MODELING ANNUAL AVERAGE DAILY TRAFFIC FOR LOCAL FUNCTIONALLY CLASSIFIED ROADS

by

Sonu Mathew

A dissertation submitted to the faculty of
The University of North Carolina at Charlotte
in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in
Infrastructure and Environmental Systems

Charlotte

2020

Approved by:

____
Dr. Srinivas S. Pulugurtha

____
Dr. Martin R. Kane

____
Dr. Rajaram Janardhanam

____
Dr. Shen-En Chen

____
Dr. William Graves

ABSTRACT

SONU MATHEW. Modeling Spatial Prediction of Annual Average Daily Traffic for Local Functionally Classified Roads. (UNDER THE GUIDANCE OF DR. SRINIVAS S. PULUGURTHA)

The rapid increase in population, the growth in demand for travel, and the subsequent traffic congestion and road safety challenges call for better utilization of existing road infrastructure. A federally funded state-administered program known as Highway Safety Implementation Program (HSIP) was instituted for state agencies to adopt a data-driven and performance-based approach to improving safety on public roads. One of the requirements of HSIP is for state agencies to report annual average daily traffic (AADT) for all functionally classified major, minor, and local roads.

A considerable amount of resources are spent by various transportation departments to estimate AADT on major, minor, and local road links. The available AADT data are based on traffic counts collected at selected locations on these roads. However, time, money and other resource constraints limit agencies from estimating AADT for all the roads in the transportation network. The count-based AADT is available for all major and minor road links, but available for a relatively fewer number of local road links.

The objectives of this research are: 1) to review AADT estimation methods for functionally classified major and local roads, 2) to examine the influence of road network, socioeconomic, demographic, and land use characteristics on local roads AADT, 3) to develop sustainable and repeatable methods to estimate AADT on local functionally classified roads, and 4) to validate and calibrate the models to improve their predictability.

To achieve the aforementioned objectives, this research examined five different modeling approaches to estimate AADT for all local roads. They include traditional ordinary least square (OLS) regression, geographically weighted regression (GWR), and geospatial interpolation techniques such as Kriging, inverse distance weighted (IDW) interpolation, and natural neighbor interpolation. The available count-based AADT data at 12,899 traffic count locations on local roads in North Carolina during the years 2014, 2015, and 2016 was used as the dependent variable when developing the models. The road, socioeconomic, demographic, and land use characteristics for the year 2015 were considered as the explanatory variables. The explanatory variables were screened to minimize multicollinearity by computing and comparing the Pearson correlation coefficients.

The model development was carried out in two levels: the statewide AADT estimation and county-level AADT estimation. The speed limit, road density, distance to the nearest nonlocal road, the count-based AADT at the nearest nonlocal road, and population density are significant explanatory variables used to develop the statewide models. The validation results indicated that the GWR model performed relatively better when compared to other considered statistical and geospatial methods. GWR can accommodate the spatial variations in AADT data, by geographic location, when estimating the local road AADT. The errors in estimated local road AADT are lower for locations with a higher number of nearby traffic count stations.

Ten counties were considered for county-level analysis and modeling. The quality of land use data, population density, road density, and the number of local road traffic count stations available in the county were used in the selection process. The county-level models were observed to estimate local road AADT relatively better than the statewide models. The inclusion of land use variables for modeling can be mainly attributed to the improved performance of county-level models. The developed county-level models were used for estimating AADT at non-covered locations in each selected county.

The median prediction errors associated with statewide and county-level models were compared and assessed to recommend future sampling requirements to improve the model predictability. The median prediction errors are higher for urban local roads and for local roads with a speed limit greater than 25 mph and less than 50 mph. In most of the cases, the median prediction error seems to depend on the number of available local road traffic count stations and county characteristics. These findings indicate that count-based local road AADT data from spatially distributed traffic count stations in North Carolina can improve the predictability of models.

The prediction errors were also low at local road traffic count stations near single-family residential units, multi-family residential units, and the commercial area. Contrarily, they are relatively higher at local road traffic count stations near schools, institutions, government, office, and industrial land uses. This could be attributed to differences in the number of local road traffic count stations by land use area type (more the number of local road traffic count stations, lower the prediction error).

Samples sizes were estimated based on the coefficient of variation in the available count-based local road AADT data and the number of local road links by the speed limit and link connectivity for each county at a 70% confidence level. A 15% prediction error rate was considered acceptable for local roads and used to estimate the sample sizes. A sampling plan based on the number of local road locations, functional classification type, speed limit ranges, and road connectivity type like dead-ends is recommended. To expand the local road traffic data collection program and estimate spatially distributed count-based local road AADT, sample data must be collected at around 12,000 (based on the speed limit) to 22,000 (based on the link connectivity type) different stations in North Carolina biennially. The simple random sampling criterion can be used when selecting locations based on the speed limit and link connectivity, in a county, while ensuring that they are geographically distributed in the county.

This research proposes the use of county-level growth factors based on available count-based local road AADT for future AADT estimations. The count-based local road AADT and growth factor for the reporting year, for the county in which the local road is located, must be used if the count-based AADT was available for the previous year(s). For non-covered locations, the estimated AADT for the base year (2015 in this research) and growth factors from the base year to the reporting year must be used.

It is recommended to update the base year local road AADT estimation model to 2020 once the statewide travel demand model is updated or census 2020 data (block-level) is available. Overall, the application of the proposed AADT estimation method and growth factors minimize the costs associated with lapses in traffic count data collection programs and plans. The estimated or actual AADT for each local road link can be used to compute the VMT for each local road link. The findings from this research can be used to proactively identify solutions and plan, design, build, and maintain the local roads.

DEDICATION

to My Family

ACKNOWLEDGEMENTS

I would like to express my sincere gratitude to my advisor Dr. Srinivas S. Pulugurtha, for the constant support, the guidance he has provided throughout the Ph.D. program. I will be eternally grateful to him for teaching me the importance of being earnest and doing work with utmost honesty. His encouragement and personal guidance have given me the greatest learning curve I could imagine as a person. Thank you, Dr. Srinivas, for providing me an opportunity to pursue a Ph.D. under your guidance.

I would like to thank my committee members: Dr. Rajaram Janardhanam, Dr. Martin Kane, Dr. Shen-En Chen, and Dr. William Graves for their encouragement, insightful questions, and suggestions to improve my dissertation.

I would like to thank the North Carolina Department of Transportation (NCDOT) for providing financial support for this dissertation. Also, I would like to express my special thanks to Mr. Kent L. Taylor, State Traffic Survey Engineer, NCDOT, for providing excellent support, guidance, and valuable inputs for the successful completion of this dissertation.

A special word of thanks to Mr. Chaithanya Bhure and Mr. Chandan Mannem for their help in processing the data for this research. I thank my fellow student researchers: Sarvani Duvvuri, Raghu Gouribhatla, Raunak Mishra, Swapneel Kodupuganti, and Sravya Jayanthi for their support and help throughout the Ph.D. program.

I would like to thank my master's advisors: Dr. Ashish Dhamanaiya, Dr. Shriniwas Arkatakar, and Dr. G. J. Joshi from SVNIT Surat, India, for expanding my horizon towards research. A big thanks also to my friends who have always been the most important source of support throughout this journey: Ritika Mishra, Srijith Balakrishnan, Ajinkya Mane, Vasudev Narayanan, Vinayaraj V.S., and Nipjyoti Bhardwaj.

A very special word of thanks for my family, my father (Mathew P.C.), mother (Laila George), brother (Manu Mathew), sister (Merlin Mathew), sister-in-law (Sharon Rose), and brother-in-law (Albert John), for always being there with me throughout this journey.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

AADT	Annual average daily traffic
ANN	Artificial neural networks
ATR	Automatic traffic recorder
GWR	Geographically weighted regression
HSIP	Highway Safety Implementation Program
IDW	Inverse distance weighting
NCDOT	North Carolina Department of Transportation
OLS	Ordinary least square
TAZ	Traffic analysis zones
VMT	Vehicle miles traveled
WLS	Weighted least squares

CHAPTER 1: INTRODUCTION

1.1 Motivation and Background

Rapid growth in population over the past two decades has led to an increase in travel demand, resulting in congestion, safety, and environmental issues. As traffic increases with growth in population, the conflicts that arise because of human interaction, off- and on-network characteristics, and other associated factors also increase. Understanding the causes of crashes, identifying appropriate solutions, and proactively adopting or implementing countermeasures helps improve traffic safety. Federal agencies have made reducing crashes a top priority by considering safety every time and at every stage of a project. For this purpose, a federally funded state-administered program known as the Highway Safety Implementation Program (HSIP) has been instituted. The goal of HSIP is to achieve a significant reduction in fatalities and serious injuries on public roads (Gross, 2017). One of the requirements of HSIP for state agencies is to report annual average daily traffic (AADT) on all paved public roads (FHWA, 2018) and develop safety performance measures. The AADT also helps estimate vehicle miles traveled (VMT) at state-, area-, and link-level (route-level). The accurate estimation of the AADT is a pivotal point as it is a central factor for the performance evaluation and the planning process of various transportation projects.

Field data are collected by agencies based on need or as a part of traffic count programs. The Traffic Survey Group of the North Carolina Department of Transportation

(NCDOT) currently counts traffic on all functionally classified roads. They cover a major portion of functionally classified major roads, but only a small portion of functionally classified local roads. A comprehensive traffic volume data collection is not economical in the case of local roads, even though they constitute a major proportion of the roads in the state. AADT must be estimated at these locations, which also helps estimate AADT for the coming years, but resource constraints limit various DOTs from expanding their traffic count data collection and monitoring efforts.

Many researchers have broadly explored estimating the AADT in urban/local areas using various statistical methods, time series modeling methods, and density-based/gravity-based geospatial methods. The estimations for unknown locations from past research are established based on the available count-based AADT data and incorporating additional explanatory variables related to road characteristics and socioeconomic attributes of the study area. Moreover, most of the current research methods help estimate AADT for functionally classified major road links due to the availability of traffic counts for these roads (either AADT or Annual Daily Traffic, ADT). The efforts to estimate AADT for local functionally classified paved roads open to the public have been very limited in the present research. Even the regional travel demand forecasting models typically ignore local roads. Hence, there is a need to develop methods to estimate AADT for local roads.

Several factors influence the predictability of AADT on local roads. Considering the sample counts from an area along with road characteristics and socioeconomic characteristics, a few researchers estimated AADT on local roads. Most of these researchers ignored the local travel characteristics and development density related indicators in their predictions. As local roads are designed for land access, most daily travel is oriented from the land being accessed to the nearest higher functionally classified roads. Knowing the characteristics of land use in the vicinity of local roads is, therefore, important for the accurate estimation of AADT.

The goal of this research is to estimate AADT for local roads. The research findings will minimize the cost associated with traffic count data collection programs. Also, it will assist with the computation of safety performance functions, resource allocation, and prioritization of infrastructure projects for future improvements.

1.2 Problem Statement

The recent Fixing America's Surface Transportation (FAST) Act legislation requires states to generate a database containing AADT for all paved roads open to the public. Reliable estimation of AADT is central to road improvement and funding prioritization, safety performance assessment, and calibrating/validating travel demand forecasting models. A significant amount of time and money is spent to collect traffic counts and estimate AADT on a major portion of functionally classified major roads, but only a small portion of local functionally classified road links.

The local roads constitute most of the road network in a state. The state of North Carolina has about 77,000 miles of local functionally classified roadways. As traffic volumes on local roads are low compared to other functionally classified roads, collecting traffic counts at all local roads in a state is not economically feasible. With the increased emphasis on HSIP, AADT is a necessary variable for safety performance evaluation. Considering resource constraints, there is a need to collect surrogate data and/or develop methods/models to estimate AADT of local functionally classified roads.

