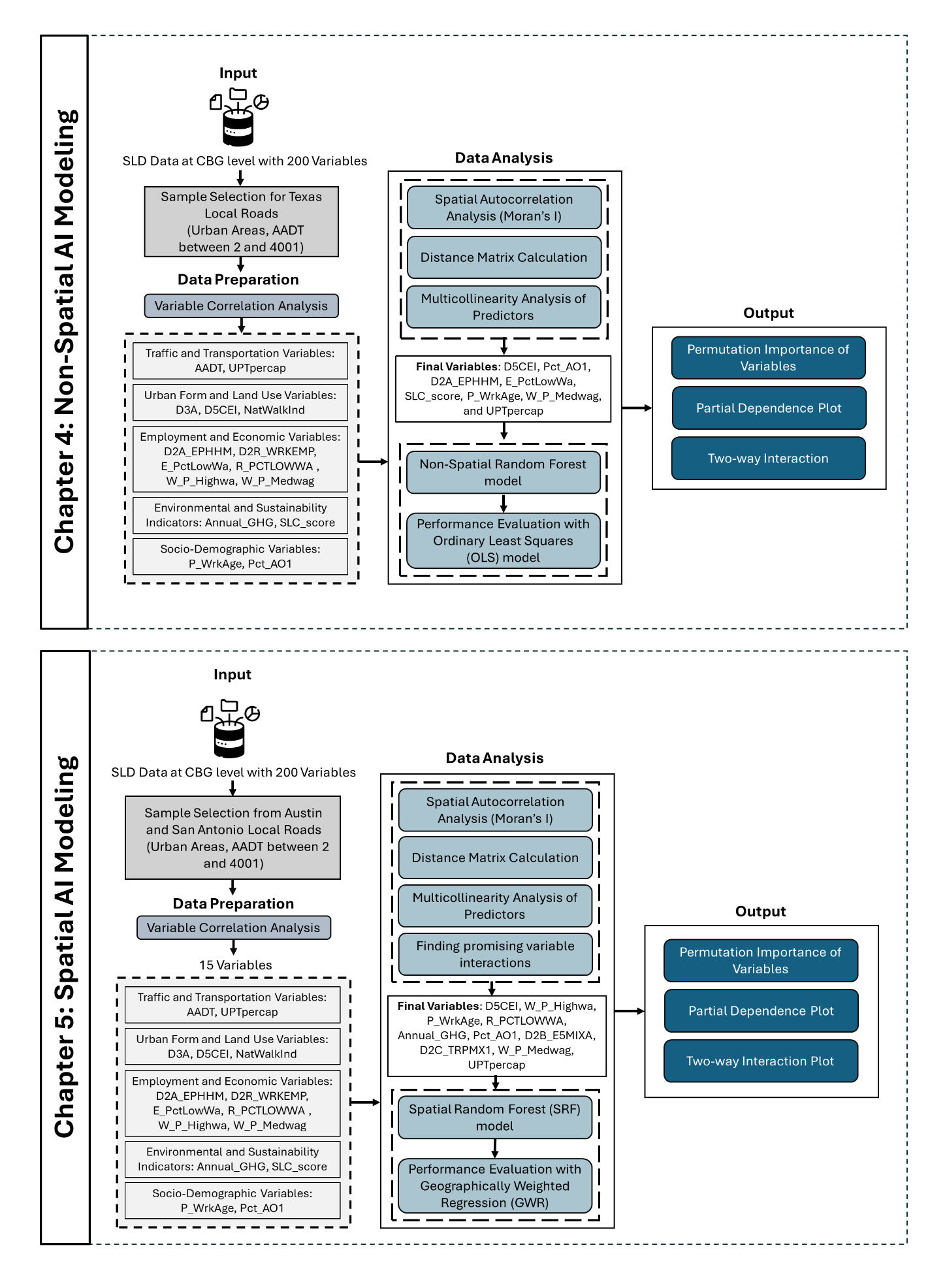


Figure .. Organization of the next chapters.

# NON-SPATIAL AI MODELING

**IV.**

This chapter presents a comprehensive modeling framework for estimating AADT on local roadways using non-spatial machine learning techniques. It begins by outlining the data preparation process, including variable selection, multicollinearity checks, and spatial diagnostics. The chapter then introduces the non-spatial RF model and evaluates its predictive performance using standard metrics such as R² and RMSE. To contextualize the model’s performance, comparisons are made with baseline OLS models, both linear and nonlinear, to highlight the limitations of traditional regression approaches. The chapter concludes with a discussion on model interpretability, residual analysis, and implications for planning applications.



## Motivation

Understanding traffic volumes on local roadways is a cornerstone of sustainable transportation planning, yet it remains one of the most underexplored and data-deficient areas in the field. Unlike highways and major arterials, local roads often lack the infrastructure for continuous traffic monitoring, making it difficult to gather the detailed data necessary for informed decision-making around roadway design, traffic safety, environmental assessment, and infrastructure investment. This gap is especially problematic given that local roads represent the majority of the roadway network and play a critical role in supporting everyday mobility, access to public transit, and neighborhood connectivity. Traditional estimation methods, reliant on short-term counts and generalized adjustment factors, are not only cost-prohibitive but also unable to capture the detailed and dynamic traffic patterns unique to local contexts. In response to these challenges, this study applied a Non-spatial RF model, an advanced machine learning approach capable of handling nonlinearities and complex variable interactions to predict AADT using diverse geospatial, socioeconomic, and transportation-related data. By focusing specifically on local roadways, this research fills a significant methodological and practical void, offering a scalable, data-driven solution that can improve traffic estimation accuracy where it is needed most. The findings have the potential to reshape how transportation agencies monitor and plan for local mobility, ultimately informing the development of more equitable, efficient, and sustainable urban transportation systems.

## Methodology

This section outlines the methodological framework employed to estimate AADT on local roadways using a Non-spatial RF model. It details the data preprocessing steps, variable selection process, and sampling strategy used to construct the modeling dataset.

### Study Design

The analytical framework used in this study (as shown in Figure 4.1) follows a structured approach, beginning with data acquisition from the SLD. The data preparation phase involves variable correlation analysis to identify key predictors and ensure that the selected features contribute meaningfully to the subsequent modeling process. This study includes a comprehensive range of factors that influence the AADT estimation. The variables considered in this study span multiple domains, including traffic and transportation (AADT, UPTpercap), urban form and land use (D3A, D5CEI, NatWalkInd), employment and economic indicators (D2A\_EPHHM, D2R\_WRKEMP, E\_PctLowWa, R\_PCTLOWWA, W\_P\_Highwa, W\_P\_Medwag), environmental and sustainability measures (Annual\_GHG, SLC\_score), and socio-demographic attributes (P\_WrkAge, Pct\_AO1).

Following data preparation, the analysis proceeds to spatial analysis and modeling. The first step involves spatial autocorrelation analysis using Moran’s I, which assesses the degree of spatial dependence among the variables in the dataset. This is followed by the calculation of a distance matrix, which quantifies spatial relationships between observations. Additionally, a multicollinearity analysis of predictors is conducted to identify redundant variables that could affect model estimates. Once the exploratory analysis is complete, key candidate variables, such as transit ridership, land-use centrality, greenhouse gas emissions, and workforce composition, are selected for further analysis. Two modeling approaches are employed. The first is the non-spatial RF model, which is conducted without explicitly incorporating spatial dependencies. Later, the OLS method is employed, which is a baseline for evaluating the predictive performance of the non-spatial RF model.

Later, the results are compared between the two models. The outputs of this modeling framework include permutation importance and local variable importance, which identify the most influential predictors in non-spatial contexts. Additionally, response curves and surfaces provide insight into how individual predictors influence the dependent variable across different spatial scales. Finally, partial dependence plots illustrate the marginal effect of specific variables on the outcome, facilitating a deeper understanding of variable interactions and their implications. These outputs collectively enhance the interpretability of the model and support data-driven decision-making.

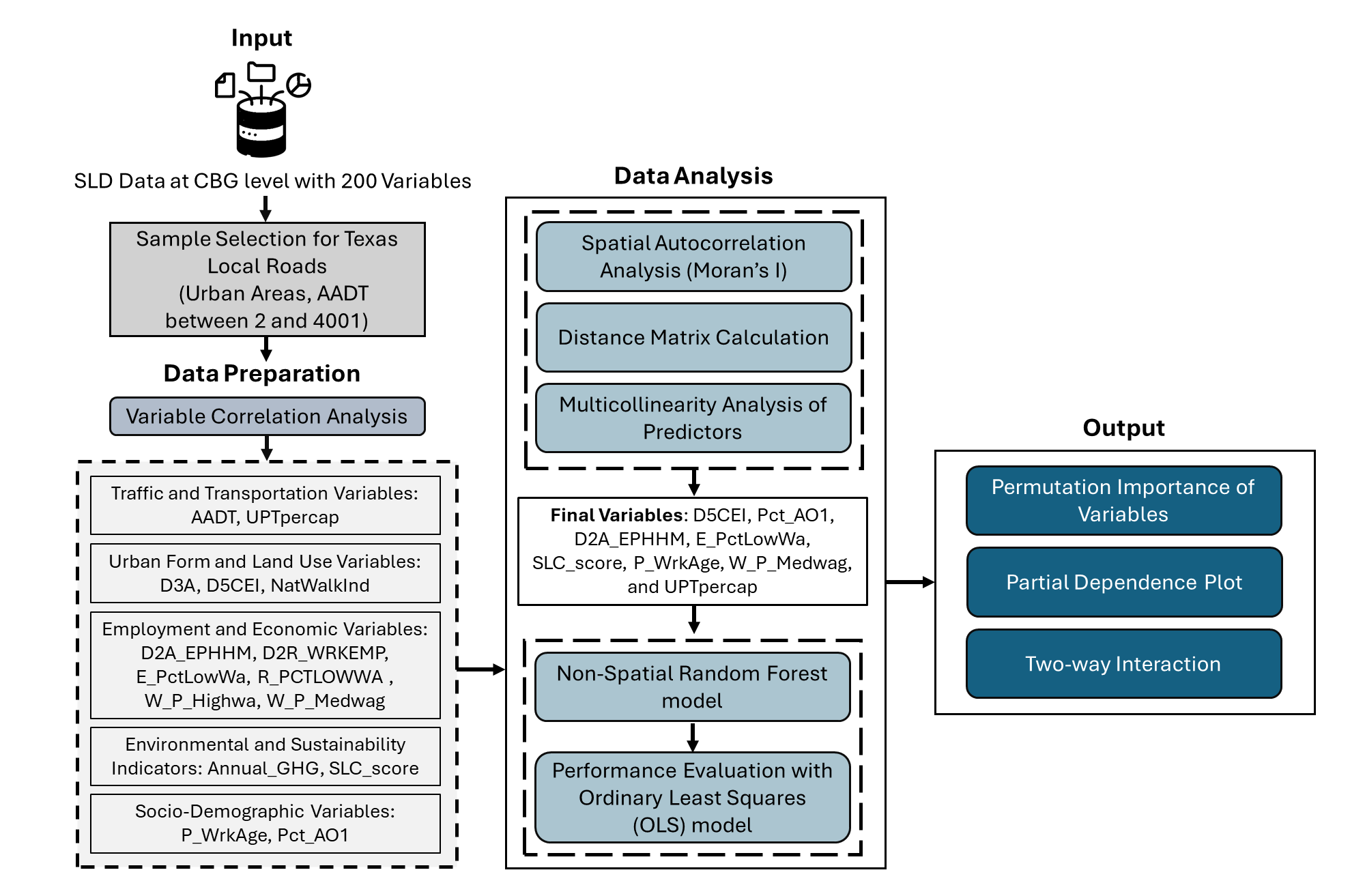


Figure .. Study design flowchart.

### Data Preparation

The dataset utilized in this study was collected from SLD. It is specifically curated to support the estimation of AADT on local roadways. Understanding the distribution and characteristics of the key variables is crucial in analyzing AADT and its influencing factors. The dataset contains variables related to socioeconomic indicators, transportation accessibility, greenhouse gas emissions, land use characteristics etc. These variables are essential for developing a non-spatial RF model to estimate AADT. The dataset comprises 48,470 data points collected between 2019 and 2023, of which 22,984 correspond to local roads defined by AADT values below 4,001. For the estimation of AADT on local roadways, a random sample of 5,000 data points was selected for all of Texas, covering an AADT range of 2 to 4,000. The sample distribution graph shown in Figure 4.2 illustrates the density estimation of AADT for two datasets, Original and Sample, represented by overlapping density plots. The distribution reveals a right-skewed pattern, indicating that lower AADT values are more frequent, while higher values occur less often. The close alignment between the two distributions suggests that the sample dataset retains the key characteristics of the original dataset, with slight variations in density across different AADT ranges. This comparison validates the representativeness of the sample in preserving the overall AADT distribution and ensures that the analysis reflects the broader trends in local road traffic volume.