Most of the current research methods help estimate AADT for functionally classified major roads due to the availability of traffic counts for these roads (either AADT or Annual Daily Traffic, ADT). Very few researchers have worked on estimating AADT on local functionally classified roads. A few researchers in the past explored statistical methods and machine learning approaches to estimate AADT. Although statistical models are relatively easier and provide quick estimates of AADT, these models generally provide results on a global level. However, the characteristics of a road segment and demographics of that area may vary over space or vary at a local level. Hence, it is envisaged that models accounting the spatial variability in dependent and independent variables may give reliable estimates of AADT in the study area. The estimates not only help planners develop safety performance measures and compute local road VMT but also assist to plan, propose, and prioritize infrastructure projects for future improvements and in air quality estimates.

1.3 Research Objectives

The objectives of this research, therefore, are:

1) to review AADT estimation methods for functionally classified major and local roads,
2) to examine the influence of road network, socioeconomic, demographic, and land use characteristics on local roads AADT,
3) to develop sustainable and repeatable methods to estimate AADT for local functionally classified roads, and,
4) to validate and calibrate the models to improve their predictability.

1.4 Organization of the Report

The remainder of this report is comprised of ten chapters. A review of existing literature on different methods to estimate AADT on local roads and how selected state DOTs are evaluating the AADT of local roads are discussed in Chapter 2. Chapter 3 illustrates the data collection and processing involved estimating the AADT on local roads. Chapter 4 outlines the methodological framework adopted for this research. Chapter 5 covers the descriptive analysis of count-based AADT data. Statewide model AADT estimation results are discussed in Chapter 6, while the county-level model AADT estimation results are presented in Chapter 7. Chapter 8 details the modeling errors and sampling requirements to improve accuracy. The model accuracy assessment based on count-based AADT range is illustrated in Chapter 9. Conclusions from this research study and scope for future research are presented in Chapter 10.

CHAPTER 2: LITERATURE REVIEW

The functional classification of roads is mainly intended at determining the role of the road in serving the mobility and accessibility needs of the people and goods. It defines the function of the road before designing their width, speed limit, intersection control, and other design features. In other words, the mobility need is explained in terms of various elements such as the operating speed, the level of service, and the riding comfort. Accessibility is measured in terms of access to various land use activities. The functional classification of roads based on their hierarchy as per FHWA guidelines are the interstate system, other arterials, collectors, and local streets. As the focus of this research is to estimate AADT and VMT of local functionally classified roads, the classification concepts, criteria, and procedures for this category are summarized next.

As per the FHWA guidelines, the roads that provide access to the residential areas, businesses, farms, or other abutting property are classified as local roads (FHWA, 2013). In most of the cases, local roads connect to other local streets and collectors. The local functionally classified roads are further classified into the urban and rural local roads. Also, local roads do not carry any through traffic movement. As per the NCDOT guidelines, local roads are designed specifically to provide better accessibility and to connect to the collector and arterial roads (NCDOT, 2014). It consists of all the roads which are not defined as arterials or collectors. A review of past literature on estimating AADT is presented in this Chapter

2.1 AADT Estimation Methodologies

The researchers in the past have developed various methods and models to estimate AADT when count data from the field are not available for a road link. These include statistical models based on area type such as urban and rural (Mohamad et al., 1998; Xia et al., 1999; Seaver et al., 2000; Smith et al., 2002), time series methods (Xia et al., 1999; Zhao and Chung, 2001; Tang et al., 2003; Fricker et al., 2008), and density-based and gravity-based geospatial methods (Wang and Kockelman, 2009; Selby and Kockelman, 2011; Pulugurtha and Kusam, 2012; Duddu and Pulugurtha, 2013; Kusam and Pulugurtha, 2015). On the other hand, literature also documented the application of Geographic Weighted Regression (GWR) (Selby and Kockelman, 2011), Kriging (Selby and Kockelman, 2011), Inverse Distance Weighting (IDW), Natural Neighbor and trends techniques, considering traffic counts within the vicinity, to estimate the AADT. A brief overview of the state-of-the-art AADT estimation methods is summarized in different sections: statistical methods, geospatial methods, artificial neural network, and other methods. This task was followed by a comparison of different methods to estimate AADT.

2.1.1 Statistical methods

The general Ordinary Least Squares (OLS) regression models are widely adopted to model the relationship between a dependent variable and the explanatory variables. The general form of an OLS regression model is shown in Equation (1).

$$Y_i = \beta_1 + \beta_2 X_2 + \cdots \beta_n \beta_n + \varepsilon \quad (1)$$

where Y_i is the dependent variable; X_1, X_2, … X_n are the explanatory variables; β_1, β_2…. β_n, are the coefficients; and ‘ε’ is the residual error.

Neveu (1983) introduced a quick-response method to estimate traffic volume on rural state highway systems in New York. They used an elasticity-based formulation to estimate future year traffic volume as a function of present year traffic volume and influenced by various demographic factors. The accuracy of the estimated traffic volume is highly depended on the accuracy of the input variable. The applicability of this model to other areas and the assumption of constant elasticities over time are the major limitations of this research.

Saha and Fricker (1986) proposed aggregate- and disaggregate-level models to estimate AADT on rural locations of Indiana state road networks. In their study, state- and national-level demographic and economic variables were used for the estimation. It can be considered as a basis for many other studies in rural road AADT estimation. Xia et al. (1999) proposed a multiple regression model for the prediction of AADT on non-state roads in the urbanized areas in Florida. They employed Geographic Information Systems (GIS) to aggregate various data elements and quantify the spatial effect (buffer width, 0.25 miles to 3 miles) of various parameters like population, employment, and accessibility on non-state road traffic generation. The findings from their research depict that road characteristics like the number of lanes, functional classification, and area type were the potential regressors in the developed model, whereas socioeconomic factors were insignificant. This research benefited from comprehensive statistical measures to address the general problems associated with linear models like multicollinearity.

Seaver et al. (2000) estimated traffic volume on the rural roads based on the road type with data from 80 counties in Georgia using statistical methods. Several regression equations were developed based on the 45 different characteristics for estimating ADT.

Zhao and Chung (2001) modified the model developed by Xia et al. (1999) using a larger dataset, including all the AADTs for state roads in Florida. They performed extensive spatial analyses to derive land use (employment) and accessibility (direct access to expressway) measurements for the new multiple regression models. They incorporated the effect of regional economic activity on the traffic on a road in the model development process. However, findings from their research are not transferrable to other locations because details of the urban form are involved in the modeling process.

Li et al. (2004) identified various factors affecting the seasonal variations in traffic patterns using regression analysis. The causes of these repetitive patterns in traffic were studied by considering land use, demographic, and socioeconomic variables which also contains resident's and tourist's inflow and outflow during various seasons, retail and employment characteristics of the study area, etc. They illustrated the direct estimation of the seasonal factors for short-period traffic counts based on land use, demographic, and socioeconomic variables. Finally, the generated seasonal groups were assigned to the short-term traffic counts based on the similarity in land use, demographic, and socioeconomic characteristics of the study area.

Goel et al. (2005) proposed a method to improve the estimation of AADT on highway links from coverage counts (24 hours of continuous count). The Monte Carlo simulation was employed to compare the performance of correlation-based methodology (which is compatible with the generalized least squares estimation) with the traditional approach (OLS estimation). The results from their study showed that when there is a high correlation between AADT counts, the predictive accuracy of the correlation-based method was better over a conventional approach. The lower correlation between the volumes of the section, however, led to similar estimates for both the methods.

Apronti et al. (2016) developed regression models for estimating ADT of low volume roads in Wyoming based on socioeconomic, demographic, and geometric variables such as road width, surface type, land use, access to highway, census population, and tax revenue. They compared the linear regression model with the logistic regression approach. The predictive accuracy of logistic regression models (the probability of a road belonging to the predefined AADT threshold) was good compared to linear models.

Staats (2016) developed a non-linear regression model to estimate AADT on local roads in the state of Kentucky. Three different models were developed based on geographical and socioeconomic variability across the state. The explanatory variables considered for each model include probe count, residential vehicle registration, and curve rating.

Jayasinghe and Sano (2017) incorporated a two-way approach to estimate the AADT on roads in metropolitan areas. Their proposed methodology uses “multiple centrality” and “weighted link cost” to estimate the AADT at the link level. This method helps to capture road type variables with global and metric distances.

Raja et al. (2018) conducted a study on the estimation of AADT on low-volume roads by developing a regression model using the existing AADT values, socioeconomic data, and location data. OLS regression models were developed using 70% of the available data. They also considered and explored the applicability of quadratic and logarithmic transformations. The validation of the model was conducted using the value of the Nash-Sutcliffe coefficient. The validation results indicated that the linear and quadratic models performed at the same level while the logarithmic model generated a lower value of the coefficient than the other two. They concluded by suggesting the use of a linear or quadratic model for the estimation of AADT on low-volume roads.

2.1.2 Geospatial methods

GWR was first proposed in 1996 (Brunsdon et al., 1996). It is an extension of the traditional regression framework that can spatially estimate the regression coefficients which will be centered on a point in the dataset. The general form of the GWR model is shown in Equation (2).

$$Y_i = \beta_o(u_i v_i) + \sum_{k=1}^{p} \beta_k(u_i v_i) X_{ik} + \varepsilon_i \quad (2)$$

where 'i' denotes the location in which the coefficients are estimated. Y_i is the dependent variable, X_{ik} is the k^{th} explanatory variable, (u_i, v_i) indicates the regression parameters of the k^{th} explanatory variable, finally, $\mathcal{E}_i$ is the residual error for the i^{th} spatial location.

Zhao and Park (2004) have employed the GWR method to estimate the AADT in Broward County, Florida. One OLS model and two GWR models were developed and compared in their research. The explanatory variables such as the number of lanes, accessibility to employment, population, and employment within the vicinity of a count station, and direct access to expressways were considered in the modeling process. Like the study conducted by Xia et al. (1999), a limited number of variables were explored in their study. It was also noted that the choice of weighting function plays a pivotal role in the GWR model performance (Zhao and Park, 2004).

Du and Mulley (2006) studied the applicability of the GWR model to examine the relationship between transportation accessibility and land value. They concluded that GWR provides a better understanding of spatially varying relationships like land value and transportation accessibility. Chow et al. (2006) explored the spatial variability in the relationship between public transit use for a home-based work trip and potential transit use predictors using the GWR. The results from their research indicate that the applicability of GWR models is better than the OLS models.

Gadda et al. (2007) examined the uncertainties associated with the AADT estimates from short-duration traffic counts in a spatiotemporal perspective. They quantified the changes in factoring errors, spatial errors, and temporal errors by day-of-the-week, month-of-the-year, functional class, the number of lanes, and duration and distance to nearest SPTC station. Their results indicated that the spatial errors increase drastically beyond 5 miles from the traffic count stations in the urban areas, and 1 mile in the rural areas.

Yang et al. (2017) used GWR models to estimate the possible interaction between active mode of travel demands (walking trips) and ambient built-environment attributes such as population density, transit accessibility, characteristics of the intersection, and the road network. Their results explicitly pointed out the higher predictive accuracy of the GWR model over the OLS model.

Recent research initiatives also explored the Kriging method that is based on the spatial interpolation of observations. This technique consists of the estimation of the parameters by calculating the "weighted average" of the available data and use it to estimate the unknown values (Selby and Kockelman, 2013). Kriging considers the surrounding measured location values to estimate the unmeasured location. The general form of the Kriging is shown in Equation (3).

$$Z(S_o) = \sum_{i=1}^{N} W_i Z(S_i) \qquad (3)$$

where $Z(S_i)$ is the count-based AADT at the location "i" and W_i is the unknown weight for the count-based AADT at the i^{th} location, S_o is the prediction location, and N is the number of traffic count stations.

Wang and Kockelman (2009) analyzed the prediction of AADT at non-covered locations using the traffic count data over seven years in Texas and the Kriging method. Using the temporal extrapolation, the counts were estimated followed by a spatial interpolation to the non-covered locations. Eighty percent of the data was used for the analysis, and the rest was used for the validation. The median of the errors was 33%, which seems to be reasonable. The results indicate that the Kriging method can be used for the estimation of traffic conditions at unmeasured locations.

Similarly, Selby and Kockelman (2011) estimated ADT in Texas through the application of Euclidean distance and network distance-based Kriging methods. Even though universal Kriging was found to perform better than the non-spatial regression techniques, errors are observed to be higher at locations with a few traffic counts and/or in less measurement-dense areas.

Selby and Kockelman (2013) explored the spatial estimation of AADT in Texas using two methods: GWR and universal Kriging. The model inputs included the existing counts, the highway data, and other parameters such as the demographic and employment data. Universal Kriging model parameters were obtained using the weighted least squares (WLS) regression, and the corresponding model was divided into two parts: local trend and spatial function to compute the error terms. The data-generation process was termed "stationary" due to the dependence of the model on the location's distances but not on its absolute position in the space. Both Euclidean distances and the network distances were considered for the prediction. On the other hand, the GWR also used WLS regression for estimation, but the GWR is “mathematically simpler” than the Kriging. The results indicate that the universal Kriging yielded better estimates (in terms of errors) than GWR. The errors were relatively lower in areas with high count values. The county-level employment density parameter did not have much effect on the estimation of the AADT. On the other hand, parameters such as the road type, the speed limit, the number of lanes, and the population had a significant impact on AADT.

Pulugurtha and Kusam (2012) extracted off-network characteristics, such as demographic, socioeconomic, and land use characteristics, over multiple buffer bandwidths around a road link to estimate AADT on functionally classified roads. The effect of an explanatory variable on the AADT of a link decreases with an increase in the distance from the subject link (Duddu and Pulugurtha, 2013). Spatial variations in the variables such as land use characteristics, on- and off-network characteristics, etc. play a major role in the AADT estimation process. The buffer width to capture data was observed to vary by the functional class; smaller buffer widths would help capture data to generate more meaningful outputs for lower functional class roads (Kusam and Pulugurtha, 2015). Further, the neighboring link characteristics (upstream and downstream) observed to influence the AADT on the subject link.

The Southeast Michigan Council of Governments (SEMCOG) developed an algorithm to estimate AADT at non-covered locations in a GIS environment. The data obtained from the local agencies were used to estimate AADT on roads with unknown traffic volume as a weighted average of AADT on surrounding road links (Holik et al., 2017). Similarly, the Virginia Department of Transportation (VDOT) adopted the trip generation method to estimate AADT at the link level for local roads. Google aerial images were used to determine construction activities and network connectivity, length of the network, etc. for assigning the number of trips generated to estimate AADT (Tsapakis et al., 2017).

2.1.3 Artificial neural networks and other Machine learning

Machine learning has received constant attention in the field of transportation engineering over the past few decades. Among different computational algorithms, ANN has been widely employed in studying traffic forecasting and traffic pattern analysis. Later, supervised learning methods like the support vector machine learning approach were adopted by various researchers (Chowdhury et al., 2006; Castro-Neto et al., 2009; Ma et al., 2012).

Sharma et al. (1999) used 48-hour coverage counts in Minnesota to estimate AADT using the artificial neural network method. A traditional method using data from automatic traffic recorder (ATR)-equipped links was also incorporated for comparison of performance. Their results from comparison indicate that when single 48-hour coverage counts are correctly assigned to a factor group, the traditional method is observed to produce better AADT estimates than the neural network approach. Sharma et al. (2001) extended the neural network approach to estimate AADT on low-volume roads. They applied the ANN to compute the AADT of low-volume roads from the existing volumes of short period counts. Their results indicated that 48-hour duration counts are preferable to the 24-hour or 72-hour duration counts.

Zhong et al. (2004) employed genetically designed neural network models and regression models, factor models, and time series models to estimate the missing traffic count data from the permanent traffic counters. The results from their research indicated the predictive accuracy of genetically designed regression models over the other models mentioned above. In a before-after comparison (data from before and after the failure of permanent counters), average errors were reported to be insignificant in the case of genetically modified regression models. Sun and Das (2015) developed an AADT estimation methodology for rural non-state roads in Louisiana. Statistical and pattern recognition methods were explored to estimate the AADT on such roads. Their findings indicate that the predictability of support vector regression (SVR) models is better than count-based models such as Poisson and Negative Binomial models in the AADT estimation for low-volume roads. Sabla (2016) employed ANN and support vector regression models to estimate AADT on different road functional classes in South Carolina. They illustrated the advantages of SVR models over traditional linear models in estimating AADT.

Das and Tsapakis (2019) employed the support vector machine learning approach in estimating AADT on local roads. According to their findings, the population density and the work area characteristics density are the best predictors in estimating AADT. The accuracy of the machine learning model was also found to be better than traditional linear models. Finally, they proposed the top five decision rules to improve the predictive accuracy of the developed model.

2.1.4 Other methods

A few researchers proposed a means to estimate AADT based on contemporary ground images (McCord et al., 2003; Jiang et al., 2006). They suggested converting hourly volume to daily volume using hourly factors. Further, daily volume was converted to

AADT using seasonal factors. In addition to the ground image, McCord et al. (2009) combined the aerial image information with the information available in the traffic count database, and the combination of aerial information and ground database improved the accuracy of the AADT estimation.

Wang et al. (2013) conducted a parcel level travel demand analysis to estimate the AADT on roads in Broward County, Florida. Their developed model consisted of four steps: network modeling, parcel-level trip generation, parcel-level trip distribution, and parcel-level trip-assignment. The gravity model was used for trip generation, and the all-or-nothing assignment was used in the trip assignment process for the local roads with a value of AADT lesser than 30,000 vehicles per day. The developed model was compared with the regression model. The results implied that the regression model tends to over-estimate the AADT. Using the Mean Absolute Percentage Error (MAPE), the model was validated, and the proposed method seems to have a lower estimation error.

Lingras et al. (2000) applied time series analysis based on different types of road groups for predicting daily traffic volumes. Both statistical and neural network models were developed for predicting daily traffic volumes for comparison purposes. Neural network models are observed to outperform autoregressive models with higher prediction errors.

2.2 Comparison of Methods to Estimate AADT

Smith et al. (1997) developed four models including historical average, time-series, neural network, and nonparametric regression models to estimate freeway traffic flow that represents 15-minute future traffic volume on the Northern Virginia Capital Beltway. From the Wilcoxon signed-rank test conducted, they revealed that the nonparametric models are easy to implement, proved to be portable, and experienced significantly lower errors than other considered models.

Smith et al. (2002) compared the performance of parametric and nonparametric regression models using the seasonal autoregressive integrated moving average (ARIMA) for traffic flow forecasting. The findings from their research indicate a characteristically stochastic nature of traffic condition data as opposed to chaotic.

Zhao and Park (2004) compared the predictability of the OLS model and the GWR models in the AADT estimation process. They concluded that GWR models perform better than the OLS model, due to their inherent capability to account for the variability in data. Similarly, Eom et al. (2006) considered spatial dependency for the estimation of AADT of non-freeway roads. The study was carried out with three data elements: AADT, road characteristics, and census information. For the analysis, AADT for the year 1999 was used and models were developed for the Raleigh, North Carolina and Wake County, North Carolina. Their results showed that Kriging performed better than the OLS regression method for Wake County, North Carolina while the OLS regression method performed better for Raleigh, North Carolina.

Tang et al. (2003) conducted a study comparing four modeling techniques for estimating AADT. The four models were time series, nonparametric regression, neural network, and Gaussian maximum likelihood. The results from their research indicate that nonparametric regression and Gaussian maximum likelihood yielded lower errors than the other two methods. It was concluded in their study that the Gaussian maximum likelihood model is applicable compared to the other models.

Lam et al. (2006) applied a nonparametric regression model and the Gaussian maximum likelihood model for short-term traffic volume forecasting. Historical traffic data collected for the annual traffic census in Hong Kong was used for the modeling process. Their study results and comparison favored the use of a nonparametric regression model over the Gaussian maximum likelihood model for traffic volume forecasting.

Duddu and Pulugurtha (2015) worked on estimating the AADT as a function of land use characteristics extracted using the principle of demographic gravitation. According to the principle, the effect of a variable on the AADT of a link decreases with an increase in the distance from the subject link. Mathematical and computational models based on learning algorithms were developed to estimate the AADT and were compared for performance evaluation. The proposed methodology helps estimate the AADT with improved performance compared to traditional methods and does not require data from the ATRs. Their findings indicate that the ANN models have better predictive capability compared to the statistical models.

Selby and Kockelman (2016) performed a comparative assessment between spatial interpolation methods (Universal Kriging and GWR methodology) and the OLS method for the prediction of traffic levels at non-covered locations in Texas. Like previous findings, the performance of the spatial regression methods surpassed the OLS method.

2.3 AADT Estimation Methods by DOTs

Various online reports and resources were reviewed to identify notable practices followed by various DOTs in estimating AADT and VMT. Most DOTs estimate missing AADT counts using methods set out in FHWA's Traffic Monitoring Guide (FHWA, 2016). An online survey was conducted to gather information on how selected other state DOTs are estimating AADT on local roads. Notable research initiatives conducted by six states are summarized in the following subsections.

2.3.1 Kansas

Kansas DOT (KDOT) collects a sample of traffic counts on roads that are functionally classified as local. The local roads are further divided into three categories: urban, county, and small city. Kansas has a total of 98,000 miles of local roads– 83,200 miles in the county group, 4,800 miles in small cities (rural corporate), and 10,000 miles in the urban areas. Within each group, the total local mileage is assigned the average local ADT to produce an aggregate VMT.

Each of the urban areas has an ADT based on counts from a mix of CBD, residential, and non-city (“HPMS donut area”) roads. The county average includes non-corporate roads both paved and unpaved. The small city averages are based on a selection of 3-8 cities within each maintenance district in different population groups.

This leaves some corner cases: roads in state parks are assigned an ADT/VMT based on visitation, suburban areas of urban cities (reverse donut) are assigned either the urban, county, or largest small city ADT as deemed appropriate by a traffic analyst. Undeveloped roads are typically assigned a marginal ADT value as they likely do not have regular daily traffic.

KDOT updates the local road counts on a 9-year cycle; the rural and urban ADTs are updated on the same cycle; the small city ADTs are updated every three years due to the sampling schedule. This provides an adequate Local VMT for Kansas for Highway Performance Monitoring System (HPMS) reporting.

2.3.2 Kentucky

The Kentucky Transportation Cabinet (Ketch) has developed a new method to estimate local VMT. KyTC had selected counts from randomly identified local road links. Since KyTC has complete count coverage on nonlocal roads (arterials and collectors), they modeled local road AADT based on count-based AADT on the connected nonlocal roads. Their approach and major findings are summarized as follows.

1. Randomly selected 28 counties to sample from rural and urban areas for each highway district to assure the spatial and socioeconomic distribution.
2. Estimated the minimum number of samples from each county to develop the model.
3. Collected and processed the traffic counts to determine the factored ADT.
4. Estimated the average local ADT for each sampled county and modeled the relationship between average collector ADT and local ADT.
5. A relationship exists between local and collector ADT.
6. The power function with exponent less than one best matched with the average of new counts.