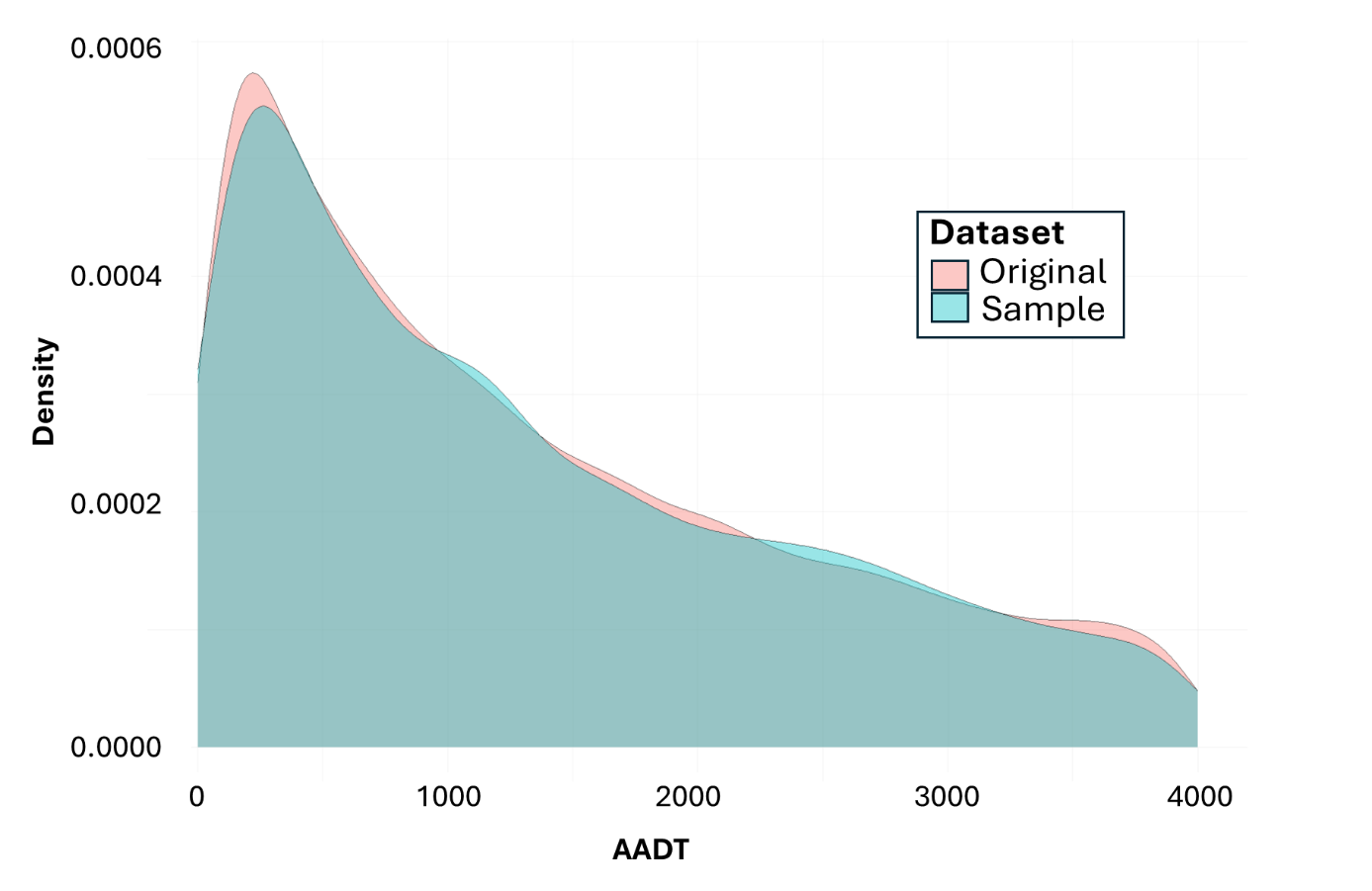


Figure .. Distribution of AADT in original and sample datasets.

Figure 4.3represents the spatial distribution of the variables, with yellow representing the highest value and purple representing the lowest value.

|  |  |  |
| --- | --- | --- |
| **A map of texas with purple dots  AI-generated content may be incorrect.** | **A map of texas with purple dots  AI-generated content may be incorrect.** | **A map of texas with red and yellow dots  AI-generated content may be incorrect.** |
| **A map of texas with purple dots  AI-generated content may be incorrect.** | **A map of texas with red and purple dots  AI-generated content may be incorrect.** | **A map of texas with purple dots  AI-generated content may be incorrect.** |
| **A map of texas with red and yellow dots  AI-generated content may be incorrect.** | **A map of texas with purple dots  AI-generated content may be incorrect.** | **A map of texas with different colored dots  AI-generated content may be incorrect.** |
| **A map of texas with red dots  AI-generated content may be incorrect.** | **A map of texas with purple dots  AI-generated content may be incorrect.** | **A map of texas with different colored spots  AI-generated content may be incorrect.** |
| **A map of texas with purple dots  AI-generated content may be incorrect.** | **A map of texas with purple dots  AI-generated content may be incorrect.** | **A map of the state of texas  AI-generated content may be incorrect.** |

Figure .. Spatial distributions of the variables.

The scatterplots in Figure 4.4 illustrate the relationship between AADT and various predictor variables derived from the dataset. Each panel represents a predictor variable plotted against AADT, where the density of points is indicated by color intensity. These visualizations provide insights into how different factors influence traffic volume, identifying potential trends, patterns, or outliers in the data. D5CEI exhibits a positive association with AADT, suggesting that areas with higher regional accessibility experience greater traffic volumes. This is consistent with the expectation that more centrally located areas attract higher levels of vehicular movement due to their connectivity within the broader transportation network. A similar trend is observed for Pct\_AO1, where regions with a higher proportion of single-vehicle households tend to have elevated traffic volumes. The relationships between D2A\_EPHHM and D3A with AADT reveal more intriguing patterns. Higher values of employment and household entropy, which reflect a more balanced mix of residential and employment land uses, are associated with moderate increases in traffic volumes. This trend aligns with the notion that mixed-use developments can generate localized travel demand. The D3A variable, representing the density of the road network, also exhibits a weak positive correlation with AADT, reinforcing the idea that higher road density facilitates greater vehicle mobility.

Conversely, variables related to socioeconomic characteristics, such as R\_PCTLOWWA and E\_PctLowWa, demonstrate weaker or slightly negative associations with AADT. These findings suggest that areas with a higher proportion of low-wage workers experience lower traffic volumes, potentially due to greater reliance on non-motorized travel modes or public transit. Similarly, the SLC\_score measures how well a location supports sustainable transportation. Higher SLC scores may correlate with lower vehicle volumes in some locations, reflecting improved access to alternative modes of transportation. The plot for Annual\_GHG reveals a strong positive relationship with AADT, indicating that regions with higher traffic volumes generate greater emissions. This finding aligns with prior research linking vehicular travel demand with environmental impacts. In contrast, D2R\_WRKEMP shows a weak relationship with AADT, implying that variations in employment-to-residence balance may not significantly influence overall traffic levels.

Workplace wage characteristics also exhibit varied relationships with AADT. P\_WrkAge, W\_P\_Medwag, and W\_P\_Highwa display non-linear associations, suggesting that the influence of workplace income levels on traffic volume is context dependent. Higher-wage employment centers may not always correspond with increased traffic volumes, possibly due to the availability of alternative commuting options such as transit. A noteworthy observation arises in the scatter plot for UPTpercap, which does not correlate strongly with AADT. This finding suggests that transit ridership levels alone may not directly influence traffic volumes, likely due to variations in transit accessibility, service quality, and land-use patterns. Likewise, NatWalkInd does not display a strong negative correlation with AADT, indicating that while walkability improvements may contribute to localized reductions in vehicle use, they do not necessarily translate into broader declines in traffic volumes.

**A screenshot of a graph

AI-generated content may be incorrect.**

Figure .. Relationship between AADT and predictor variables.

### Variable Correlation Analysis

A correlation analysis was conducted among the variables. Highly correlated variables are removed, and the rest are presented in Figure 4.5. The correlation matrix visually represents the strength and direction of these associations, where blue tones indicate positive correlations, and red tones signify negative correlations. The size of the circles in the correlation plot reflects the magnitude of the relationship, with larger circles denoting stronger correlations. This analysis is critical for assessing multicollinearity, identifying key contributing factors, and selecting appropriate variables for modeling AADT.

D5CEI exhibits moderate positive correlations from the correlation matrix with D3A and NatWalkInd. This suggests that areas with high road network density and better walkability tend to have higher regional centrality for automobile access. D3A and NatWalkInd are also moderately correlated, indicating that dense road networks often coincide with higher walkability scores. SLC\_score also shows a moderate positive correlation with D3A and walkability measures, reinforcing its role in evaluating accessibility and urban development patterns. A notable positive correlation is observed between D2R\_WRKEMP and Pct\_AO1, indicating that regions with higher employment density tend to have a more significant proportion of single-car households. Conversely, E\_PctLowWa has a negative correlation with Annual\_GHG, suggesting that regions with a higher proportion of low-wage workers tend to have lower overall emissions.

Some variables exhibit moderate to strong negative correlations, for example, UPTpercap, a measure of transit usage, is negatively correlated with W\_P\_Highwa, suggesting that higher-income employment centers are associated with lower transit ridership per capita. This may reflect automobile dependence in high-income employment zones, where access to personal vehicles is more common. Similarly, Annual\_GHG has a negative correlation with SLC\_score, indicating that locations with better sustainable transportation infrastructure tend to have lower emissions, reinforcing the environmental benefits of transit-oriented development.

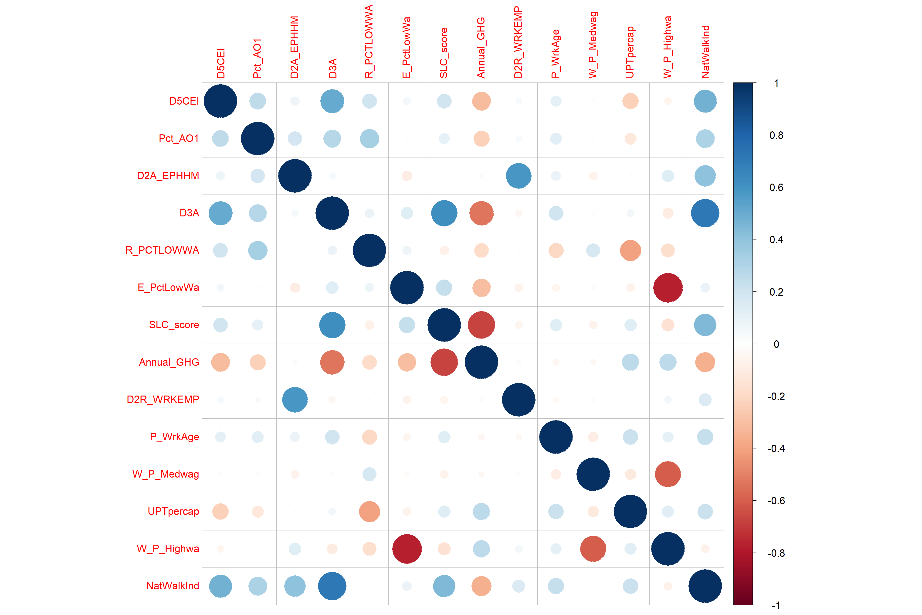
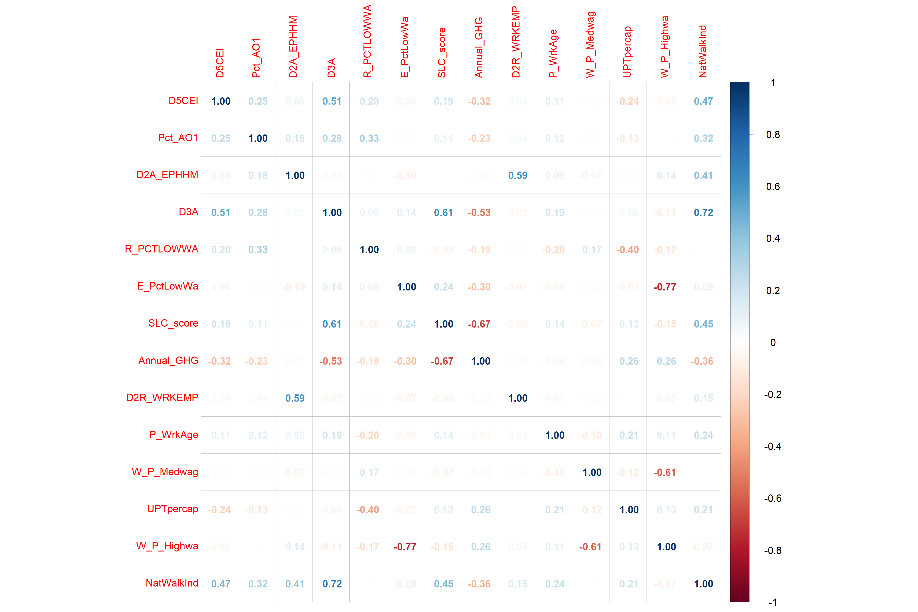
 

Figure .. Correlation analysis of predictor variables for AADT estimation.

### Reducing multicollinearity in the predictors

A multicollinearity analysis was conducted to ensure the stability and interpretability of the model by identifying and removing highly correlated predictor variables. This process involved two key steps using the spatialRF package: first, a pairwise correlation check with a threshold of 0.8 was applied to eliminate strongly correlated variables, giving preference to those with higher importance scores; second, a variance inflation factor (VIF) analysis was performed with a conservative threshold of 1.2 to remove further variables exhibiting multicollinearity. Out of the initial 14 predictors, 8 variables were retained after this filtering process. The following variables, D5CEI, Pct\_AO1, D2A\_EPHHM, E\_PctLowWa, SLC\_score, P\_WrkAge, W\_P\_Medwag, and UPTpercap, were selected as the final set for subsequent model runs and analysis. These selected predictors exhibited minimal redundancy and low interdependence, ensuring that the model could effectively capture the independent contributions of each variable without being distorted by overlapping information. Later, these variables were used for the non-spatial RF model.

### Theory/Model

***Non-Spatial RF Model***

RF is an ensemble machine learning algorithm introduced by (Breiman, 2001a), designed to enhance predictive accuracy and control overfitting by combining multiple decision trees. It operates by constructing each tree from a bootstrap sample of the data and splitting nodes using a random subset of predictors, thus ensuring model diversity. In regression problems, the final prediction is obtained by averaging the output of individual trees.

Given a dataset , where are the predictors and are the target values (in this case, AADT), the prediction for a new observation is obtained as the average of predictions from all decision trees in the ensemble:

Where:

* *T* is the total number of trees,
* is the prediction from the *t*-th decision tree.

Each tree uses a bootstrap sample, selecting a random feature subset at each split to find the best division. This randomness enhances the variety within trees, reducing variance and improving generalization (Breiman, 2001a). At each split in a tree, a random subset of predictors (out of total) is selected. The optimal split minimizes Mean Squared Error (MSE):

The RF minimizes the out-of-bag (OOB) error as an internal estimate of generalization performance. Additionally, variable importance is quantified via: Mean Decrease in Impurity (MDI) and **Permutation-based importance**, calculated by randomly permuting a variable and measuring the increase in prediction error.

In this study, the RF model is implemented using the *rf()* function from the *spatialRF* R package (Benito, 2023), which wraps the high-performance *Ranger* package (Wright and Ziegler, 2017). The model was trained using a selected set of predictors to estimate AADT, a continuous target variable representing transportation demand. The non-spatial RF model does not explicitly incorporate spatial autocorrelation in the data, treating each observation as independent in the modeling process. However, the *spatialRF::rf()* function accepts a spatial distance matrix and user-defined distance thresholds to evaluate the spatial autocorrelation of residuals, typically using Moran’s I statistic (Moran, 1950). These diagnostics help assess whether spatial structure remains unexplained by the model, but they are not used for prediction or model training. Spatial diagnostics are conducted post hoc using a user-supplied **distance matrix**  and distance thresholds to assess **spatial autocorrelation** in model residuals, typically via the following equation of **Moran’s I statistic**:

Where, and are residuals at locations *i* and *j*; is the spatial weight (often 1 if within distance threshold, 0 otherwise); and is the mean residual. Non-spatial RF is effective in capturing nonlinear relationships and interactions among predictors. Unlike traditional parametric methods such as OLS or GWR, RF does not assume linearity, normality, or homoscedasticity (Cutler et al., 2007). This makes RF especially well-suited for modeling complex, nonlinear, and high-dimensional relationships commonly found in spatial and environmental data (Hengl et al., 2018a). Moreover, RF is robust to multicollinearity and outliers and is capable of modeling interactions implicitly without requiring manual specification. However, it may overlook spatial dependencies inherent in geographic data, which can lead to biased predictions in spatially correlated environments (Hengl et al., 2018a; Talebi et al., 2022a). Its performance and residual spatial diagnostics provide a useful reference point for comparing with spatially explicit models such as spatial RF or GWR.

***Ordinary Least Square (OLS) Model***

The OLS model is a foundational linear regression technique used to model the relationship between a dependent variable and independent variables. OLS estimates the parameters that minimize the sum of squared residuals between observed and predicted values. The OLS regression model is expressed as:

Where, is vector of observed values, is matrix of predictors (including a column of 1s for the intercept), is vector of coefficients, ε is the vector of error terms. The estimated coefficients are calculated by minimizing the residual sum of squares (RSS):

The OLS model is based on the assumption that residuals are independent and identically distributed (i.i.d.), normally distributed with a mean of zero, and exhibit constant variance (homoscedasticity). It also assumes no multicollinearity among the predictors. OLS is easy to interpret and computationally efficient, but can be limited in predictive performance when the relationship between predictors is nonlinear, or when multicollinearity and spatial autocorrelation are present (Chatterjee and Hadi, 2015).

## Analysis and Results

This section summarizes the model results, including the important variables for AADT estimation, response curves, and non-linear impacts of the predictor variables.

### Predictive Performance of the Model

The Non-spatial RF model was trained on a dataset comprising 5000 observations and 14 predictor variables to estimate AADT. The model was fitted using the ranger package, employing 500 trees and a minimum node size of 5 to optimize predictive accuracy. The model achieved an R² of 0.7207, indicating that the predictor variables explain approximately 72.07% of the variance in AADT. Additionally, the pseudo R² of 0.8489 suggests a strong overall predictive capability. The root mean squared error (RMSE), a measure of model error, was 627.99, indicating that the model demonstrates reasonable predictive accuracy. The residuals representing the differences between observed and predicted AADT values were analyzed to assess model fit. The mean residual is -4.53, indicating that the model does not exhibit a strong bias in its predictions.

To evaluate the predictive performance of traffic volume estimation, the non-spatial RF model is compared with two OLS models: a basic OLS model and an OLS model with nonlinear terms. The basic OLS model, using eight predictors, achieved a low R² of 0.032, indicating that only about 3.2% of the variance in AADT was explained. Introducing polynomial (squared) terms in the improved OLS model slightly raised the R² to 0.039, highlighting some nonlinear effects such as diminishing returns in the predictor variables. However, the explanatory power remained weak overall. Additionally, both OLS models showed significant heteroskedasticity (Breusch-Pagan p-values < 0.01), violating OLS assumptions and casting doubt on the validity of coefficient inferences. In contrast, the non-spatial RF model substantially outperformed both OLS models in both predictive accuracy and residual behavior. With an R² of 0.721, the RF model explained nearly 72% of the variance in AADT and demonstrated strong predictive capacity. The RMSE dropped significantly to 627.99. Residual diagnostics further validate the model’s robustness: although the residuals were not perfectly normally distributed (Shapiro-Wilk W = 0.974, p < 0.001), the spatial autocorrelation analysis revealed no significant clustering beyond 20,000 meters. Beyond this distance, Moran's I values became statistically insignificant, suggesting that the model successfully accounted for local spatial patterns through its predictors. This contrasts with the OLS models, where residual spatial dependence is more likely to persist. The RF’s ability to capture complex nonlinearities and variable interactions, without assuming constant error variance or linearity, offers a clear methodological advantage. These results support the use of ensemble-based machine learning approaches, such as RF, over traditional regression methods for modeling traffic volume, especially when spatial and nonlinear patterns are present but difficult to specify manually.

Table .. Comparison of Model Performance Metrics.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Model** | **R²** | |  | | --- | |  |   **RMSE** | **Heteroskedasticity (BP test p-value)** |
| OLS (Linear) | 0.0319 | 1069 | 0.0057 |
| OLS (with Nonlinear Terms) | 0.0393 | 1065 | 5.91e-05 |
| Non-Spatial RF (ranger) | 0.7207 | 627.99 | Not Applicable |

### Importance of Factors in Estimating AADT

The importance of predictor variables was assessed using the mean decrease in impurity (Gini importance). The top contributing variables to AADT prediction include D5CEI, UPTpercap, and NatWalkInd, highlighting the significant role of urban accessibility, transit ridership, and walkability in influencing traffic volume. The high importance of D5CEI is consistent with the expectation that central, well-connected locations experience higher vehicular movement. Similarly, the influence of UPTpercap suggests that areas with greater transit usage tend to have higher traffic volumes, potentially reflecting multimodal transportation environments.

Conversely, variables such as E\_PctLowWa (Percentage of Low-Wage Workers), Percent of medium wage workers (W\_P\_Medwag), and Percentage of Working-Age Population(P\_WrkAge) exhibit lower importance, suggesting that workforce-related and economic characteristics have a relatively weaker direct influence on AADT compared to factors like land use, accessibility, and transit ridership. The relative importance of all variables is summarized in Figure 4.6.

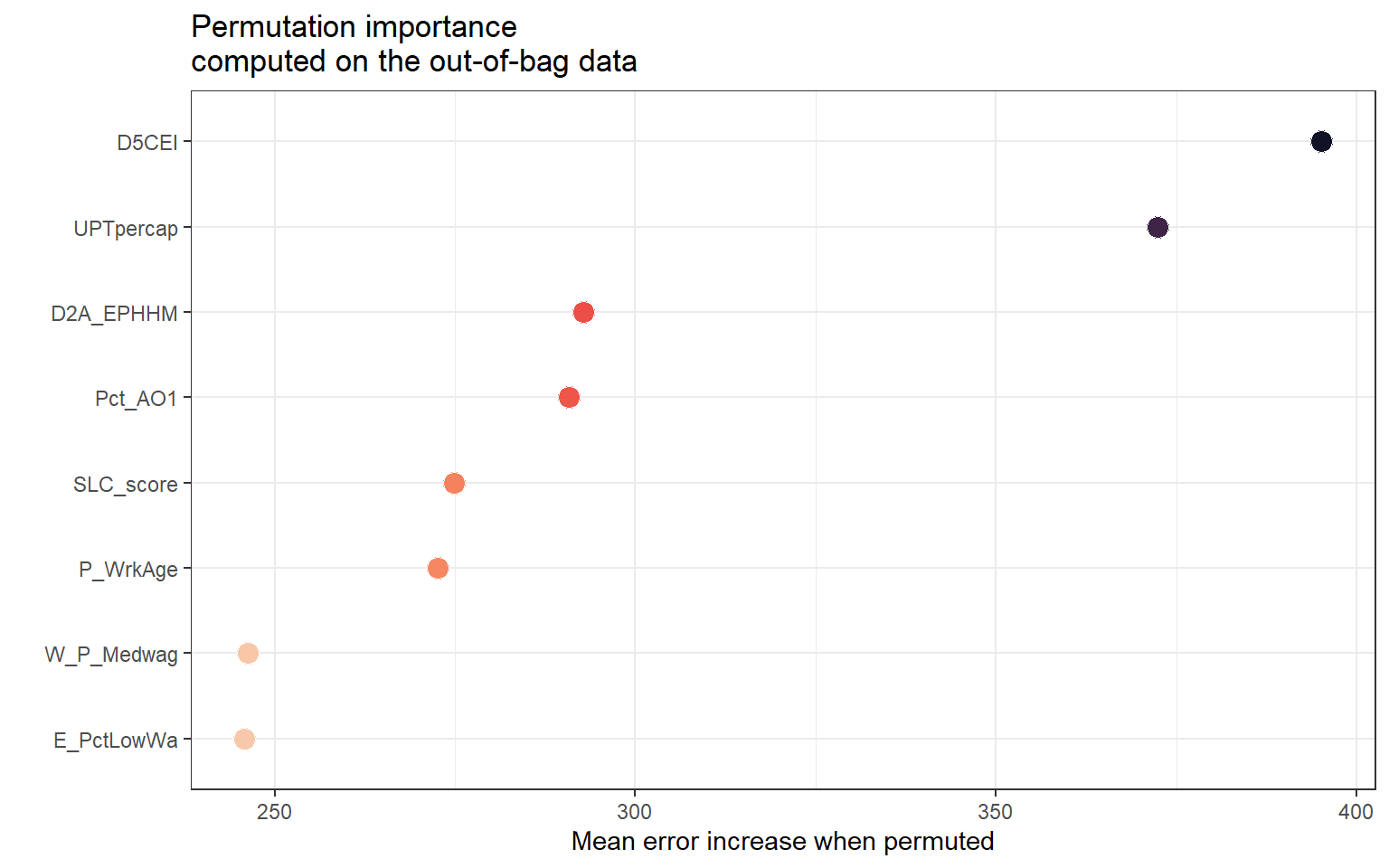


Figure ..Important variables from the Non-spatial RF model.

### Non-linear Impacts of AADT Predictors

The response curves highlight the complex, nonlinear interactions between dependent and predictor variables. The inclusion of quantile-based trends enables a more meaningful interpretation, accounting for variations across different levels of the predictor distributions. The response curves in Figure 4.7 illustrate the nonlinear relationships between AADT and key predictor variables in the non-spatial RF model. These partial dependence plots (PDPs) capture how AADT varies with each predictor while holding all other variables at their median values. The results indicate complex and non-monotonic response patterns, emphasizing the advantages of machine learning approaches in capturing complex relationships that traditional linear models may overlook. D5CEI shows a significant increase in AADT up to the mid-range, after which the effect stabilizes, indicating that more central locations tend to have higher traffic volumes, likely due to better accessibility and more travel demand. These findings align with Cui et al., (2022) and Wang et al., (2020), suggesting that areas with higher accessibility experience greater vehicular movement due to improved connectivity. UPTpercap indicates a rise in AADT at higher levels, implying that regions with significant transit usage might also experience increased vehicle traffic. This could reflect densely populated, multimodal urban centers where driving and transit operate together (Verbavatz and Barthelemy, 2019). D2A\_EPHHM exhibits a U-shaped trend, demonstrating reduced AADT at mid-range values and elevated traffic at both extremes. This pattern may indicate varying commuting behaviors influenced by land-use balance. Highly specialized or highly mixed-use areas experience higher traffic. These might be dense urban zones with significant commercial-residential overlap, attracting both local and through traffic, possibly even congestion (Ewing et al., 2011). Pct\_AO1 indicates a steady rise in AADT, suggesting that higher percentages of one-car households are associated with increased AADT, indicating greater vehicle dependency and more frequent vehicle use in areas with limited travel alternatives.

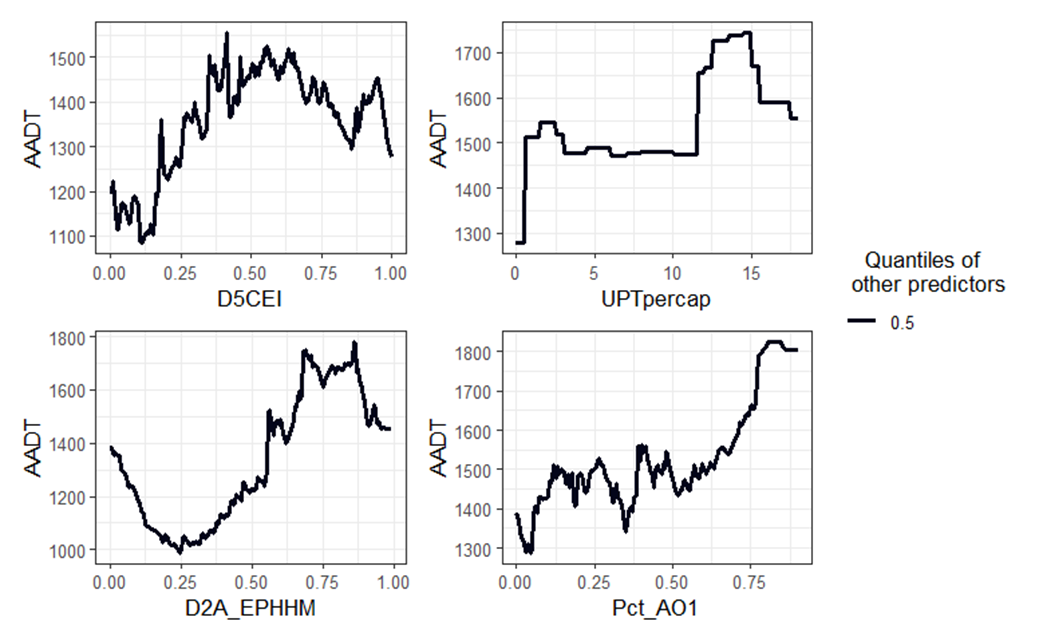


Figure ..Response curves of the top predictor variables.

### Two-Way Interaction Effects on AADT Prediction

The interaction between D5CEI and UPTpercap reveals that high centrality consistently contributes to elevated AADT, with its effect becoming stronger in areas with moderate to high transit usage, as shown in Figure 4.8 (a). Notably, AADT remains relatively high even in zones with lower transit ridership if D5CEI is substantial, suggesting that land-use centrality remains a critical factor for traffic generation, independent of transit reliance. Figure 4.8 (b) indicates that areas with a stronger employment-household balance (D2A\_EPHHM) and high transit usage tend to exhibit higher predicted AADT values. However, there is considerable variability in predictions across the range of D2A\_EPHHM values when UPTpercap is low, suggesting that employment-residence balance alone may not fully account for traffic patterns without considering transit availability. Figure 4.8 (c) reveals that areas with a higher share of working-age population (P\_WrkAge) and a moderate to low percentage of low-wage earners (E\_PctLowWa) tend to have higher predicted AADT. The pattern suggests that workforce composition plays a role in driving traffic demand, with the effect of low-wage concentration diminishing when a large proportion of the population is of working age. Figure 4.8 (d) illustrates a clear interaction between sustainability land-use characteristics (SLC\_score) and regional centrality. AADT increases with higher SLC scores but reaches the highest levels when combined with strong centrality. This suggests that even areas designed with sustainability in mind (e.g., walkability, mixed-use) can still experience high traffic volumes when centrally located. The highest AADT values are predicted for areas where both D5CEI and D2A\_EPHHM are high, as shown in Figure 4.8 (e), emphasizing that the synergy between central location and balanced land use leads to more intense traffic. In contrast, areas with low values for both factors show much lower AADT predictions. Figure 4.8 (f) indicates that while sustainable land-use scores and transit use are individually linked to higher AADT, their combined effect shows saturation at specific levels. As both SLC\_score and UPTpercap increase, AADT stabilizes, possibly indicating a shift towards multimodal transport environments where vehicular traffic no longer increases linearly.

|  |  |
| --- | --- |
|  |  |
| **(a)** | **(b)** |
|  |  |
| **(c)** | **(d)** |
| A screen shot of a graph  AI-generated content may be incorrect. | A screen shot of a graph  AI-generated content may be incorrect. |
| **(e)** | **(f)** |

Figure ..Two-way interaction effects of varying urban and transportation variables on AADT prediction.