A sample plot showing the relationship between local sample ADT and collector AADT is shown in Figure 1. Currently, KyTC adopted this methodology for HPMS submittals. Also, they are proactively involved in efforts to improve the traffic volume reporting.

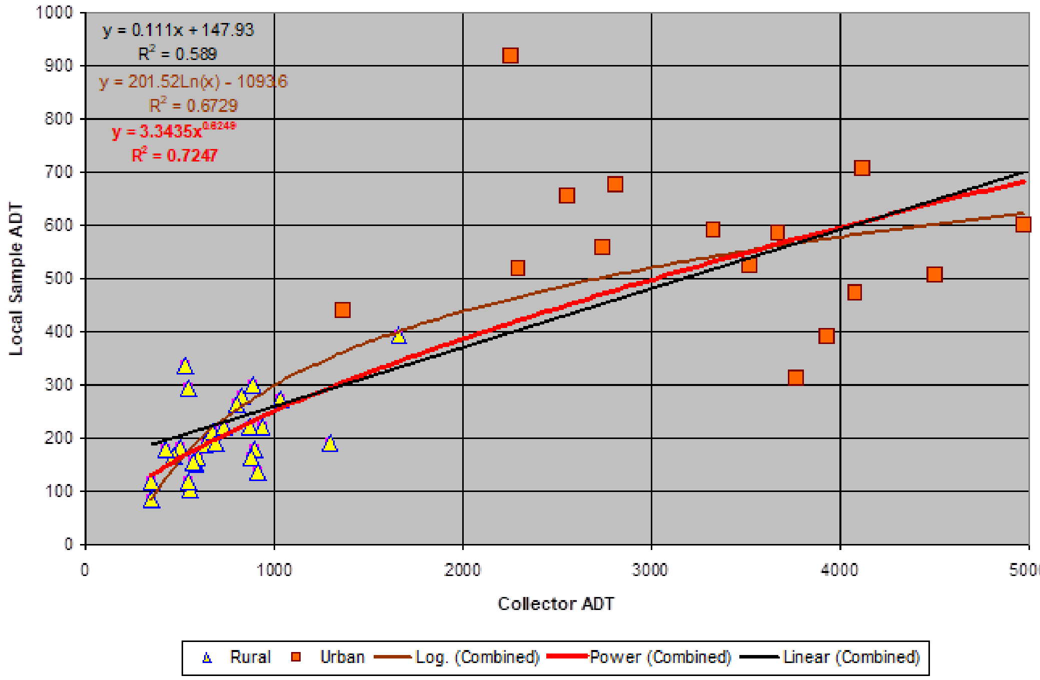


Figure 1 Comparison of local ADT to collector ADT (Source: KyTC)

2.3.3 New York

The New York State Department of Transportation (NYSDOT)'s Highway Data Services Bureau is responsible for annually reporting the state's VMT to the FHWA for the HPMS reporting. The traffic volume data is collected using 177 permanent count stations and portable short counts taken at approximately 12,000 locations per year. The portable traffic count program, also known as short counts, is comprised of inventory counts taken for minor collectors and local roads. These counts are 2-7 days in duration and are adjusted to represent annual averages using factors developed from the continuous counters. Using this process, NYSDOT develops a "current year estimate" of the AADT for all locations where counts have been taken within the prior 15 years.

A tabular matrix file that contains all locations for which NYSDOT-accepted traffic count data has been collected in the past 15 years is used for the VMT estimation. To complete all 15 years in the matrix, years for which there are no counts are filled in with an estimated AADT or a predicted AADT. An estimated AADT is an estimated value between two years with traffic count data. A predicted AADT is a value estimated using 'NYSDOT's Traffic Data Forecaster' tool which is based on a grouped linear regression approach.

To improve the estimates on local roads, 8,000 additional counts were taken during 2015 and added to the matrix table. The locations were randomly selected utilizing the existing road inventory. The result was more mileage covered by traffic counts with a statewide total as summarized next.

1. Rural minor collectors – counts on 70% of the mileage
2. Rural local roads – counts on 21% of the mileage
3. Urban local streets – counts on 11% of the mileage

2.3.4 South Carolina

The South Carolina Department of Transportation (SCDOT) currently uses default values if no count data is available to estimate local VMT. Each year, they calculate a percent growth for volume factor groups using all count data available for that year. The percent growth is then applied to the routes they are unable to collect traffic counts. However, their ongoing research on "cost-effective strategies for estimating statewide AADT" is mainly aimed at developing models for predicting AADT at non-coverage locations. Based on their work plan, SCDOT is exploring Kriging models to estimate AADT on local roads. This spatial interpolation technique uses nearby counts to estimate AADT at non-coverage locations. They proposed to develop an excel-based tool that will automatically calculate the AADT for all non-coverage locations using available count-based AADT.

2.3.5 Texas

The Texas Department of Transportation (TxDOT) estimates AADT on local roads using a statistical sampling process developed by the Texas Transportation Institute (TTI). The method is mainly aimed at assigning statistically valid median traffic volumes to non-covered locations. The methodological framework starts with grids overlaid on maps showing the functional classification of the road in a selected area. Sequential numbers are then assigned to each grid cell while random numbers are generated using Microsoft Excel. The grid cells corresponding to the random number are identified. Each iteration at which the grid cell contains a local street is marked as a count location on the map. This procedure was repeated to identify enough locations. The statistical analysis is performed to determine the number of count locations necessary to provide the representative samples in an area, based on population. According to their findings, the aforementioned procedure has resulted in median traffic count volumes on local streets that more realistically represent the variety of local streets that exist. The FHWA approved this random traffic count selection process for use.

2.3.6 Washington

Washington State Department of Transportation (WSDOT) collects traffic counts for all arterials and collectors. They have very limited traffic counts for local roads. For local roads, WSDOT estimates the VMT based on the total VMT for the arterials and collectors. In the case of rural local roads, 7% of the arterial and collector VMT total is considered. In the case of urban areas, WSDOT breaks down for each urbanized area and groups the small urban areas. The urban local roads, 11% of the total arterial, and collector VMT are considered. Within each of these groups (rural, small urban group, and individual urbanized areas), they take the total local road VMT and divide it by the local road miles to estimate ‘AADT per length’ (factor) for that group. This AADT ‘factor’ is used to determine the VMT of a road link.

2.3.7 Summary

Some DOTs that participated in the survey are currently involved in developing models to estimate AADT on local roads. Based on the survey response, some DOTs have conducted (some ongoing) noteworthy research initiatives to assess AADT at non-coverage locations.

2.4 Limitations of Previous Research

In the case of local roads, estimating AADT from a short-period perspective or along the selected links has been the usual practice. Installation of ATRs or permanent traffic counters on all functionally classified road links is not economical in terms of cost and benefit. Due to resource constraints, the estimation of AADT for the road links with little or no AADT continues to pose a challenge for agencies. Hence, an efficient AADT estimation model can also be a solution to reduce the cost and time required while ensuring good prediction of the AADT on local roads.

The local roads are designated for land access. Most travel is oriented from the land being accessed to the nearest nonlocal road (higher functionally classified road). However, most of the previous researchers did not consider the land use variables in the AADT estimation. While looking into the type of land use, parcel-level land use information will give indications about the number of trips generated by each parcel type. Apronti et al. (2016) considered the effect of land use characteristics on local road AADT. However, they considered land use characteristics as an indicator variable (binary variable) in their model. Thus, they assessed AADT based on the land being accessed and the type of land use. It is envisaged that considering the land use along with its coverage may give better insight into the AADT generation. This can be considered an advantage of assessing the AADT in response to changes in land use characteristics.

The locations with limited land use data, where road density is defined as the mileage of roads within a standard distance to the assessing road link (0.25 mile – 0.5 mile), is considered an indicator of how heavily the area is developed. Most of the previous studies considered accessibility as an indicator variable. They analyzed whether the local road had direct access to other higher functionally classified roads. However, it is a general notion that higher functionally classified roads with higher AADT have higher interaction with local roads. Hence, the distance to other higher functionally classified roads and AADT at those links can also be considered as potential explanatory variables.

There are many limitations of statistical methods for estimating the AADT. One of the main problems is that the parameters used in statistical methods are typically estimated for the entire study area. However, each variable varies with respect to space. In other words, the relationship between the dependent and explanatory variables is not stationary over space. Spatial statistical methods are used to improve the model accuracy by accounting for spatial variations in the explanatory variables. Based on the literature review, geospatial methods like GWR and Kriging can integrate variability in the explanatory variables (non-stationarity or heterogeneity) and the possible correlation of this variability to the AADT. The difference in GWR and OLS is that the explanatory variable is a function of location. Moreover, the predictability of GWR and Kriging was found to be better than the statistical models.

One of the advantages of spatial interpolation methods is that the data can be updated easily in the GIS platform. Further, these methods can be used for other jurisdictions by using their spatial map, existing AADT, socioeconomic factors, land use, and road characteristics. Overall, the spatial distribution of AADT values and other explanatory variables can be better utilized for the estimation of the AADT on local roads.

A few studies explored GWR and Kriging methods to estimate AADT. However, those studies considered major roads (interstates and other primary arterial roads) due to the availability of traffic counts for these roads. Also, the study area in their research was limited to certain counties. Apart from the statewide models, this research will also develop AADT estimation methods at the county-level. A comparative assessment of errors associated with each model will indicate the smallest spatial area for modeling AADT on local roads. Also, most of the previous studies considered a limited number of samples to estimate AADT on local roads. The present research uses available count-based AADT from a relatively larger number of traffic count stations (12,899 counts) for model development and validation. Overall, the previous efforts to estimate AADT for local functionally classified paved roads open to the public have been very limited.

CHAPTER 3: DATA COLLECTION AND PROCESSING

This chapter presents data collection and data processing methods adopted in this research.

3.1 Data Collection

The state of North Carolina, USA, is the study area of this research. This research examined four types of data for AADT estimation: available count-based AADT data, road data, socioeconomic and demographic data, and parcel-level land use data. All the data for this research was obtained from the NCDOT.

3.1.1 Count-based AADT data

The NCDOT's Traffic Survey Group gathers statewide traffic data to monitor the state's road planning, construction, and maintenance needs. The traffic data is comprised of the observations associated with traffic count stations in all of North Carolina between 2002 and 2017. The geospatial file contains traffic data for 44,378 counting stations in North Carolina. While looking into the local roads, traffic counts are collected on a biennial basis. This research uses available count-based AADT data for 2015 as only 2010 and 2015 socioeconomic data are available for the state. Additionally, as the traffic counts are collected biennially at selected stations on local roads, the average of available 2014 and 2016 count-based AADT data are also considered in the modeling and assessment process. The final database includes count-based AADT for 36,957 locations in 100 counties.

Figure 2 shows the distribution of traffic count stations among different counties in the state of North Carolina for the year 2015.