## Key Findings

The key findings from the non-spatial RF model emphasize the critical role of urban accessibility, transit ridership, and walkability in shaping traffic patterns. Among the most influential predictors, the regional centrality index (D5CEI) and road network density (D3A) demonstrate strong positive associations with AADT, reinforcing the notion that well-connected and highly accessible areas experience greater vehicular movement. Notably, the response curves reveal nonlinear patterns, with traffic volume increasing sharply at mid-range values of these predictors before stabilizing, indicating a saturation effect. Additionally, transit ridership per capita (UPTpercap) and the walkability index (NatWalkInd) show irregular relationships with AADT, suggesting that areas with higher transit use and pedestrian-friendly environments may still experience high vehicle traffic, likely due to commercial and mixed-use developments that attract both transit users and drivers.

The two-way interaction analysis provides further insights into the interplay between urban and transportation variables in predicting AADT. The interaction between D3A and D5CEI highlights that while both factors independently contribute to higher traffic volumes, their combined effect is particularly strong in areas with high road density and regional accessibility. Similarly, the interaction between public transit ridership (UPTpercap) and road network density (D3A) suggests threshold effects, where transit availability helps manage congestion but does not necessarily reduce traffic beyond a certain level. The relationship between road density and greenhouse gas emissions (Annual\_GHG) reveals an inverse trend at high emission levels, suggesting that areas experiencing higher emissions may implement congestion mitigation strategies. Additionally, the interaction between employment-residential balance (D2A\_EPHHM) and regional centrality confirms that well-balanced, centrally located areas generate substantial traffic demand. Overall, these findings underscore the complex and nonlinear nature of urban mobility dynamics, highlighting the importance of integrated transportation and land-use planning strategies to balance accessibility, sustainability, and congestion management.

## Summary

Accurate AADT estimation is crucial for transportation planning and decision-making. This chapter includes the application of a Non-Spatial RF model to estimate AADT using key socioeconomic and transportation-related variables, addressing data limitations in conventional methods. The model revealed that factors such as regional centrality, road network density, and transit ridership significantly influence traffic volume. Additionally, response curves captured non-linear effects, where traffic volumes increase sharply at mid-range predictor values before stabilizing. Two-way interaction analysis further highlighted the combined impact of accessibility, sustainability, and transportation infrastructure on AADT.

This research provides a novel machine-learning-based approach for traffic volume estimation, emphasizing the complex interdependencies between transportation and urban development factors. The findings suggest that road infrastructure, walkability, and transit accessibility contribute to varying levels of traffic demand, underscoring the importance of multimodal transportation strategies. Additionally, insights from two-way interactions highlight the need for balanced land-use policies that accommodate vehicular and non-motorized mobility while minimizing congestion and emissions.

Despite its strong predictive performance, this study is limited by the exclusion of spatial autocorrelation effects, which could further refine AADT estimates. Future research should explore spatial machine learning approaches, integrate dynamic traffic count data, and assess the long-term impacts of urban policies on traffic volume. Transportation agencies can develop more adaptive and data-driven traffic management solutions by incorporating these elements.

# SPATIAL AI MODELING

**V.**

This chapter advances the traffic volume estimation framework by integrating spatial modeling techniques to estimate AADT on local roadways. Building upon prior analyses using non-spatial machine learning, this chapter introduces a spatial RF model that explicitly accounts for spatial dependencies in the data. The chapter details the modeling pipeline, including spatial diagnostics, variable selection, and the incorporation of spatial autocorrelation structures. To benchmark the effectiveness of spatial RF, a GWR model is also employed as a comparative baseline. Model performance is evaluated using standard metrics such as R² and RMSE, alongside spatial analysis of residuals. This comparison provides valuable insights into the advantages of spatially explicit modeling frameworks over traditional regression-based and non-spatial machine learning approaches, particularly in capturing local variations in traffic patterns.

## Motivation

Accurate estimation of traffic volumes on local roadways requires not only rich data but also modeling techniques that account for the inherently spatial nature of transportation systems. While non-spatial models have shown promising results in handling complex variable interactions, they often overlook spatial heterogeneity and autocorrelation. Local roads, in particular, exhibit spatially varying relationships between traffic volumes and influencing factors such as land use, transit access, and demographic patterns. Failing to account for these variations can lead to biased or inefficient estimates. This chapter addresses these limitations by introducing spatially informed modeling approaches, namely spatial RF and GWR, which allow coefficients or decision structures to vary by location. By incorporating spatial structure into the predictive framework, this study aims to enhance the accuracy, interpretability, and policy relevance of AADT estimation for local roadways, providing better understanding of traffic behavior across diverse urban contexts.

## Methodology

This section includes the methodological framework employed to estimate AADT on local roadways using a Spatial RF model. It details the data preprocessing steps, variable selection process, and sampling strategy used to construct the modeling dataset.

### Study Design

The analytical framework used in this study (as shown in Figure 5.1) follows a structured approach, beginning with data acquisition from the SLD. The data preparation phase entails the analysis of variable correlations to identify significant predictors and to ensure that the selected features contribute meaningfully to the subsequent modeling process. This study encompasses a comprehensive array of factors that influence the estimation of AADT. The variables considered in this study encompass multiple domains, including traffic and transportation (AADT, UPTpercap), urban form and land use (D3A, D5CEI, NatWalkInd), employment and economic indicators (D2A\_EPHHM, D2R\_WRKEMP, E\_PctLowWa, R\_PCTLOWWA, W\_P\_Highwa, W\_P\_Medwag), environmental and sustainability measures (Annual\_GHG, SLC\_score), and socio-demographic attributes (P\_WrkAge, Pct\_AO1).

The Data Analysis phase consists of four main steps. First, spatial autocorrelation analysis using Moran’s I is performed to assess spatial structure. Then, distance matrix calculation helps account for spatial dependencies. Multicollinearity analysis is used to eliminate redundant predictors based on variable importance and variance inflation factor thresholds. Next, promising variable interactions are identified to capture nonlinear synergistic effects between predictors. This results in a final refined set of 12 variables. These variables are then used to train the Spatial RF model, a spatially aware ensemble learning method that effectively captures spatial heterogeneity and complex interactions. For benchmarking, the model’s predictive performance is evaluated against the GWR model. The Output section includes key interpretative tools derived from the SRF model: (1) permutation importance plots to identify influential predictors, (2) partial dependence plots to visualize nonlinear effects, and (3) two-way interaction plots to explore how combinations of variables impact AADT. Together, this workflow highlights a comprehensive, spatially informed approach to traffic volume estimation. This methodological approach ensures that spatial dependencies are adequately considered while maintaining the integrity of predictive modeling based on machine learning. By combining spatial analysis with RF modeling, the study thoroughly assesses the factors influencing the dependent variable, offering valuable insights into urban planning, transportation policy, and sustainability initiatives.

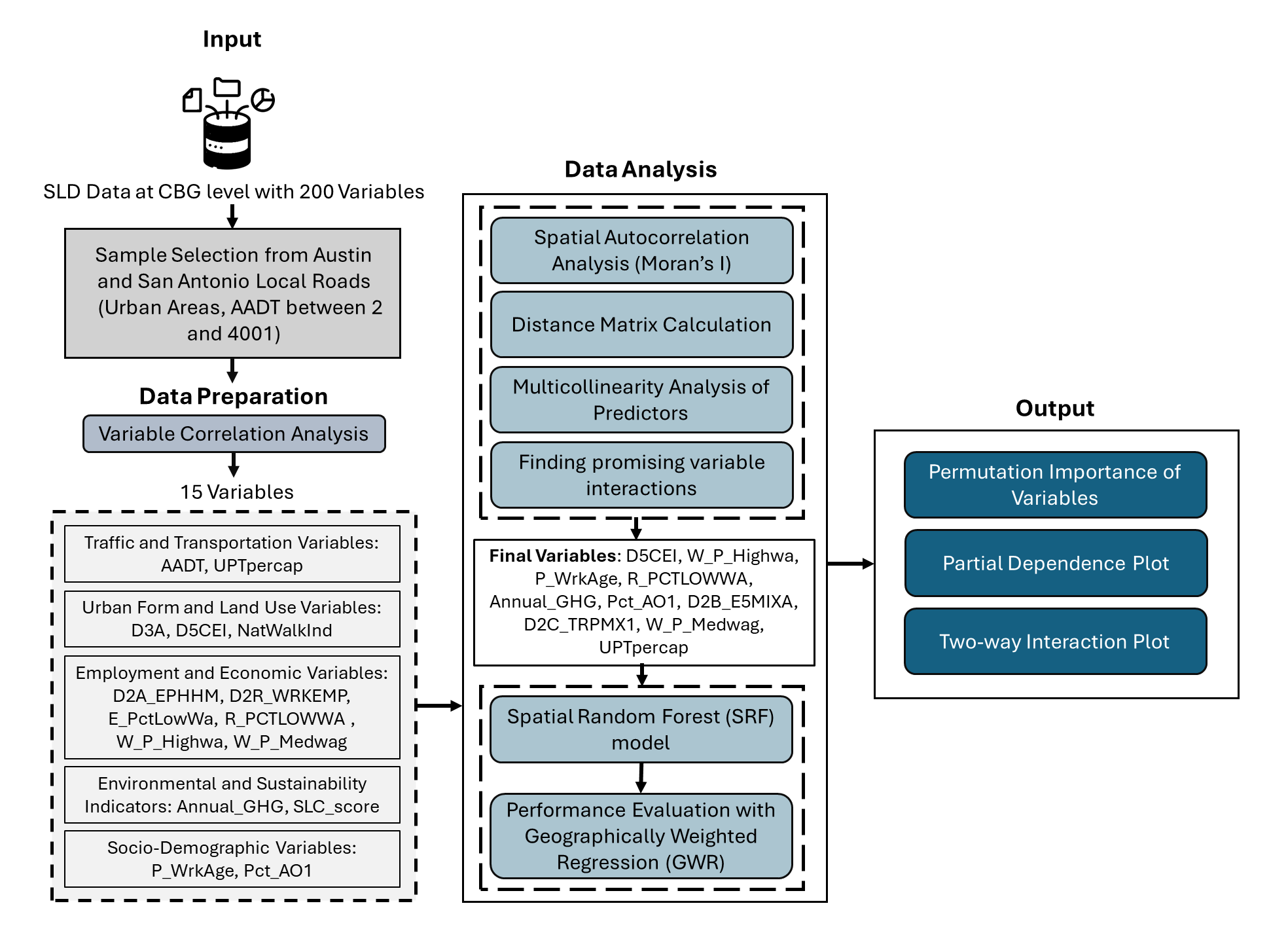


Figure ..Study design flowchart.

### Data Preparation

The dataset used in this study comes from the SLD and was selected specifically to support the estimation of AADT on local roadways. Since a range of factors influences AADT, it was important to start by understanding the distribution and characteristics of key variables in the dataset. These include indicators related to socioeconomic conditions, transportation accessibility, land use patterns, and greenhouse gas emissions, factors that are commonly linked to traffic volume. The full dataset contains 48,470 data points collected between 2019 and 2023. Out of these, 22,984 represent local roads, which are defined in this study as roads with an AADT value under 4,001. For the spatial modeling, the focus was narrowed to two major cities in Texas: Austin and San Antonio. Together, these cities contributed 6,116 data points for local roadways. To make the analysis more manageable while maintaining diversity in the data, a random sample of 5,000 observations was selected, covering AADT values ranging from 2 to 4,000. Figure 5.2 shows a density plot comparing the full dataset and the selected sample. Both follow a right-skewed distribution, meaning most roads have lower AADT values, with higher volumes being less common. The close alignment between the two distributions suggests that the sample is a good representation of the full dataset, allowing the analysis to reflect broader traffic trends on local roads in these cities.

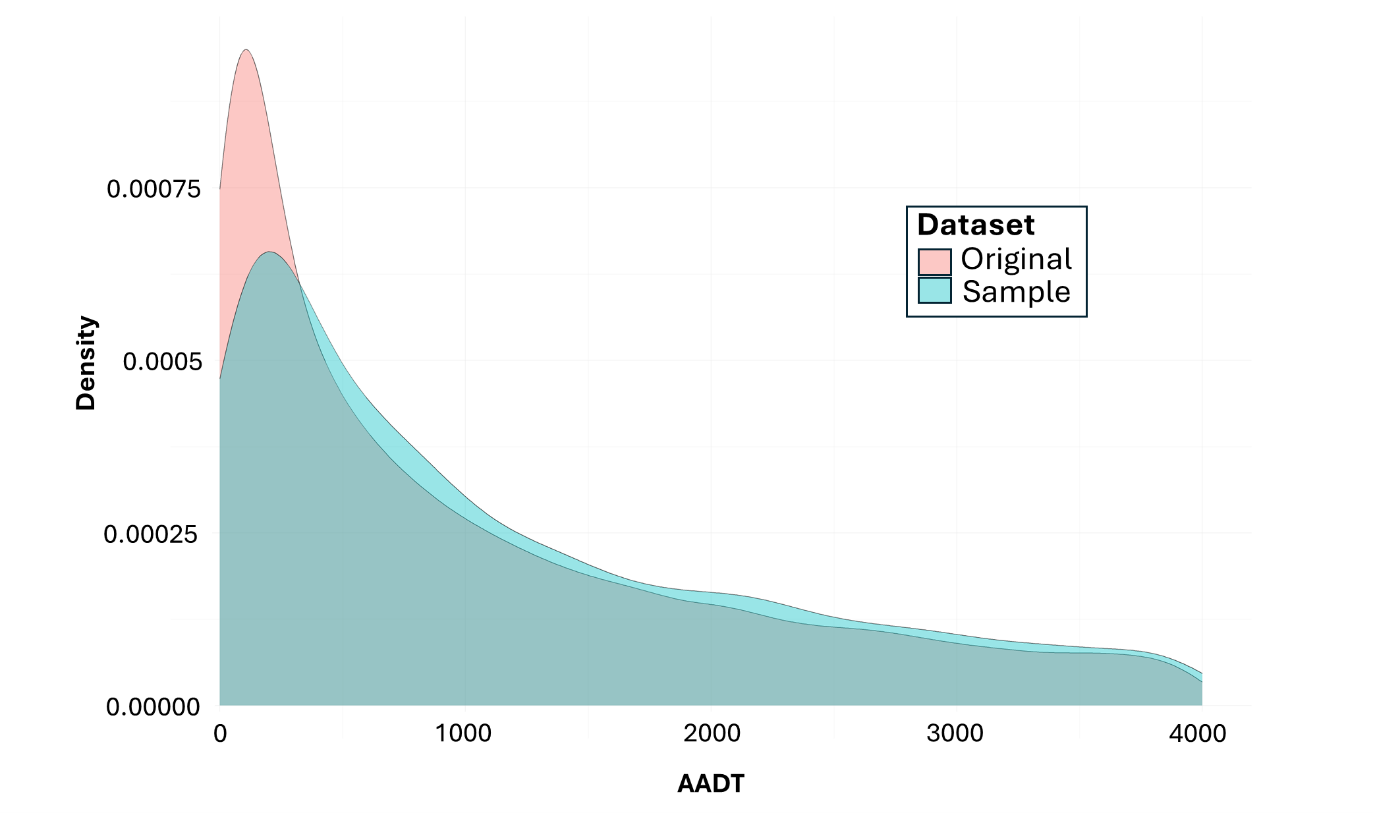


Figure ..Distribution of AADT in original and sample datasets.

These spatial distribution maps in Figure 5.3 provide a comprehensive view of how key predictor variables are geographically distributed across the study area, specifically focused on cities like Austin and San Antonio, with deep purple representing the highest value and light peach representing the lowest value. The concentration of higher values (shown in darker shades) in the urban cores for several variables, such as D5CEI (centrality), D3A (road density), and UPTpercap (transit ridership per capita). These concentrations indicate that more centrally located and transit-rich areas typically correspond with higher land-use intensity and accessibility. For example, variables like W\_P\_Highwa (high-wage employment), D2A\_EPHHM (employment-residence entropy), and SLC\_score (Smart Location Choice) also exhibit strong clustering patterns in these urban zones, suggesting that economic and infrastructure-related indicators are geographically co-located in high-density, high-accessibility regions.

Conversely, some variables, such as R\_PCTLOWWA (percentage of low-wage workers) and E\_PctLowWa (employment in low-wage sectors), appear more dispersed and less concentrated in urban centers, potentially reflecting a spatial mismatch between affordable housing and employment opportunities. The spatial heterogeneity of variables like Annual\_GHG (greenhouse gas emissions) and NatWalkInd (walkability index) further illustrates how environmental, and mobility characteristics vary within and between cities.

|  |  |  |
| --- | --- | --- |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  | |

Figure ..Spatial distributions of the variables.

Figure 5.4 displays a series of scatterplots illustrating the relationships between AADT and a set of key predictor variables derived from geospatial, socioeconomic, land use, and transportation domains. Each subplot visualizes one predictor variable plotted against AADT, with data points colored by density to highlight clustering. The visualizations allow for an initial assessment of variable influence, nonlinearity, and variance distribution across the traffic volume spectrum.

Several variables, such as D5CEI (regional centrality index) and Annual\_GHG (greenhouse gas emissions), exhibit pronounced positive associations with AADT, where traffic volumes tend to increase rapidly with rising values of these predictors before leveling off, suggesting threshold effects or saturation points in urban accessibility and emission-related impacts. D3A, representing road network density, also shows a moderate positive trend, consistent with the expectation that well-connected street grids facilitate higher vehicular flows.

On the other hand, socioeconomic variables such as R\_PCTLOWWA (percentage of low-wage workers) and E\_PctLowWa (employment-based low-wage share) show little to no strong correlation, potentially indicating that income levels may influence mode choice rather than traffic volume directly. UPTpercap (transit ridership per capita) exhibits a skewed distribution, with most observations clustered near zero, rendering its influence on AADT less interpretable through a simple linear model. Similarly, SLC\_score, intended to measure the sustainability and livability of locations, does not show a clear relationship with traffic volume.

The variables Pct\_AO1 (percentage of single-vehicle households) and D2A\_EPHHM (employment-household entropy) show upward trends in AADT with increasing values, reinforcing the influence of household vehicle ownership and land-use mix on travel demand. Other variables, such as P\_WrkAge, W\_P\_Highwa, and W\_P\_Medwag, display nonlinear patterns with wide AADT variance across the spectrum, suggesting that workforce characteristics contribute to traffic demand.

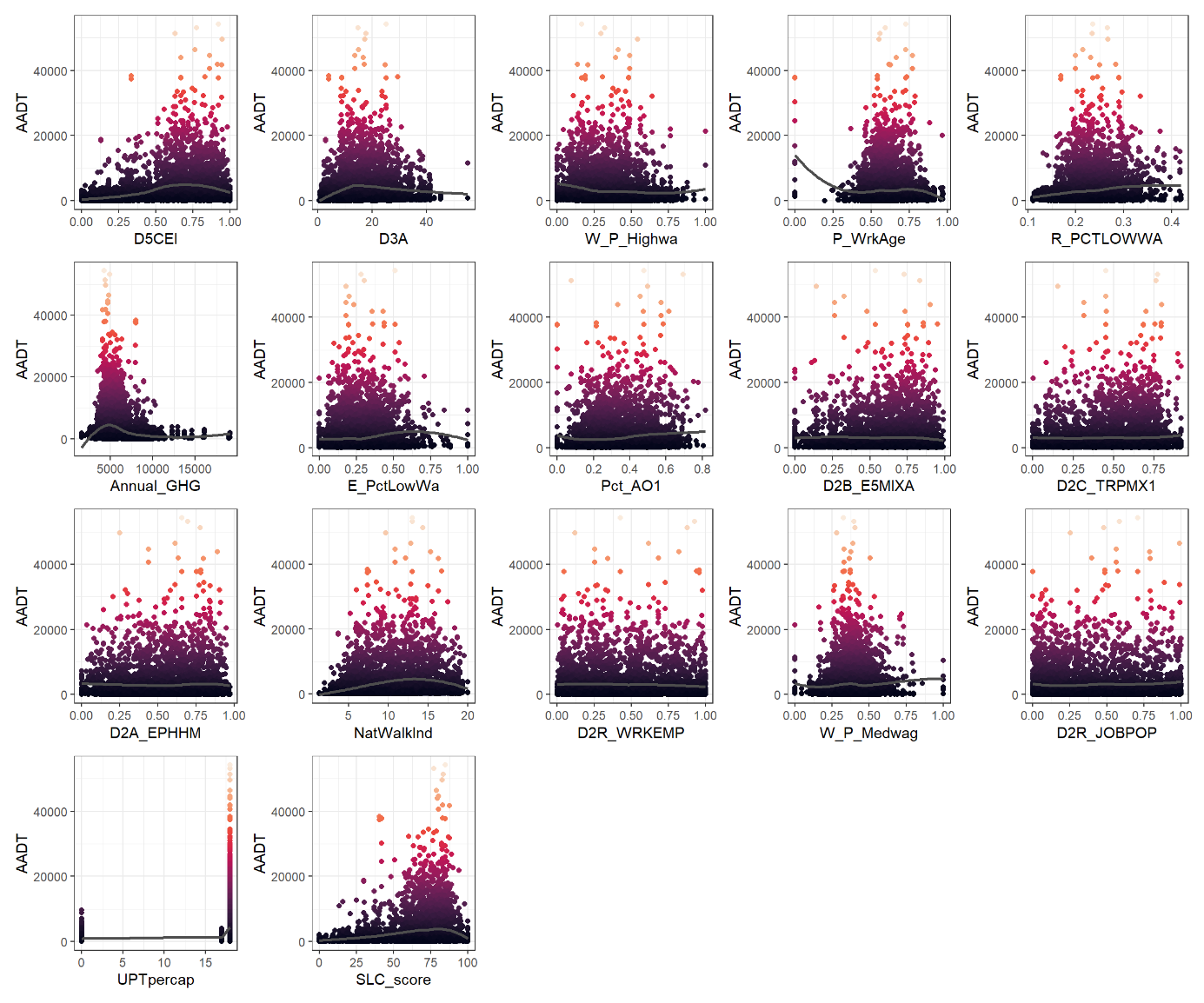


Figure ..Relationship between AADT and predictor variables.

### Spatial Autocorrelation Analysis of Predictors

To assess the presence of spatial structure in both the response variable (AADT) and the predictor variables, Moran’s I statistic was calculated at multiple distance thresholds (0 to 20,000 meters). Figure 5.5 visualizes the spatial autocorrelation patterns across 20 predictor variables and the AADT outcome, with darker shades indicating higher Moran’s I values. Black outlines mark statistically significant values (p < 0.05), highlighting where spatial clustering is meaningful.

The analysis reveals that AADT exhibits strong and statistically significant spatial autocorrelation from 1,000 meters onward, with Moran’s I values exceeding 0.5 at multiple thresholds, justifying the use of spatial modeling techniques. Several predictor variables, including D5CEI, UPTpercap, R\_PCTLOWWA, and Annual\_GHG, also demonstrate high spatial autocorrelation, particularly at medium to large distances (5,000–20,000 meters). These results suggest that these variables are not randomly distributed in space but follow discernible spatial patterns, making them suitable for spatial regression models.

Notably, D5CEI shows consistently high and significant Moran’s I values at all thresholds beyond 1,000 meters, indicating strong regional clustering of centrality. UPTpercap, a proxy for transit usage, also shows a similar spatial clustering effect, which aligns with the urban concentration of transit services. In contrast, variables like D2B\_E5MIXA and P\_WrkAge exhibit weaker spatial patterns, with lower Moran’s I values and insignificant p-values at most thresholds.

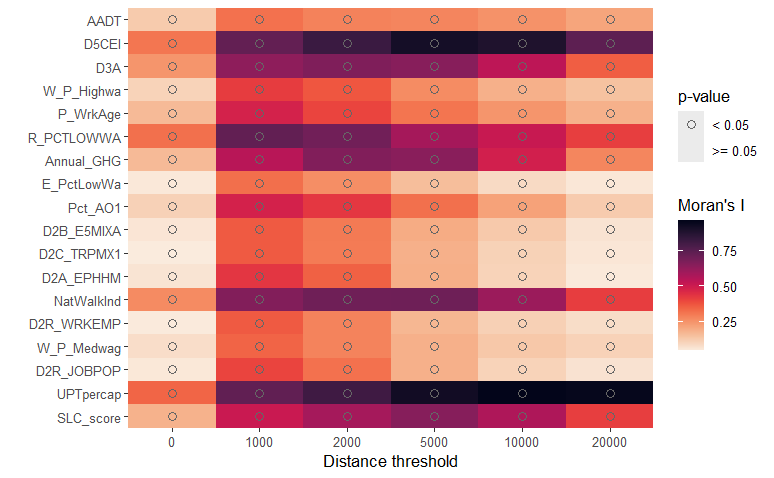


Figure . Moran’s I heatmap of spatial autocorrelation across predictors at varying distance thresholds.

### Reducing multicollinearity in the predictors

To enhance model interpretability and stability, multicollinearity among predictor variables was systematically addressed prior to spatial modeling. Multicollinearity occurs when two or more predictors exhibit high intercorrelation, leading to inflated variance in model estimates, reduced predictive performance, and misleading interpretations of variable importance. This study employed a two-step diagnostic procedure using the correlation thresholding and variance inflation factor (VIF) analysis.

First, pairwise correlations were evaluated with a conservative threshold of 0.85. No variables exceeded this threshold, indicating the absence of strong bivariate collinearity. However, this does not rule out multivariate redundancy. Therefore, a follow-up analysis was performed with a VIF threshold of 2. The VIF test revealed several variables with high redundancy in variance explained by the remaining predictors. Specifically, the variables D3A (road density), E\_PctLowWa (employment in low-wage jobs), D2A\_EPHHM (employment-household entropy), NatWalkInd (walkability index), D2R\_WRKEMP (employment-residential ratio), D2R\_JOBPOP (job-population balance), and SLC\_score (sustainable location criteria) were removed from the modeling dataset. The preference order for retention was determined based on prior permutation importance scores, ensuring that more influential variables were prioritized. This preprocessing step eliminated multicollinear predictors, enhancing the robustness of the spatial RF model and mitigating potential overfitting from redundant or overlapping predictor effects.

### Theory/Model

***Spatial RF***

Spatial data analysis involves the examination of geographically distributed data, where spatial dependencies and autocorrelations play a crucial role in predictive modeling. Traditional machine learning methods, such as classical RF, assume that observations are independent and identically distributed, which does not hold in spatial contexts due to inherent spatial structures (Breiman, 2001b). To address this, Spatial RF integrates higher-order spatial statistics to model spatial dependencies effectively (Talebi et al., 2022b)​. The RF algorithm creates a collection of decision trees, with each tree trained on a randomly selected subset of the training data through a bootstrap sampling method. Predictions are aggregated using majority voting (classification) or averaging (regression).

Given an input dataset , where represents the predictor variables and represents the target variable, the RF prediction is computed as:

Where is the prediction of the j-th tree.

SRF model builds upon the classical RF algorithm by incorporating spatially aware decision trees. This approach accounts for spatial autocorrelation, where observations closer in space tend to be more similar than those farther apart.

**Vectorized Spatial Patterns:** Instead of treating observations as independent, SRF leverages vectorized spatial patterns to capture spatial dependencies (Talebi et al., 2022b):

Where, represents the vectorized spatial patterns of input data at the location . is the observed response variable at ​. Each spatial tree in the SRF model is trained on bootstrap samples ​, and a random subset of predictors is selected at each split.

**Spatial Splitting Criteria:** The SRF algorithm modifies the classical RF splitting criterion by incorporating local mean and variance-based impurity measures (Talebi et al., 2022b). For regression, the optimal split is chosen by minimizing the variance within nodes:

Where represents the mean response variable at the node to be split. The Gini impurity index is used for classification. The equation is:

Where ​ is the proportion of class k in the node.

**Prediction in SRF:**

For a new spatial pattern , the SRF prediction is given by:

For classification:

Where *I* is an indicator function.

**Out-of-Bag (OOB) Estimation:** The generalization error in SRF is estimated using Out-of-Bag (OOB) predictions as shown in the equations below (Talebi et al., 2022b).

where is the number of trees for which the pattern is out-of-bag​.