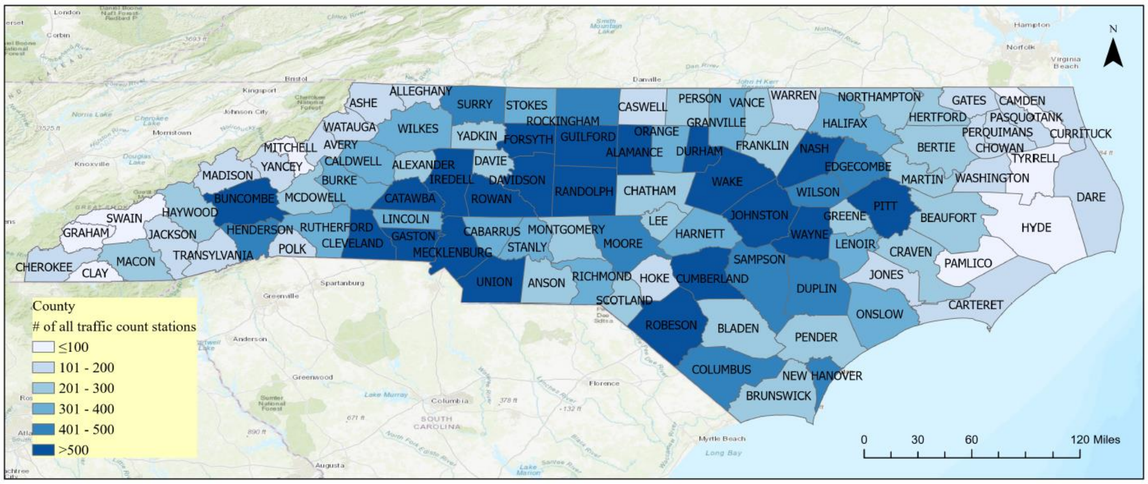


Figure 2 Distribution of traffic count stations in the state of North Carolina

From Figure 2, the distribution of the number of traffic count stations varies across different counties. The number of traffic count stations is comparatively higher in the central part (piedmont region) of North Carolina; however, the number of traffic count stations are lower at the western (mountains region) and eastern (coastal plain region) part of the state. The total number of traffic count stations ranges from a low 63 in Tyrell County to 1,678 in Wake County.

3.1.2 Road characteristics

The road network-related information was obtained in a geospatial format. This is a digital file from the road inventory database of the NCDOT that describes a subset of road attributes characteristic of the state road network. The state road system consists of interstates, US and NC routes, secondary roads, ramps, and all non-state roads maintained in North Carolina. This database includes speed limit, number of lanes, functional class, length of the link, etc.

3.1.3 Socioeconomic data

The shapefile of socioeconomic data contained information at the Traffic Analysis Zone (TAZ) level. TAZs are boundaries that contain socioeconomic data used as the foundation for trip-making in the travel model. There are 2,741 TAZs in the state of North Carolina. The data is based on the 2010 US Census. The TAZ file was a TransCAD geographical file consisting of variables like area type (urban/rural), population density, and employment-related information for the year 2015. Figure 3 illustrates the TAZ-level population data for the state of North Carolina.

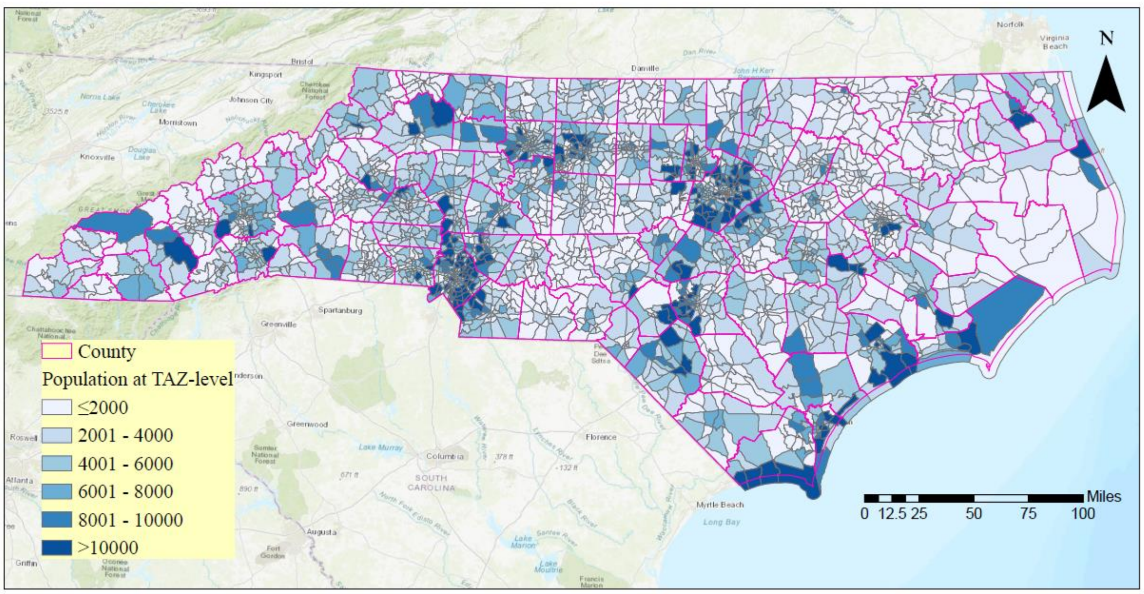


Figure 3 TAZ-level population data for the state of North Carolina

3.1.4 Land use

Information on land use development was collected from the parcel-level dataset ("nconemap" platform) for the entire state. This dataset does not provide statewide information on land use due to conflicting definitions of land use, incomplete data for many counties, and missing heated area information. Therefore, for the evaluation process, ten counties with high-quality data on land use were used when developing county-level models. The selected counties are shown in Figure 4.

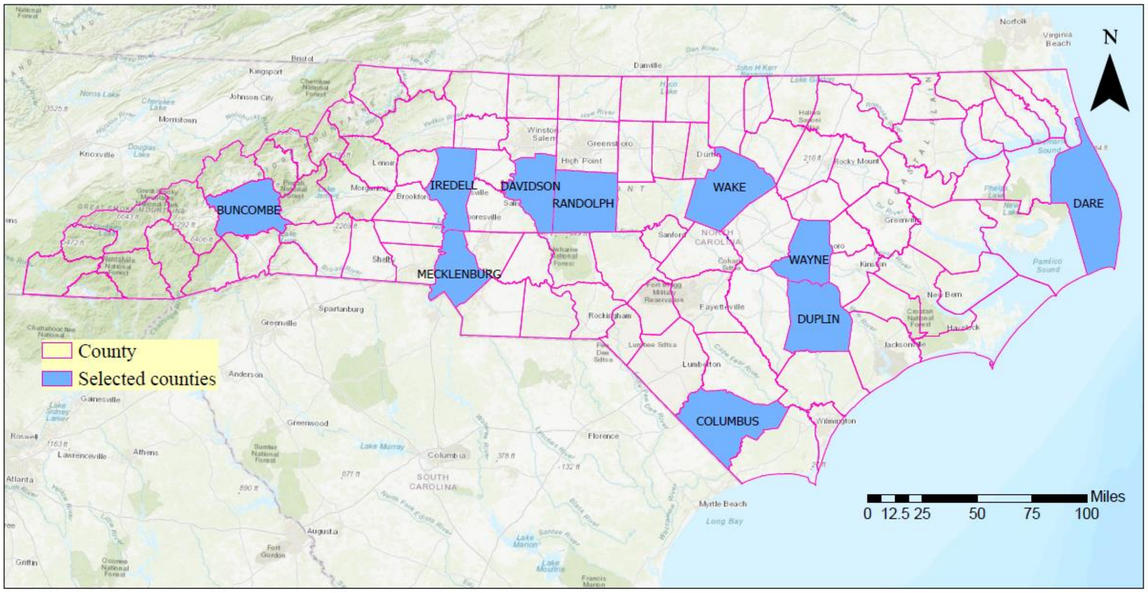


Figure 4 Selected counties for land use-based modeling

3.2 Data Processing

The data processing was carried out at various levels. Software tools such as ArcGIS 10.6.2, ArcGIS Pro, and Microsoft SQL were used for data processing. The data processing framework adopted for this research is outlined in Figure 5.

3.2.1 Count-based AADT data

The available count-based AADT data were processed to identify local roads for modeling and assessment. The AADT shapefile was overlaid over the road characteristics data obtained from NCDOT. A single shapefile with count-based AADT and road information was generated using the spatial join feature in ArcMap. This research considered only those local road links with AADT values lower than 5,000.

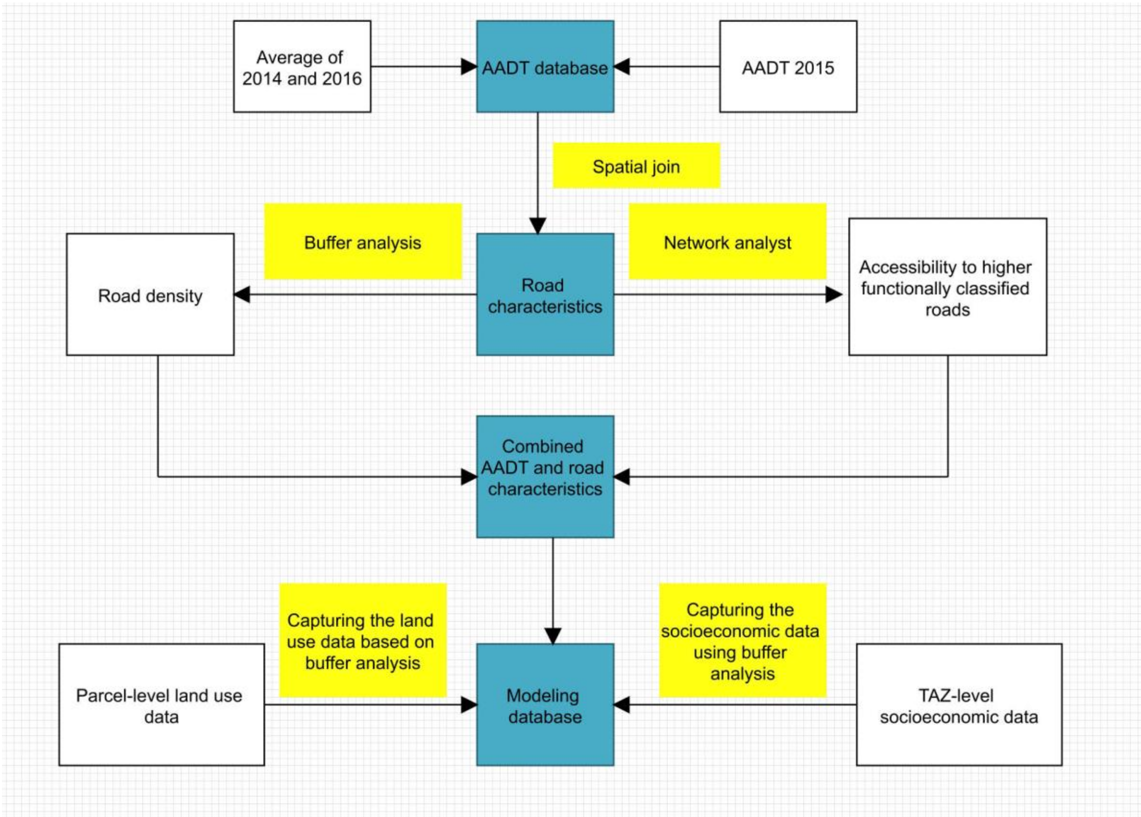


Figure 5 Data processing

Furthermore, the available count-based AADT data were classified into two categories: 1) local roads, and 2) higher functionally classified roads. The available count-based AADT data at 12,899 local road traffic count stations were considered based on the criteria. Figure 6 shows the distribution of the 12,899 local road traffic count stations among different counties in the state of North Carolina.

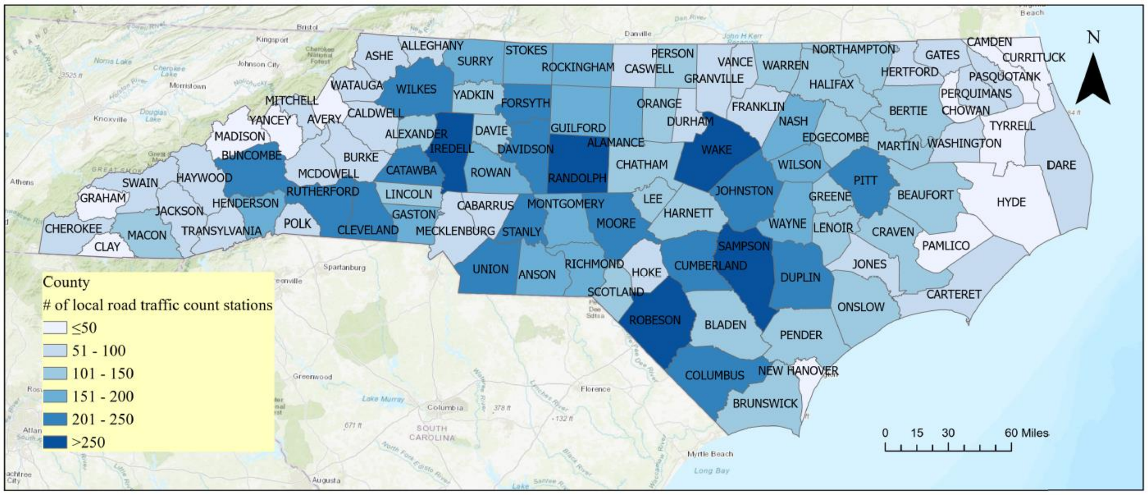
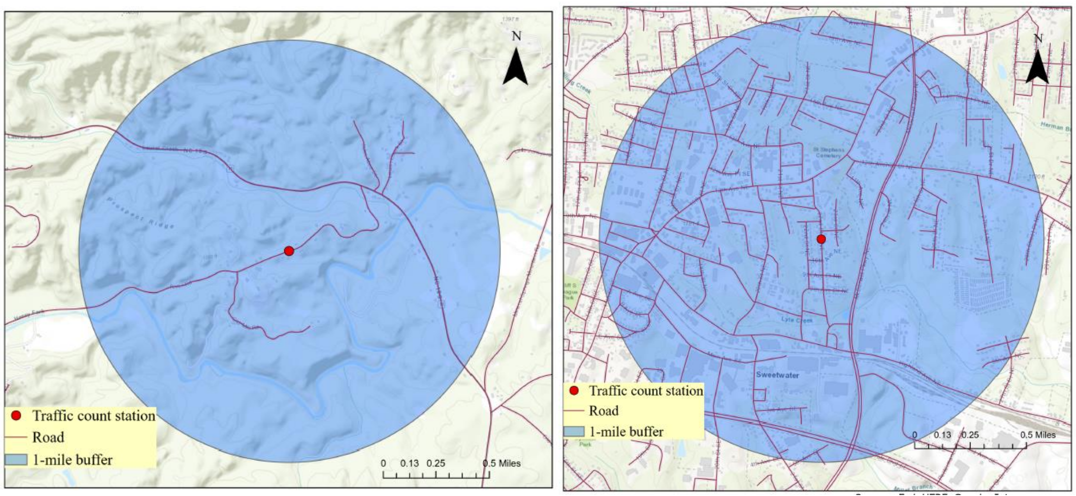


Figure 6 Distribution of local road traffic count stations in the state of North Carolina

3.2.2 Road characteristics

The road density (length of road/square mile of the area) in an area generally indicates how heavily the area is developed (U.S. Environmental Protection Agency, 2017; Meijer et al., 2018). As the land use data is limited to some counties in the study area, road density is considered as an indicator of development in this research. A buffer of 1-mile has been created on each traffic count station in the study area. Further, the intersect feature in ArcMap was employed to capture the road density within a buffer, as shown in Figure 7.

To estimate the shortest path (path distance), “network analyst” tools in ArcGIS were employed. A new network dataset for the state has been created. The road characteristics shapefile obtained from NCDOT was used for creating the network dataset. The one-ways are separately identified and inputted into the network dataset. The intersection points in each higher functionally classified road were located using the intersect feature in the ArcGIS. The intersections in the higher functionally classified roads were extracted and added as a new feature class. To find the distance between the local roads and the nearest higher functionally classified road (collector roads and above), the ‘closest facility’ analysis and ‘origin-destination cost matrix’ were performed. Both tools measure the cost of traveling (in terms of distance and time) between an origin and destination.



(a) Low road density area (b) High road density area

Figure 7 Road density within 1- mile buffer

The closest facility analysis tool in ArcGIS Pro measures the path distance between ‘incidents’ and ‘facilities. In this research, ‘incidents’ are entered as traffic count stations on the local roads, and ‘facilities’ are coded as intersection points on the higher functionally classified road. This tool can calculate the best route between incidents and facilities as shown in Figure 8, returning travel distance and the travel time as output.

Similarly, the origin-destination cost matrix solves and measures the lowest cost path along with the network from multiple origins and destinations (Figure 9).

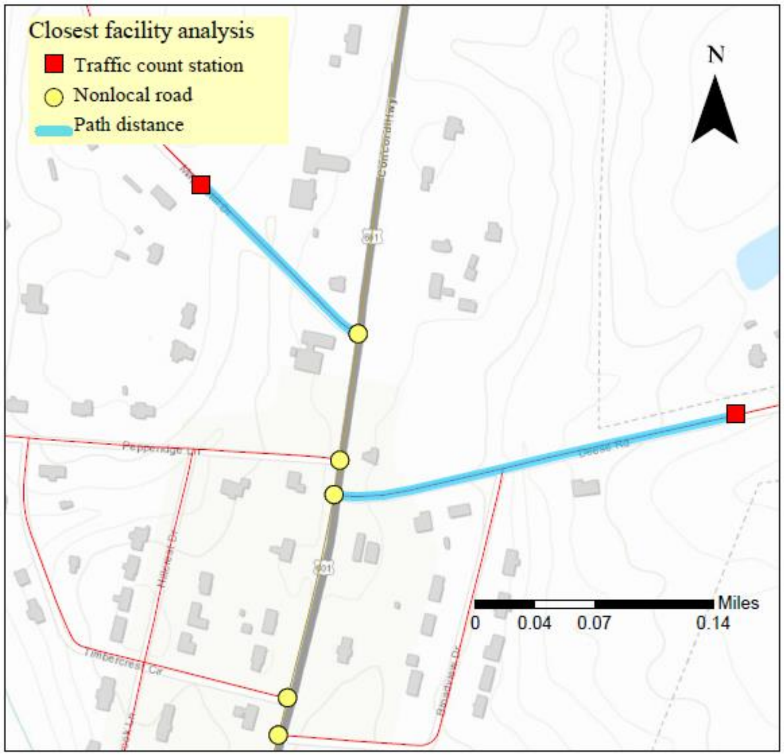


Figure 8 Closest facility analysis

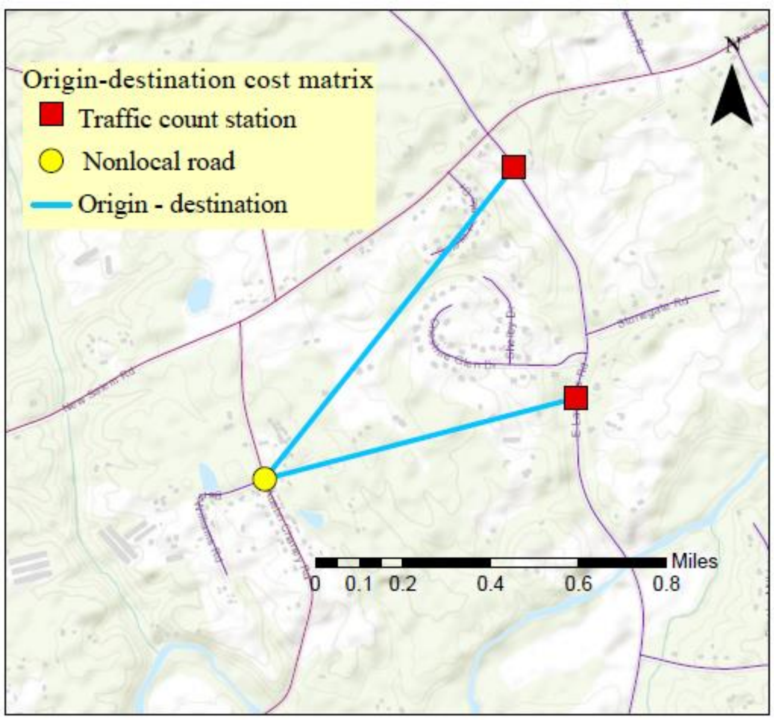


Figure 9 Origin-destination cost matrix

The traffic stations are the origins, and intersection points at the higher functionally classified roads (collector and above) are considered as destinations. Also, while solving, the direction of travel (toward the facility). Compared to the closest facility analysis, origin-destination cost matrix analysis reduces the computational time. However, the closest facility analysis gives the true shapes of the routes as output. Finally, the count-based AADT at the nearest higher functionally classified road was estimated from the statewide count-based AADT data (all functionally classified roads).

3.2.3 Socioeconomic data

The next step in data processing is to capture the socioeconomic data in the study area. The TAZ-level data from the statewide travel demand model was used as the areal unit of measurement. Many researchers use TAZ as their basic geographical unit for the aggregation of socioeconomic data to estimate AADT (Staats, 2016; Zhong and Hanson, 2009; Apronti et al., 2016). In general, each TAZ represents the spatial unit containing similar land use and commuter patterns (US Census Bureau, 2010).

The statewide TAZ-level data contains socioeconomic and other attributes such as region (coastal plain, piedmont, and mountains), area type (urban, suburban, and rural), density, population, household income, workers, different categories of employees (industrial, high industrial, retail, high retail, office, service, government, educational, and hospital), and total employees. Buffers of 50 feet, 100 feet, 330 feet, 660 feet, and 1,320 feet were generated along the road link, as shown in Figure 10.

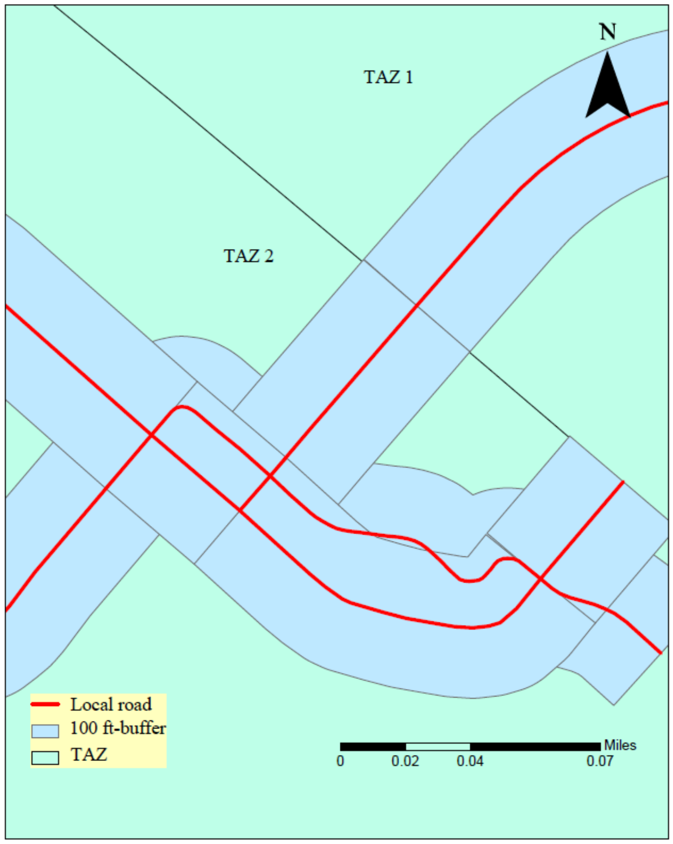


Figure 10 Extracting population within a 100 feet buffer

Further, the 'intersect' feature in ArcGIS was used to extract socioeconomic data by overlaying buffers over the TAZs. It was assumed that the socioeconomic variables are uniformly distributed over each TAZ. The weighted average population in the buffer of a subject road link was estimated using Equation (4).

$$P_i = \sum_j \frac{A_{j,i}}{A_j} * P_j \tag{4}$$

where, P_i = population of buffer 'i', $A_{j,i}$ = actual area of TAZ 'j' in buffer 'i', A_j is the area of the TAZ 'j', and P_j is the population of TAZ 'j'.