**Handling Spatial Autocorrelation:** SRF explicitly accounts for spatial autocorrelation by incorporating spatial relationships into the modeling process. One approach is through local spatial-spectral information, which captures neighborhood characteristics and enhances the model’s ability to learn spatial heterogeneity. By leveraging spatial proximity, SRF ensures that geographically close observations are more likely to share similar properties, improving prediction accuracy in spatially dependent datasets (Hengl et al., 2018b). This technique is particularly useful in land-use classification, traffic modeling, and environmental monitoring applications, where local spatial patterns influence the target variable (Georganos et al., 2021b). Another key feature of SRF is multi-resolution data handling, which allows predictors from different spatial scales to be integrated into the model. Unlike traditional RF, which assumes uniform resolution across all predictors, SRF can combine high-resolution and low-resolution spatial data, making it more versatile for real-world geospatial applications (Meyer et al., 2018). This capability ensures that relevant spatial features are incorporated regardless of their scale, enhancing model robustness and applicability across diverse datasets. Additionally, SRF effectively manages missing values without requiring explicit imputation. Traditional machine learning models often struggle with missing spatial data, necessitating preprocessing techniques such as interpolation or mean imputation. However, SRF can seamlessly handle missing observations by leveraging spatial dependencies and decision tree structures, allowing predictions to remain reliable even in incomplete datasets (Georganos et al., 2021b). This feature is particularly beneficial for geospatial studies where data availability varies across regions, reducing the need for extensive preprocessing while maintaining model accuracy.

***Geographic Weighted Regression (GWR)***

GWR represents a localized variant of linear regression that allows model parameters to vary across space, making it particularly suitable for analyzing spatially heterogeneous relationships. Originally introduced by Brunsdon et al. (1996), GWR addresses one of the key limitations of traditional global regression models, namely, the assumption that relationships between dependent and independent variables are stationary across space. Instead, GWR recognizes that these relationships may differ at different geographic locations due to underlying contextual, environmental, or socioeconomic factors. The basic formulation of GWR is expressed as:

Where, is the dependent variable at location *i*; ​ are the independent variables, are the location-specific parameters estimated at spatial coordinates , is the random error term.

GWR develops an individual regression equation for every site in the dataset by assigning weights to nearby observations. These weights are determined using a kernel function, typically Gaussian or bisquare, based on the spatial proximity between observations. The bandwidth of the kernel function controls the degree of spatial smoothing and can be fixed or adaptive. The optimal bandwidth is typically selected via cross-validation or by minimizing a model selection criterion such as the corrected Akaike Information Criterion (AIC) (Fotheringham et al., 2002). One of the strengths of GWR is its ability to reveal spatial non-stationarity, i.e., the idea that regression coefficients change across geographic space. This is particularly valuable in transportation, environmental, and urban studies, where localized factors often influence system behaviors. However, GWR is not without limitations. It is computationally intensive for large datasets and may suffer from issues like multicollinearity, overfitting, and interpretational complexity if too many predictors or small bandwidths are used (Wheeler and Tiefelsdorf, 2005). Additionally, GWR assumes that the observed spatial variation is entirely captured by the spatial coordinates, which may not account for unmeasured spatial processes or non-spatial dependencies.

## Analysis and Results

This section summarizes the spatial RF model results, including the important variables for AADT estimation, response curves, and non-linear impacts of the predictor variables.

### Predictive Performance of the Model

The Spatial RF model was fitted using 500 decision trees and 14 predictor variables across 5,000 observations. This ensemble-based machine learning model yielded strong predictive performance. The OOB R² value was 0.244, indicating reasonable internal validation accuracy. When comparing observed and predicted values, the model achieved an R² of 0.615 (shown in Table 5.1), which means approximately 61.5% of the variation in AADT was captured by the model. The RMSE of 3,211.7 reflects moderate prediction error, and the normalized RMSE of 1.15 suggests that while the model is better than a simple baseline, some variability remains unexplained. Importantly, the residuals show minimal spatial autocorrelation beyond the smallest distances, as indicated by insignificant Moran’s I values at 10,000 meters and beyond. This indicates the model recognizes spatial patterns in the data. However, the Shapiro-Wilk test indicates non-normality in residuals, which is expected in tree-based models and does not violate their assumptions.

When compared to the GWR model, the Spatial RF demonstrates superior predictive accuracy. The GWR model produced a much lower R² value of 0.138, explaining only 13.8% of the variability in AADT. Moreover, the GWR model’s RMSE of 4,784.67 is considerably higher than that of the Spatial RF model (3,211.70), indicating poorer predictive performance. Although GWR offers location-specific coefficient estimates and interpretable spatial variation, it is inherently limited by its linear structure and assumptions of localized linearity. In contrast, the Spatial RF captures nonlinear relationships and complex variable interactions across space without requiring explicit spatial coefficients. It leverages spatial cross-validation and local permutation importance to reflect spatial heterogeneity in a flexible manner. Furthermore, Spatial RF avoids the issue of overfitting common in GWR when too many predictors or small bandwidths are used (Chang Chien et al., 2020). The reduction in residual spatial autocorrelation in the Spatial RF model supports the idea that spatial patterns were effectively learned by the model.

Table .. Comparison of Model Performance Metrics.

|  |  |  |
| --- | --- | --- |
| **Metric** | **Spatial RF** | **GWR** |
| R² (Observed vs Predicted) | 0.615 | 0.138 |
| RMSE | 3,211.70 | 4,784.67 |
| Residual Spatial Autocorrelation | None beyond 10km | — |
| Residual Normality (Shapiro-Wilk) | W = 0.742, p < 0.001 | — |
| Interpretation | Stronger fit, nonlinear, less residual clustering | Weaker fit, linear, spatial coefficients |

### Promising Variable Interactions

The variable interaction analysis revealed several promising combinations of predictors (shown in Figure 5.6 and Table 5.2) that enhance the model’s ability to explain variations in AADT. Among these, the most influential interaction was between UPTpercap (transit ridership per capita) and D5CEI (regional centrality index). This interaction achieved the highest relative importance score (100%) and resulted in an R² improvement of 0.024, indicating that areas with both high centrality and high transit ridership contribute significantly to elevated traffic volumes. This suggests a synergistic effect, where transit-accessible and centrally located regions tend to exhibit high multimodal transportation activity, reinforcing vehicle travel demand rather than replacing it.

Another strong interaction was observed between D5CEI and R\_PCTLOWWA (regional share of low-wage workers). This combination produced an R² improvement of 0.018 and had a relatively high correlation (0.76) with individual predictors. The interaction implies that regional centrality amplifies traffic volumes, particularly in areas with a concentrated low-wage workforce, potentially due to commute-related travel from residential zones to employment centers. This supports previous literature noting that economically constrained populations often rely more on car travel in poorly connected transit areas, despite being concentrated in central locations.

The third notable interaction involved UPTpercap and Pct\_AO1 (percentage of single-vehicle households), evaluated through a principal component transformation. This pair yielded an R² gain of 0.011, suggesting that household-level vehicle access modifies how transit usage influences traffic volume. In urban areas where transit is available, but households predominantly rely on a single car, AADT may rise due to overlapping modal choices for different trip types. Similarly, the interaction between D5CEI and Annual\_GHG, with an R² increase of 0.0155, demonstrates the interconnectedness of centrality and emission patterns. High-emission zones are often those with intensified traffic activity, and this interaction highlights the nonlinear relationship between vehicular emissions and traffic density, especially in highly connected regions.

These interaction effects offer valuable insights into the underlying spatial and behavioral patterns shaping local roadway usage. The bottom-right violin plot in Figure 5.6 further confirms the utility of incorporating interaction terms: models with interactions achieved consistently higher R² scores across 10 spatial folds compared to those without. This highlights the importance of considering complex variable relationships to enhance traffic volume estimation, especially in diverse urban environments.

|  |  |
| --- | --- |
|  |  |
|  |  |
|  | |

Figure .. Top Variable interactions and their impact on AADT prediction performance in the Spatial RF model.

Table .. Promising Variable Interactions Identified for AADT Estimation Using Spatial RF.

|  |  |  |  |
| --- | --- | --- | --- |
| **Interaction** | **Importance (% of max)** | **R-squared improvement** | **Max correlation with predictors** |
| UPTpercap..x..D5CEI | 100.0 | 0.0240 | 0.840 |
| D5CEI..x..R\_PCTLOWWA | 94.5 | 0.0180 | 0.760 |
| UPTpercap..pca..Pct\_AO1 | 75.4 | 0.0110 | 0.720 |
| D5CEI..x..Annual\_GHG | 70.5 | 0.0155 | 0.667 |

### Importance of Factors in Estimating AADT

The variable importance plot in Figure 5.7, derived from the Spatial RF model, highlights the most influential predictors of AADT across local roadways. The importance value is given in Table 5.3. The most significant contributor is the interaction between public transit usage per capita and regional centrality (UPTpercap × D5CEI), suggesting that areas with high central accessibility and greater public transit presence tend to experience increased traffic volumes. This is followed closely by the individual contributions of D5CEI and UPTpercap, reinforcing the notion that accessibility and transit usage are central to understanding traffic demand. Additional interaction terms, such as D5CEI with the proportion of low-wage workers (R\_PCTLOWWA), and D5CEI with greenhouse gas emissions (Annual\_GHG), also ranked highly, indicating that spatial accessibility modifies the influence of socioeconomic and environmental factors in shaping traffic flow.

Lower on the importance scale are variables like the proportion of one-vehicle households (Pct\_AO1), employment entropy, and workplace wage characteristics. While these factors still contribute meaningfully to the model, their impact on prediction accuracy is more limited compared to access-related metrics. The importance scores are based on permutation error increases, meaning the model’s performance noticeably declines when high-ranking variables are randomly permuted, demonstrating their critical role in prediction. The spatial RF model’s ability to capture these nonlinear interactions provides richer, more realistic insights into the drivers of traffic volume compared to traditional linear approaches.

A graph with red dots and numbers

AI-generated content may be incorrect.

Figure .. Important variables from the Spatial RF model.

Table .. Variable Importance from the Spatial RF Model.

|  |  |  |
| --- | --- | --- |
| **Rank** | **Variable** | **Importance (Mean Error Increase)** |
| 1 | UPTpercap × D5CEI | 3297.166 |
| 2 | D5CEI | 3079.507 |
| 3 | UPTpercap | 2922.533 |
| 4 | D5CEI × R\_PCTLOWWA | 2391.889 |
| 5 | Annual\_GHG | 2069.174 |
| 6 | R\_PCTLOWWA | 2065.101 |
| 7 | D5CEI × Annual\_GHG | 2014.849 |
| 8 | UPTpercap (PCA) × Pct\_AO1 | 1920.701 |
| 9 | Pct\_AO1 | 1864.137 |
| 10 | W\_P\_Highwa | 1551.263 |
| 11 | D2C\_TRPMX1 | 1395.051 |
| 12 | P\_WrkAge | 1363.609 |
| 13 | D2B\_E5MIXA | 1362.251 |
| 14 | W\_P\_Medwag | 1322.853 |

### Non-linear Impacts of AADT Predictors

The plots in Figure 5.8 illustrate the non-linear relationships between key predictor variables and AADT as captured by the Spatial RF model. Notably, the interaction term between public transit usage per capita (UPTpercap) and regional accessibility (D5CEI) reveals a strong upward trend in AADT, especially as both values increase, indicating that high accessibility in areas with greater transit usage leads to increased traffic volumes. This pattern aligns with findings from Zhao and Park (2004b), who demonstrated that accessibility metrics significantly influence AADT, especially in urban contexts. The marginal effect of D5CEI itself shows a clear non-linear increase in AADT, plateauing at higher centrality levels, suggesting a saturation point where further increases in centrality contribute little to traffic volume, a trend also observed in study by Lämmer et al. (2006), which emphasized diminishing returns of land-use centrality on traffic intensity.

UPTpercap alone demonstrates an increase in AADT at extreme values, which may signify areas of exceptionally high transit ridership coinciding with multimodal transport hubs or densely populated urban centers. Meanwhile, Annual Greenhouse Gas emissions (Annual\_GHG) display a declining relationship with AADT beyond a specific threshold, possibly reflecting congestion zones or effective emission mitigation strategies. The relationship between R\_PCTLOWWA (percentage of workers with low wages) and AADT is more complex, exhibiting a U-shaped pattern that may reflect varying transportation behavior across income groups, initially lower AADT in low-wage areas due to transit reliance, then rising again due to multimodal or longer-distance commutes.

Finally, the interaction terms D5CEI × R\_PCTLOWWA and D5CEI × Annual\_GHG further emphasize the value of modeling joint effects. For instance, D5CEI × R\_PCTLOWWA indicates that regions with higher accessibility and low-wage concentrations experience increased traffic volumes, potentially due to concentrated economic activity. These interaction effects highlight the strength of spatial RF models in uncovering complex, non-linear, and spatially heterogeneous relationships, which traditional linear models often fail to capture. The use of partial dependence plots, as presented here, provides valuable insight into these detailed effects, reinforcing the growing body of literature supporting machine learning approaches for spatial traffic modeling.

A graph of a number of numbers

AI-generated content may be incorrect.

Figure .. Response curves of the top predictor variables.

### Two-Way Interaction Effects on AADT Prediction

The plot in Figure 5.9 shows a strong interaction between regional centrality (D5CEI) and the proportion of workers earning medium wages (W\_P\_Medwag) on predicted AADT values. The upper-left quadrant, characterized by high D5CEI and low W\_P\_Medwag, displays the highest predicted AADT levels, suggesting that highly accessible areas with fewer medium-wage workers experience intense traffic volumes, possibly due to the presence of commercial or high-income residential zones. On the other hand, regions with both low D5CEI and low W\_P\_Medwag exhibit significantly lower predicted AADT, reflecting lower mobility in peripheral areas with less economic activity. Notably, even with increasing W\_P\_Medwag, AADT does not increase unless D5CEI is high, highlighting centrality’s dominant influence in driving traffic volume.