A similar analysis was performed to capture the weighted average employment density and other employment categories.

3.2.4 Land use data

The study links' land use characteristics were identified using the buffer method. The North Carolina parcels geodatabase contains 5,536,606 parcels in the state. Nevertheless, there are no definitions of land use in 27% of the parcels. The research, therefore, considered selected counties for modeling based on the quality of land use data, population density, and the number of counts available in that county. In county-wide parcel data, missing, abrupt values, duplicate data points, and land use developments after the year 2015 (modeling year -2015) were removed from the dataset. The raw dataset consists of several land use categories. The descriptions of the chosen land use categories are shown in Table 1.

The total number of residential parcels (single-family residential units and multifamily residential units) and areas of other types of parcels were extracted for analysis and modeling. As 50 feet was observed inadequate to capture parcels in some cases, 100 feet was considered as a suitable buffer width to capture land use characteristics within the vicinity of the local roads. As an example, Figure 11 shows a 100 feet buffer (flat buffer) generated around a road link to extract land use characteristics.

In general, local roads are designed for land access. Most travel is oriented from the land being accessed to the nearest nonlocal road. The AADT is impacted by the amount of land being accessed, the type of land use, and the density of the development. Hence, capturing the land use characteristics is very important for the accurate estimation of AADT.

Table 1 Land use descriptions

Land use categories	Description
Agricultural	Area for agriculture use
Commercial service	Shopping mall, service station, commercial condominium, furniture showroom, supermarket, Convenience store, fast-food centers, and small sized grocery stores
Government	County, state, federal, municipal government buildings
Institutional	Public colleges, Church, day care, lab-research, and other institutional facilities for communities
Industrial	Manufacturing units, distribution centers, industrial common area, specialized industrial operations
Multi-family residential	Townhouse apartments, garden apartments, hi-rise apartments, mobile homes etc.
Office	Office condominium
Recreational/social	Theatre, night club, bowling alley/ skating rink, club – lodge, golf course, and other recreational amenities
Retail	Area utilized for retail shops
School/college	Public schools, private schools
Single-family residential	Residential parcel units either fully detached, semi-detached, row houses, or a town home
Transportation	Truck terminal, parking lots, and other transportation facilities
Warehouse	Manufacturing, wholesale trade, distribution units etc.

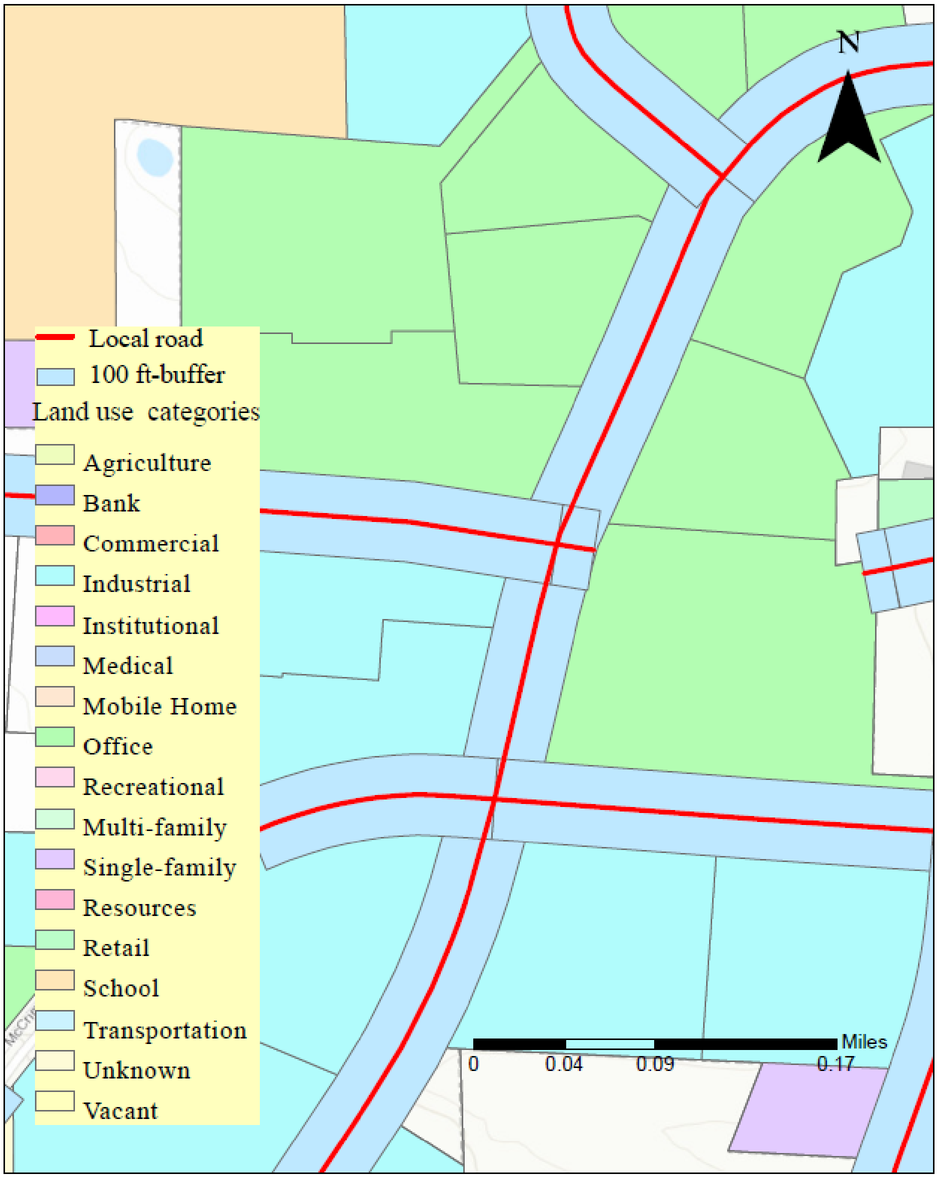


Figure 11 Extracting land use within a 100 feet buffer

CHAPTER 4: METHODOLOGY

The focus of this research is to develop a sustainable and repeatable AADT estimation method for local roads. Statistical (OLS) and geospatial analytical methods (GWR, Kriging, IDW, and natural neighbor interpolation) were explored for modeling. The results and spatial distribution of errors were assessed and compared between each modeling method. The methodological framework adopted for this research includes the following steps:

1. Descriptive analysis of local road data
2. Identifying potential explanatory variables influencing local road AADT
3. Check for multicollinearity between explanatory variables
4. Develop local road AADT estimation models
    a. Statewide
    b. County-level
5. Validate the models
6. Estimating local road AADT at non-coverage locations

4.1 Descriptive Analysis of Local Road Data

A descriptive analysis was conducted to understand the influence of selected explanatory variables on the available count-based local road AADT. The median count-based local road AADT was used as the central tendency measure since the data had a high degree of skewness. The minimum, mean, maximum, and standard deviation of local road AADT were also computed and examined.

4.2 Identifying Potential Explanatory Variables Influencing Count-based Local Road AADT

In general, AADT is impacted by the amount of land being accessed, the type of land use, and the density of the development. Also, a local road could support through traffic from other local roads. These local characteristics were considered as the potential explanatory variables influencing local road AADT.

4.3 Check for Multicollinearity Between Explanatory Variables

The Pearson correlation coefficients were computed to perform correlation analysis. The Pearson correlation coefficient illustrates the strength of the linear relationship between two variables. The Pearson correlation coefficient that fell within a 95% confidence level was classified into six categories for further assessment (Mane and Pulugurtha, 2019). They are:

1. HN - High negative correlation (less than -0.5)
2. MN - Moderate negative correlation (-0.5 to -0.3)
3. LN - Low negative correlation (-0.3 to 0)
4. LP - Low positive correlation (0 to +0.3)
5. MP - Moderate positive correlation (+0.3 to +0.5)
6. HP - High positive correlation (greater than 0.5)

One explanatory variable of two correlated explanatory variable is chosen for the modeling process.

The spatial autocorrelation was examined to determine the effect of AADT on its neighboring link (nearby AADT stations). The Moran's I in the GIS environment measures the spatial autocorrelation of the dataset. The value of Moran's I ranges from -1 to 1. The Moran's I value -1 indicates the perfect clustering of dissimilar values or negative spatial autocorrelation in the dataset. If the Moran's I value is near to zero, it indicates no spatial autocorrelation. The Moran's I value of 1 indicates the perfect positive autocorrelation or the clustering of similar data points in the study area.

4.4 Develop Local Road AADT Estimation Models

The statistical methods (OLS) and geospatial methods were explored in the modeling process. The spatial methods incorporate the effect of spatial locations into the local road AADT estimation. The geospatial analytical methods assume that locations with AADT counts close to one another are alike, and the level of correlation reduces with an increase in the distance between these locations. The predictability of the geospatial models depends on the density and spatial distribution of local road traffic count stations. GWR, Kriging, IDW, and natural neighbor interpolation were explored for the spatial modeling of local road AADT. Each modeling approach is briefly discussed in the following subsections.

The best two models (one statistical and one geospatial) were identified from the statewide modeling results and used for the county-level modeling and estimating local road AADT at non-coverage locations.

4.4.1 Ordinary least square (OLS) regression

The general OLS model is widely used to model the relationship between a dependent variable (count-based local road AADT) and the explanatory variables. The non-constant error variance problem is common in count-based predictions. This research addressed that issue by log-transforming the count-based local road AADT. The general form of the OLS regression model used in this research is expressed as in Equation (5).

$$Ln\ AADT = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots \beta_k X_k + \varepsilon \tag{5}$$

whereβ_j (j = 0,1,2…k) = set of estimated parameters (coefficients), ε = the random error, and k = number of explanatory variables.

By minimizing the sum of the squares of the residuals, this method calculates the best fitting line for the observed data.