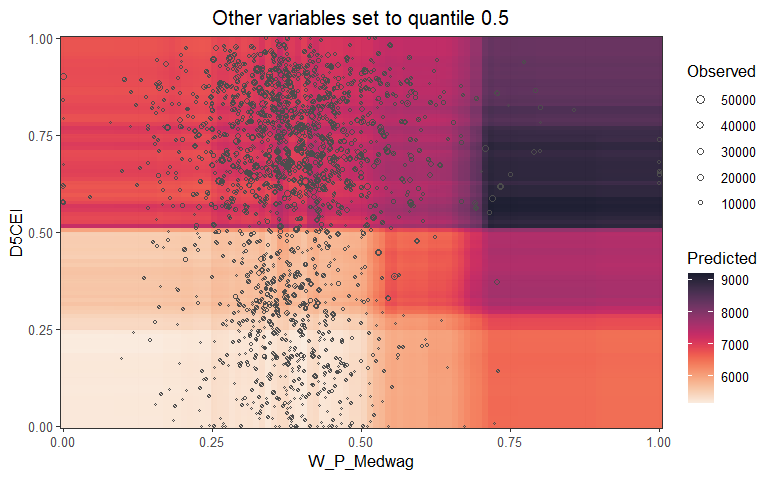


Figure .. Impact of regional accessibility and medium-income workforce on the AADT.

The interaction between D5CEI and Annual\_GHG emissions in Figure 5.10 reveals interesting spatial dynamics in traffic volume. Predicted AADT is highest in zones where both centrality and emissions are elevated, confirming that more accessible areas not only experience higher traffic but also contribute disproportionately to vehicular emissions. AADT levels remain low in areas with both low D5CEI and low GHG emissions, typically suburban or rural areas with limited connectivity and vehicle activity. The model suggests a threshold effect, beyond a certain level of emissions, traffic volumes do not increase significantly unless coupled with high D5CEI, indicating that emissions alone are not predictive without accounting for accessibility.

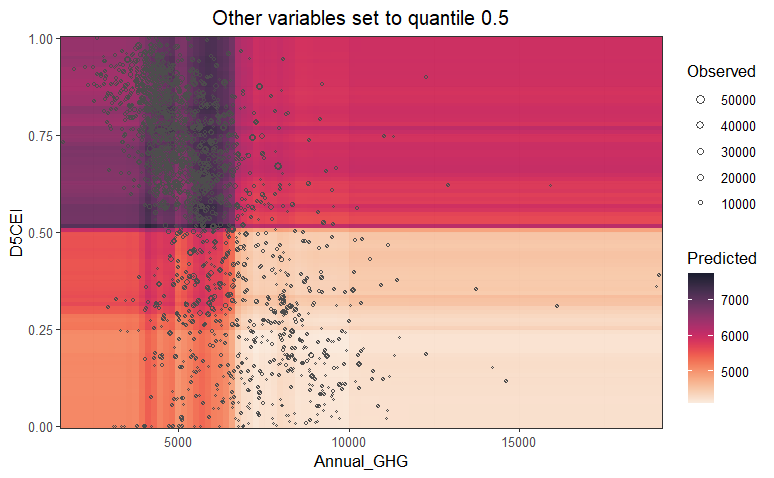


Figure .. Combined influence of centrality and greenhouse gas emissions on traffic volume.

The plot in Figure 5.11 uncovers a complex, non-linear interaction between the proportion of low-wage workers (R\_PCTLOWWA) and Annual\_GHG emissions. AADT values tend to remain moderate across the range of R\_PCTLOWWA when emissions are low, but as GHG emissions increase, predicted AADT exhibits a dip and subsequent rise, possibly indicating mixed land use areas. The observed traffic is highest in regions with low R\_PCTLOWWA and high emissions, suggesting these might be commercial or higher-income areas with dense vehicular activity. Conversely, areas with a higher proportion of low-wage workers are associated with more moderate AADT even when emissions are elevated, possibly reflecting increased reliance on non-motorized or public transport.

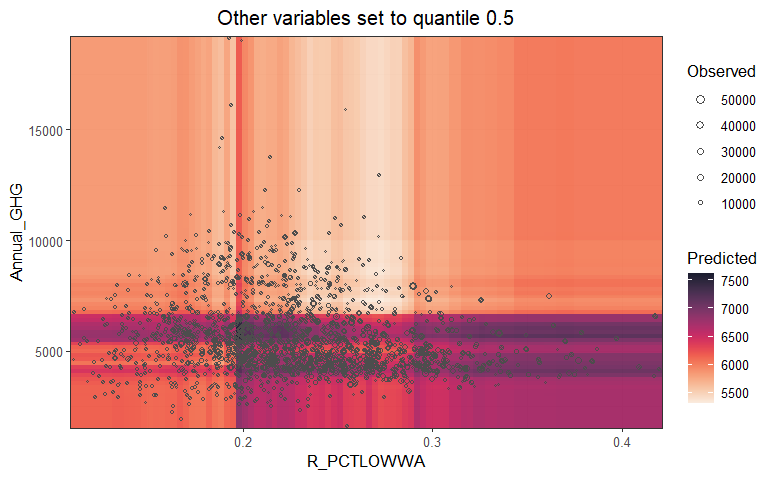


Figure .. Interaction between the low-wage workforce share and emissions in predicting AADT.

The interaction between socioeconomic composition and regional centrality is vividly demonstrated in this plot. The highest predicted AADT values are observed where D5CEI is high and R\_PCTLOWWA is low, reinforcing the idea that centrally located, higher-income zones attract greater traffic volumes. In contrast, peripheral areas with a higher concentration of lower-income populations exhibit substantially lower traffic volumes, indicating limited car ownership or reduced accessibility. Notably, central locations still exhibit elevated AADT even when low-wage populations are prevalent, implying that the built environment and intensity of urban activity in these hubs can override demographic constraints in driving travel demand.

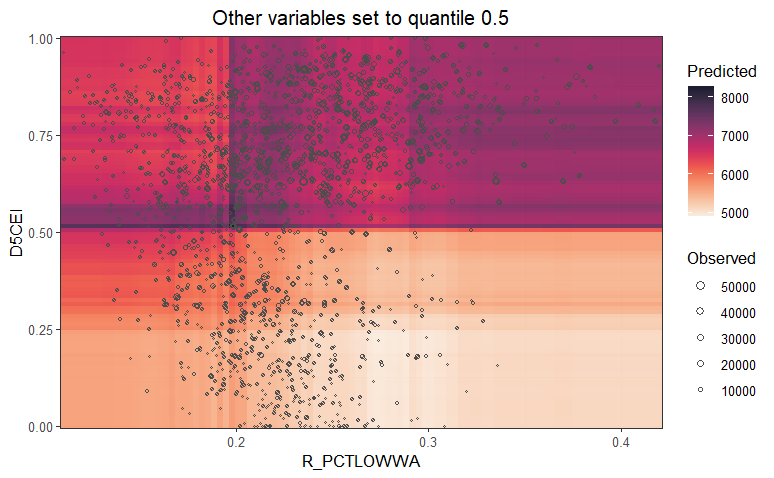


Figure .. Combined effect of socioeconomic status and centrality on traffic intensity.

## Key Findings

This study developed a Spatial RF model to estimate AADT on local roadways by incorporating a diverse set of spatial, socioeconomic, environmental, and built environment predictors. The model achieved significantly better predictive accuracy than the traditional GWR approach, explaining approximately 61.5% of the variation in AADT and producing lower prediction error. Importantly, the residuals from the Spatial RF model exhibited minimal spatial clustering at broader distances, indicating that spatial patterns were effectively captured without the risk of overfitting. These results affirm the strength of ensemble machine learning approaches for uncovering complex spatial relationships beyond the capabilities of linear spatial models.

The analysis identified several meaningful two-way interactions among predictors that substantially enhanced the model’s performance. Notably, areas characterized by both high transit usage and strong regional accessibility were found to have significantly higher traffic volumes, reflecting the reinforcing nature of multimodal connectivity and central urban form. Similarly, interactions between regional accessibility and socioeconomic composition, particularly areas with larger shares of low-income workers, highlighted how built environment characteristics can amplify or mediate demographic influences on travel demand. These patterns suggest that centrally located neighborhoods with concentrated employment opportunities or dense development can generate substantial traffic, even when income levels might otherwise imply lower vehicle ownership.

Other important interactions emerged between central accessibility and environmental indicators such as transportation-related emissions. In regions where both emissions and accessibility were high, traffic volumes were also elevated, indicating that central hubs are not only activity centers but also key contributors to environmental impacts from transportation. Additionally, interactions between household vehicle availability and transit usage indicate the behavior og vehicular dependency. Access to a single car in areas rich in transit still leads to increased roadway usage, possibly because of diverse trip purposes or multi-modal behaviors.

The variable importance analysis revealed that regional accessibility and transit ridership, both individually and in combination, were the most influential drivers of local traffic volumes. These were followed by socioeconomic status and environmental exposure measures, which exerted nonlinear effects depending on contextual conditions. For instance, the relationship between income level and traffic volume followed a U-shaped pattern, suggesting that both low- and high-income populations can generate substantial traffic under certain conditions, likely due to differing reliance on personal vehicles or commute distances. Similarly, environmental metrics showed threshold effects, where their predictive power increased only when paired with accessibility indicators.

Visualizations of two-way effects confirmed the critical role of context in shaping traffic demand. High traffic volumes were consistently observed in centrally located areas with relatively higher economic activity or limited vehicle availability, suggesting that both spatial form and social factors work in tandem to influence roadway usage. Overall, the findings highlight the importance of modeling interactions among land use, mobility resources, environmental burdens, and social conditions to accurately understand and estimate traffic patterns in complex urban systems.

## Summary

This chapter applied a Spatial RF model to estimate local roadway traffic volumes using a diverse set of predictors related to transit accessibility, socioeconomic conditions, land use, and environmental factors. The model demonstrated strong predictive performance, outperforming the traditional GWR approach by capturing nonlinear relationships and reducing spatial autocorrelation in residuals. Key predictors included measures of regional centrality, public transit usage, and household vehicle access, which collectively explained over 60% of the variation in AADT.

The analysis also highlighted the value of exploring two-way interactions between variables. Notably, high traffic volumes were associated with areas that combined strong accessibility with high transit use, or with low-income populations situated in central locations. Interactions between accessibility and emissions, as well as income composition and vehicle ownership, revealed complex travel behaviors shaped by both spatial form and demographic context. These findings highlight the importance of considering joint effects in traffic modeling, providing valuable insights for transportation planning in diverse urban settings.

# CONCLUSION

**VI**

This chapter presents the overall conclusions derived from the analyses conducted throughout the thesis. It synthesizes the key findings from both the non-spatial and spatial RF modeling approaches used to estimate AADT on local roadways. The chapter highlights the most influential factors identified, interprets their interactions, and reflects on how the incorporation of spatial heterogeneity improves model performance and insights. Additionally, it discusses the main limitations encountered in the research and outlines potential directions for future work to enhance the scope, applicability, and interpretability of traffic volume estimation models in urban planning contexts.

## Key Findings

This thesis presents a comprehensive investigation into the estimation of AADT on local roadways using RF models, emphasizing both non-spatial and spatial modeling frameworks. The non-spatial RF model served as a baseline to identify important predictors and uncover initial non-linear effects of land use, socioeconomic, and environmental variables on traffic volumes. However, it was the spatial RF model that demonstrated superior predictive capabilities by incorporating spatial heterogeneity through geographically aware resampling and localized permutation importance. The spatial model not only achieved higher accuracy but also revealed two-way interactions between factors such as regional accessibility, public transit use, emissions, and socioeconomic composition. These findings underscore that traffic demand is shaped by complex, context-specific interactions rather than isolated influences, highlighting the need for integrative approaches in transportation planning. Moreover, accessibility consistently emerged as the most dominant factor across models, reinforcing its foundational role in shaping urban mobility patterns.

## Limitations

While the study offers valuable insights, several limitations must be acknowledged. The reliance on static predictors and cross-sectional AADT data limits the ability to capture temporal dynamics or seasonal fluctuations in traffic volumes. Additionally, although the spatial RF model accounts for geographic variability, it does not produce interpretable spatial coefficients like traditional regression-based methods, which may pose challenges for policy translation. The model also depends on the quality and resolution of the input data; inaccuracies or inconsistencies in source variables (e.g., emissions estimates or socioeconomic indicators) can propagate through the modeling process.

## Future Scope

Future work could extend this analysis by incorporating longitudinal traffic data to explore temporal changes in traffic behavior, enabling dynamic forecasting under different policy or land use scenarios. Integrating real-time or high-frequency data from sensors, GPS, or mobile applications could enhance the granularity and responsiveness of the models. Furthermore, adopting interpretable machine learning techniques or hybrid frameworks (e.g., combining spatial RF with SHAP or GWR post-hoc interpretation) could help bridge the gap between predictive accuracy and policy usability. Finally, expanding the modeling framework to account for multimodal transportation behavior, especially walking, cycling, and shared mobility services, would provide a more holistic understanding of urban travel demand and support more equitable and sustainable mobility planning.

# REFERENCES

AASHTO, 2019. Guidelines for Geometric Design of Low-Volume Roads, 2nd Edition.

Apronti, D., Ksaibati, K., Gerow, K., Hepner, J.J., 2016. Estimating traffic volume on Wyoming low volume roads using linear and logistic regression methods. Journal of Traffic and Transportation Engineering (English Edition) 3, 493–506. https://doi.org/10.1016/j.jtte.2016.02.004

Baffoe-Twum, E., Asa, E., Awuku, B., 2022. Estimation of annual average daily traffic (AADT) data for low-volume roads: a systematic literature review and meta-analysis. Emerald Open Research 1. https://doi.org/10.1108/EOR-05-2023-0010

Benito, B., 2023. spatialRF: Easy Spatial Modeling with Random Forest in R.

Breiman, L., 2001a. Random Forests. Machine Learning 45, 5–32. https://doi.org/10.1023/A:1010933404324

Breiman, L., 2001b. Random Forests. Machine Learning 45, 5–32. https://doi.org/10.1023/A:1010933404324

Brunsdon, C., Fotheringham, A.S., Charlton, M.E., 1996. Geographically Weighted Regression: A Method for Exploring Spatial Nonstationarity. Geographical Analysis 28, 281–298. https://doi.org/10.1111/j.1538-4632.1996.tb00936.x

Castro-Neto, M., Jeong, Y., Jeong, M.K., Han, L.D., 2009. AADT prediction using support vector regression with data-dependent parameters. Expert Systems with Applications 36, 2979–2986.

Chang Chien, Y.-M., Carver, S., Comber, A., 2020. Using geographically weighted models to explore how crowdsourced landscape perceptions relate to landscape physical characteristics. Landscape and Urban Planning 203, 103904. https://doi.org/10.1016/j.landurbplan.2020.103904

Chatterjee, S., Hadi, A.S., 2015. Regression analysis by example. John Wiley & Sons.

Cui, Y., Yu, Y., Cai, Z., Wang, D., 2022. Optimizing Road Network Density Considering Automobile Traffic Efficiency: Theoretical Approach. Journal of Urban Planning and Development 148, 04021062. https://doi.org/10.1061/(ASCE)UP.1943-5444.0000780

Cutler, D.R., Edwards Jr., T.C., Beard, K.H., Cutler, A., Hess, K.T., Gibson, J., Lawler, J.J., 2007. Random Forests for Classification in Ecology. Ecology 88, 2783–2792. https://doi.org/10.1890/07-0539.1

Das, S., 2021. Traffic volume prediction on low-volume roadways: a Cubist approach. Transportation Planning and Technology 44, pp-93-110. https://doi.org/10.1080/03081060.2020.1851452

Das, S., Tsapakis, I., 2020. Interpretable Machine Learning Approach in Estimating Traffic Volume on Low-volume Roadways. International Journal of Transportation Science and Technology 9, pp-76-88. https://doi.org/10.1016/j.ijtst.2019.09.004

Davis, G.A., Guan, Y., 1996. Bayesian Assignment of Coverage Count Locations to Factor Groups and Estimation of Mean Daily Traffic. Transportation Research Record 1542, 30–37. https://doi.org/10.1177/0361198196154200105

Desai, H., Cunagin, W., Cunagin, K., Hoyt, D., Reel Jr, R.L., Bentz, S., 2014. Application of Seasonal Adjustment Factors to Subsequent Year Data. Transportation Research Record 2443, 143–147.

Drusch, R.L., 1966a. Estimating Annual Average Daily Traffic from Short-Term Traffic Counts.

Drusch, R.L., 1966b. Estimating annual average daily traffic from short-term traffic counts. Highway Research Record.

Eom, J.K., Park, M.S., Heo, T.-Y., Huntsinger, L.F., 2006. Improving the prediction of annual average daily traffic for nonfreeway facilities by applying a spatial statistical method. Transportation research record 1968, 20–29.

Erhunmwunsee, P.O., 1991. Estimating average annual daily traffic flow from short period counts. ITE journal 61.

Ewing, R., Greenwald, M., Zhang, M., Walters, J., Feldman, M., Cervero, R., Frank, L., Thomas, J., 2011. Traffic Generated by Mixed-Use Developments—Six-Region Study Using Consistent Built Environmental Measures. Journal of Urban Planning and Development 137, 248–261. https://doi.org/10.1061/(ASCE)UP.1943-5444.0000068

FHWA, 2022. Traffic monitoring guide. U.S. Department of Transportation.

FHWA, 2016. Highway Performance Monitoring System Field Manual.

Flaherty, J., 1993. Cluster analysis of Arizona automatic traffic recorder data. Transportation Research Record 1410, 93–99.

Fotheringham, A.S., Brunsdon, C., Charlton, M., 2002. Geographically Weighted Regression: The Analysis of Spatially Varying Relationships. Wiley, Chichester, UK.

Fouedjio, F., 2020. Exact Conditioning of Regression Random Forest for Spatial Prediction. Artificial Intelligence in Geosciences 1, 11–23. https://doi.org/10.1016/j.aiig.2021.01.001

Fricker, J.D., Xu, C., Jin, L., 2008. Comparison of annual average daily traffic estimates: Traditional factor, statistical, artificial neural network, and fuzzy basis neural network approach.

Garber, N.J., Bayat-Mokhtari, F., 1986. A Computerized Highway Link Classification System for Traffic Volume Counts.

Gastaldi, M., Rossi, R., Gecchele, G., Della Lucia, L., 2013. Annual average daily traffic estimation from seasonal traffic counts. Procedia-Social and Behavioral Sciences 87, 279–291.

Gecchele, G., Rossi, R., Gastaldi, M., Caprini, A., 2011. Data mining methods for traffic monitoring data analysis: A case study. Procedia-Social and Behavioral Sciences 20, 455–464.

Georganos, S., Grippa, T., Gadiaga, A.N., Linard, C., Lennert, M., Vanhuysse, S., Mboga, N., Wolff, E., Kalogirou, S., 2021a. Geographical random forests: a spatial extension of the random forest algorithm to address spatial heterogeneity in remote sensing and population modelling. Geocarto International 36, 121–136. https://doi.org/10.1080/10106049.2019.1595177

Georganos, S., Grippa, T., Niang Gadiaga, A., Linard, C., Lennert, M., Vanhuysse, S., Mboga, N., Wolff, E., Kalogirou, S., 2021b. Geographical random forests: a spatial extension of the random forest algorithm to address spatial heterogeneity in remote sensing and population modelling. Geocarto International 36, 121–136. https://doi.org/10.1080/10106049.2019.1595177

Goel, P.K., McCord, M.R., Ruan, S., O’Kelly, M., 2006. Bayesian estimation of statewide OD flows using link volumes estimated from combined information in remotely sensed data and ground counts, in: Applications of Advanced Technology in Transportation. pp. 418–423.

Granato, S., Linn County Regional Planning Commission, 1998. The impact of factoring traffic counts for daily and monthly variation in reducing sample counting error.

Hartgen, D.T., Lemmerman, J.H., 1983. Streamlining collection and processing of traffic count statistics. Transportation research record 928, 11–20.

Hengl, T., Nussbaum, M., Wright, M.N., Heuvelink, G.B.M., Gräler, B., 2018a. Random forest as a generic framework for predictive modeling of spatial and spatio-temporal variables. PeerJ 6, e5518. https://doi.org/10.7717/peerj.5518

Hengl, T., Nussbaum, M., Wright, M.N., Heuvelink, G.B.M., Gräler, B., 2018b. Random forest as a generic framework for predictive modeling of spatial and spatio-temporal variables. PeerJ 6, e5518. https://doi.org/10.7717/peerj.5518

Jessberger, S., Krile, R., Schroeder, J., Todt, F., Feng, J., 2016. Improved Annual Average Daily Traffic Estimation Processes. Transportation Research Record: Journal of the Transportation Research Board pp 103-109. https://doi.org/10.3141/2593-13

Jiang, Z., McCord, M.R., Goel, P.K., 2007. Empirical Validation of Improved AADT Estimates from Image-based Information on Coverage Count Segments.

Jin, L., Fricker, J.D., 2008. Applying K-nearest neighbor algorithm for statewide annual average daily traffic estimates.

Labib, S.M., 2024. Greenness, air pollution, and temperature exposure effects in predicting premature mortality and morbidity: A small-area study using spatial random forest model. Science of The Total Environment 928, 172387. https://doi.org/10.1016/j.scitotenv.2024.172387

Lämmer, S., Gehlsen, B., Helbing, D., 2006. Scaling laws in the spatial structure of urban road networks. Physica A: Statistical Mechanics and its Applications 363, 89–95. https://doi.org/10.1016/j.physa.2006.01.051

Li, M.-T., Zhao, F., Chow, L.-F., 2006. Assignment of seasonal factor categories to urban coverage count stations using a fuzzy decision tree. Journal of Transportation Engineering 132, 654–662.

Lowry, M., 2014. Spatial interpolation of traffic counts based on origin–destination centrality. Journal of Transport Geography 36, 98–105.

Lowry, M., Dixon, M., 2012. GIS tools to estimate average annual daily traffic. National Institute for Advanced Transportation Technology (US).

Lu, C., Yang, R., Yang, S., Zhao, F., 2013. Seasonal factor assignment based on the similarity of hourly traffic patterns and influential variables. Journal of Software 8.

McCord, M.R., Goel, P., 2009. Estimating AADT from combined air photos and ground based data : system design, prototyping, and testing. (No. 002OY01).

McCord, M.R., Yang, Y., Jiang, Z., Coifman, B., Goel, P.K., 2003. Estimating annual average daily traffic from satellite imagery and air photos: Empirical results. Transportation Research Record 1855, 136–142.

Meyer, H., Reudenbach, C., Hengl, T., Katurji, M., Nauss, T., 2018. Improving performance of spatio-temporal machine learning models using forward feature selection and target-oriented validation. Environmental Modelling and Software 101, 1–9. https://doi.org/10.1016/j.envsoft.2017.12.001

Mohamad, D., Sinha, K.C., Kuczek, T., Scholer, C.F., 1998. Annual average daily traffic prediction model for county roads. Transportation research record 1617, 69–77.

Moran, P.A.P., 1950. Notes on Continuous Stochastic Phenomena. Biometrika 37, 17–23. https://doi.org/10.1093/biomet/37.1-2.17

Pan, T., 2008. Assignment of estimated average annual daily traffic volumes on all roads in Florida.

PennDOT Traffic, 2014. Dirt, Gravel, and Low Volume Road Maintenance Program (DGLVRP) Traffic Count Policy. Pennsylvania Department of Transportation.

Ritchie, S.G., 1986. A statistical approach to statewide traffic counting.

Rossi, R., Gastaldi, M., Gecchele, G., Kikuchi, S., 2012. Estimation of Annual Average Daily Truck Traffic Volume. Uncertainty Treatment and Data Collection Requirements. Presented at the Procedia - Social and Behavioral Sciences, Elsevier, p. pp 845-856. https://doi.org/10.1016/j.sbspro.2012.09.800

Schneider, W.H., Tsapakis, I., 2009. Review of traffic monitoring factor groupings and the determination of seasonal adjustment factors for cars and trucks. Ohio. Dept. of Transportation. Office of Research and Development.

Selby, B., Kockelman, K., 2011. Spatial prediction of AADT in unmeasured locations by universal kriging.

Shamo, B., Asa, E., Membah, J., 2015. Linear Spatial Interpolation and Analysis of Annual Average Daily Traffic Data. Journal of Computing in Civil Engineering 29, 04014022. https://doi.org/10.1061/(ASCE)CP.1943-5487.0000281

Sharma, S., Leng, Y., 1994. Seasonal traffic counts for a precise estimation of AADT. ITE Journal 64.

Sharma, S., Lingras, P., Xu, F., Kilburn, P., 2001. Application of Neural Networks to Estimate AADT on Low-Volume Roads. Journal of Transportation Engineering 127, 426–432. https://doi.org/10.1061/(ASCE)0733-947X(2001)127:5(426)

Sharma, S.C., Allipuram, R.R., 1993. Duration and frequency of seasonal traffic counts. Journal of Transportation Engineering 119, 344–359.

Sharma, S.C., Lingras, P., Liu, G.X., Xu, F., 2000. Estimation of Annual Average Daily Traffic on Low-Volume Roads: Factor Approach Versus Neural Networks. Transportation Research Record 1719, 103–111. https://doi.org/10.3141/1719-13

Sharma, S.C., Lingras, P., Xu, F., Liu, G.X., 1999. Neural Networks as Alternative to Traditional Factor Approach of Annual Average Daily Traffic Estimation from Traffic Counts. Transportation Research Record 1660, 24–31. https://doi.org/10.3141/1660-04

Sun, X., Das, S., 2019. Estimating Annual Average Daily Traffic for Low-Volume Roadways: A Case Study in Louisiana. Presented at the Transportation Research Circular, Transportation Research Board, p. pp 190-201.

Sun, X., Das, S., 2015. Developing a Method for Estimating AADT on all Louisiana Roads (Digital/other). Federal Highway Administration.

Talebi, H., Peeters, L.J.M., Otto, A., Tolosana-Delgado, R., 2022a. A Truly Spatial Random Forests Algorithm for Geoscience Data Analysis and Modelling. Math Geosci 54, 1–22. https://doi.org/10.1007/s11004-021-09946-w

Talebi, H., Peeters, L.J.M., Otto, A., Tolosana-Delgado, R., 2022b. A Truly Spatial Random Forests Algorithm for Geoscience Data Analysis and Modelling. Math Geosci 54, 1–22. https://doi.org/10.1007/s11004-021-09946-w

Tsapakis, I., Holik, W., Das, S., Kraus, E., Anderson, P., 2020. Data Collection and Annual Average Daily Traffic (AADT) Estimation for Non-Federal Aid System (NFAS) Roads (No. FHWA-SA-20-064). FHWA, Washington, DC.

Tsapakis, I., Schneider IV, W.H., Nichols, A.P., 2011. Improving the estimation of total and direction-based heavy-duty vehicle annual average daily traffic. Transportation planning and technology 34, 155–166.

Verbavatz, V., Barthelemy, M., 2019. Critical factors for mitigating car traffic in cities. PLoS ONE 14, e0219559. https://doi.org/10.1371/journal.pone.0219559

Wang, S., Yu, D., Kwan, M.-P., Zheng, L., Miao, H., Li, Y., 2020. The impacts of road network density on motor vehicle travel: An empirical study of Chinese cities based on network theory. Transportation Research Part A: Policy and Practice 132, 144–156. https://doi.org/10.1016/j.tra.2019.11.012

Wang, X., Kockelman, K.M., 2009. Forecasting network data: Spatial interpolation of traffic counts from texas data. Transportation Research Record 2105, 100–108.

Wheeler, D., Tiefelsdorf, M., 2005. Multicollinearity and correlation among local regression coefficients in geographically weighted regression. Journal of Geographical Systems 7, 161–187.

Wright, M.N., Ziegler, A., 2017. ranger: A Fast Implementation of Random Forests for High Dimensional Data in C++ and R. J. Stat. Soft. 77. https://doi.org/10.18637/jss.v077.i01

Wu, H., Zhang, Z., 2009. Framework for estimating AADT using coclustering-based collaborative filtering.

Xia, Q., Zhao, F., Chen, Z., Shen, L.D., Ospina, D., 1999. Estimation of annual average daily traffic for nonstate roads in a Florida county. Transportation research record 1660, 32–40.

Yang, B., Wang, S.-G., Bao, Y., 2011. Efficient local AADT estimation via SCAD variable selection based on regression models. IEEE, pp. 1898–1902.

Zhao, F., Chung, S., 2001a. Contributing Factors of Annual Average Daily Traffic in a Florida County: Exploration with Geographic Information System and Regression Models. Transportation Research Record 1769, 113–122. https://doi.org/10.3141/1769-14

Zhao, F., Chung, S., 2001b. Contributing factors of annual average daily traffic in a Florida county: exploration with geographic information system and regression models. Transportation research record 1769, 113–122.

Zhao, F., Li, M.-T., Chow, L.-F., 2004. Alternatives for estimating seasonal factors on rural and urban roads in Florida.

Zhao, F., Park, N., 2004a. Using geographically weighted regression models to estimate annual average daily traffic. Transportation research record 1879, 99–107.

Zhao, F., Park, N., 2004b. Using geographically weighted regression models to estimate annual average daily traffic. Transportation research record 1879, 99–107. https://doi.org/10.3141/1879-12

Zhong, M., Bagheri, E., Christie, J., 2012. Improving group assignment and AADT estimation accuracy of short-term traffic counts using historical seasonal patterns & Bayesian statistics. Procedia-Social and Behavioral Sciences 43, 607–617.