4.4.2 Geographically weighted regression (GWR)

In GWR, the local regression is performed at the geographic space. Each parameter estimate is based on the data for a subset of local road traffic count stations. This will address the extreme heterogeneity or variability in spatial data while modeling. In other words, GWR is essentially a spatially weighted regression over space, with each regression centered on a point in the dataset. The basic mechanism of GWR depends on obtaining separate regression equations for each spatial zone in which the area-centered Kernel is adapted in such a way that the adjacent areas are weighted based on the distance decay function (Fotheringham et al., 2002). The general form of estimation is given in Equation (6).

$$\mathrm{Y} = \mathrm{X}\beta\ (s) + \varepsilon \tag{6}$$

where Y is the response outcome (local road AADT), and *X* is an 'n' by '*(k+1)'* data matrix with k explanatory variables. Y, X, and ε vary spatially. The least-square estimates and its variance at any traffic count station 'i' is provided in equations (7) and (8).

$$\widehat{\beta_i} = (X^T W_i X)^{-1} X^T W_i Y \tag{7}$$

$$VAR\left(\hat{\beta}_i\right) = (X^T W_i^{-1} X)^{-1} \tag{8}$$

where W_i is an *n* by *n* diagonal matrix of spatial weights whose off-diagonal elements are zero and diagonal elements are spatial weights (Fotheringham et al., 2002).

$$W_i = \begin{bmatrix} w_{11} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & w_{nn} \end{bmatrix}$$

This indicates that there are values of β that can be estimated for any local road segment of interest. The values vary based on the spatial weight matrix. The weights are assigned based on the distance between traffic count station 'i' and other traffic count stations in the study area. The nearby local road segment characteristics are assumed to be alike, and the influence will reduce with an increase in distance. Functions such as Gaussian and bi-squared functions, given by Foderingham et al. (2003), are used to assign weights. The function form of Gaussian and bi-squared functions, respectively, are provided in equations (9) and (10).

$$W_{jj} = exp\left[-o.5(d_{ij}/b)^2\right] \tag{9}$$

$$W_{jj} = \begin{cases} \left[1-(d_{ij}/b)^2\right] & d_{ij} \leq b \\ 0 & Otherwise \end{cases} \tag{10}$$

Where d_{ij} re the distances between the traffic count stations and b is the band width.

Another important aspect is to find the optimum bandwidth (neighborhood) for the local regression. The bandwidth can be based on either the number of nearby traffic count stations or the distance band. In the case of the number of nearby traffic count station, the neighborhood size will be smaller for dense features and larger for sparse features. However, the number of traffic count stations will be a constant for the study area when the distance band is used. The Golden search approach, which is based on minimizing the value of the Akaike Information Criterion (AIC) was adopted to find the optimum bandwidth.

4.4.3 Kriging

Traditional interpolation techniques are based on the mathematical approach, which assumes that the spatial dependence of data is “implicit.” However, the spatial variation of any variable cannot be explained by mathematical expression (Wang, 2012). Spatial variability is characterized by two main parameters – large scale variation and small-scale spatial autocorrelation (error term). The general form of spatial variability is as shown in Equation (11).

$$Z_i(S) = \mu_i(S) + \varepsilon_i(S) \tag{11}$$

where $Z_i(S)$ is the dependent variable (count-based local road AADT), $\mu_i(S)$ is the conditional mean, and the ϵ_i (S) is the error term for the traffic count station ‘S’.

Kriging considers the surrounding count based AADT values to estimate the AADT at non-covered location. The Kriging method uses a weighted sum of the data at traffic count stations to compute the non-covered location (Oliver and Webster, 1990). These weights are typically based on the spatial arrangement and the distance between the traffic count stations. Equation (12) indicates the general form of the Kriging prediction mechanism.

$$\hat{Z}(S_0) = \sum_{i=1}^{N} W_i \, Z(S_i) \tag{12}$$

where unknown weights W_i are given to each measured value $Z(S_i)$ (count-based AADT) to compute the estimate for the non-covered location, $\hat{Z}(S_0)$. To evaluate these weights in the equation, the spatial autocorrelation is to be quantified. Therefore, Kriging relies on the semi-variogram plots (variance with respect to the distance) to account for the autocorrelation factor. Semi-variance (with respect to distance ‘h’) is an average of the squared deviations of the data pairs (two nearby traffic count locations) and is computed using Equation (13).

$$Semivariogram(distance_h) = 0.5 * average((value_i - value_j)^2 \quad (13)$$

where the values for ‘i’ and ‘j’ indicate the pairs of the points (two nearby traffic count stations). The obtained variance is plotted to compute the appropriate function (linear, spherical, Gaussian, etc.) of the corresponding semi-variogram. This function is highly essential in the case of Kriging, as it influences the predictability of the whole model.

The semi-variogram model remains pivotal in the case of the Kriging method since the overall predictive capability is dependent on it. The value of semi-variance over distance is typically plotted to determine the type of variogram. The overall variogram plot is also used to examine the overall trend of count-based AADT and its influence over the distance component. Figure 12 indicates the plot of a semi-variogram with their components indicated.

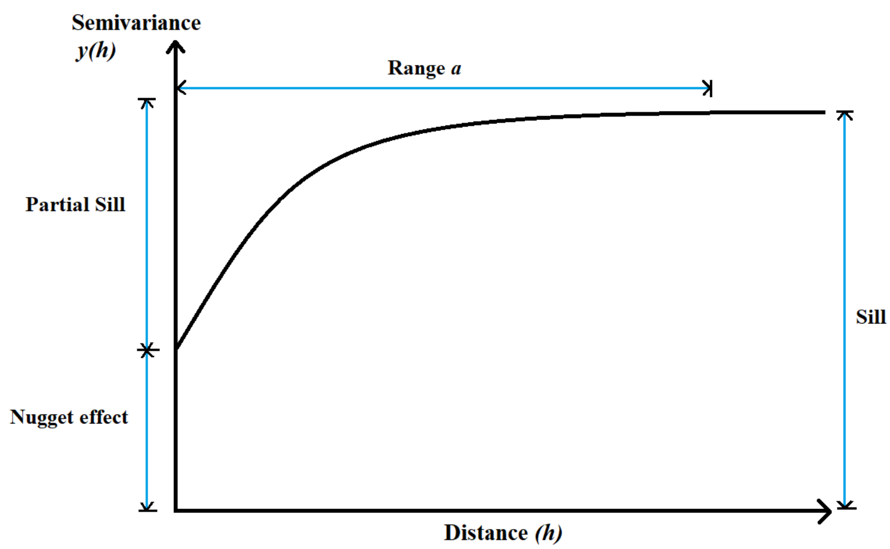


Figure 12 Semi-variogram plot (with components)

The two main components of the semi-variogram plot are "range" and "sill." The range is defined as the distance at which the model does not influence the prediction (the curve flattens) (ESRI, 2018). The corresponding y-value for the range is defined as the sill. In other words, the sill is the maximum value of the semi-variance before the curve flattens out. Therefore, a steeper curve indicates that the influence of the distance factor diminishes significantly. Nugget, on the other hand, is defined as the initial intercept (value of variance at a distance of '0') mainly attributed to measurement or spatial errors. Partial sill is defined as the difference between sill and nugget.

Based on the functionality of the estimators, the types of Kriging methods considered for this research are:

1. simple Kriging
2. ordinary Kriging
3. universal Kriging, and
4. Empirical Bayesian Kriging

The formulation for each Kriging approach is wee documented in many previous researches. The study conducted by Shamo et al., 2015 is

Simple Kriging

The simple Kriging method considers the mean of the data points (count-based AADT) to be a constant known value throughout the study area. The general form of the simple Kriging estimator is represented in Equation (14) (Shamo et al., 2015).

$$Zx(y) = \sum_{s=1}^{n(y)} Ws(y)[Z(ys) - \mu] + \mu \tag{14}$$

where W_S is the weight associated with the traffic count station y_S, Zx(y) is an estimate of value Z(y); Z(ys) is the value of the datapoint (local road AADT in this case) associated with location 'ys' and μ is the unknown constant.

Ordinary Kriging

The ordinary Kriging method considers the variation in the local mean. However, this local variation is limited by the neighborhood of the vicinity considered. Therefore, the model assumes that the mean is unknown but not fixed. Equation (15) indicates the general form of the universal Kriging method (Shamo et al., 2015).

$$Zx(y) = \sum_{s=1}^{n(y)} Ws(y)Z(ys) + \left[\sum_{s=1}^{n(y)} Ws(y)\right]\mu(y) \tag{15}$$

where Ws is the weight associated with traffic count station ys, Zx(y) is an estimate of value Z(y); Z(ys) is the value of the datapoint (local road AADT in this case) associated with location 'ys', and μ(y) is the unknown constant of the corresponding location. However, the summation of the weights ultimately adds up to 1 ($\sum_{s=1}^{n(y)} Ws(y) = 1$). Hence, Equation (16) represents the final form of the ordinary Kriging method.

$$Zx(y) = \sum_{s=1}^{n(y)} Ws(y)Z(ys) \tag{16}$$

Universal Kriging

The universal Kriging uses the mean of data points as a functional dependence corresponding to the spatial location considered (Kis, 2016). Therefore, the presence of a local trend is considered in the case of universal Kriging. There is no involvement of a mean parameter like simple and ordinary Kriging. Equation (17) indicates the general form of the universal Kriging prediction.

$$Zx(y) = \sum_{s=1}^{n(y)} Ws(y)Z(ys) \tag{17}$$

where Zx(y) is an estimate of value Z(y); Ws is the weight associated with location ys; Z(ys) is the true value of the datapoint (available count-based local road AADT in this case) associated with location 'ys'.

Empirical Bayesian Kriging

Empirical Bayesian Kriging is a geostatistical interpolation method that uses an automatic simulation process to iterate the semi-variograms to estimate at non-covered locations. Unlike other Kriging methods, Empirical Bayesian Kriging uses automatic sub-setting and simulation processes to estimate the parameters (Gribov and Krivoruchko, 2020). To estimate these parameters, Empirical Bayesian Kriging considers the error factor in the semi-variogram to produce an accurate result overall. This research considered the model fitting algorithm provided by Gribov and Krivoruchko, 2020 to develop a valid EBK model.

One of the major differences between the Empirical Bayesian Kriging and other Kriging methods includes the usage of multiple semi-variogram plots which are iterated and optimized for better prediction.

The cross-validation approach is used to identify the best Kriging model to estimate AADT on local roads. The cross-validation mechanism works by removing data for a traffic count station from the dataset and using the remaining traffic count stations in the near vicinity for estimating AADT local road AADT at the removed traffic count station. Various statistical measures are available in the software package to evaluate these cross-validation results. They include the mean prediction error (MPE), mean standardized error (MSE), average standard error (ASE), root mean square error (RMSE), Root Mean Square Standardized Error (RMSSE) (ESRI, 2018).

To find the prediction Z_S at each point i using the neighboring data Z_i, the Kriging method is used. An estimate of the prediction location, Z^*_S with variance σ^2 is computed from interpolation. Kriging error is computed as the difference in prediction and actual value, as shown in Equation (18).

$$Kriging\ error\ (E_s) = Z_S^* - Z_S \tag{18}$$

Furthermore, the standardized value at each point is computed as the ratio of the Kriging error to the standard deviation σ_S for corresponding location α (Equation (19)).

$$Standardized\ error\ (e_s) = E_S/\sigma_S \tag{19}$$

Using the computed Kriging and the standardized errors, the mean error and the mean standardized error are computed using equations (20) and (21).

$$Mean\ Error\ (ME) = \frac{1}{n}\sum_{S=1}^{n}\{Z^*{}_S - Z_S\} \tag{20}$$

$$Mean\ Standardized\ Error\ (MSE) = \frac{1}{n}\sum_{S=1}^{n}\frac{\{Z^*{}_S - Z_S\}}{Z_S} \tag{21}$$

where Z^*_S is the estimated AADT, Z_S is the count-based AADT, n is the number of values in the dataset and σ_S is the standard deviation for the corresponding traffic count station ‘s’.

The MSE value of the data represents the accuracy in the semi-variogram. Therefore, a value of zero indicates that the variogram used is accurate for the corresponding dataset. However, deviation from zero indicates that the model is either underestimating (MSE < 0) or overestimating (MSE > 0).

Average standard errors are defined as the mean of the prediction standard errors. Equation (22) represents the computation of the average standard error.

$$Average\ Standard\ Error\ (ASE) = \sqrt{\frac{1}{n}\sum_{S=1}^{n}\sigma_S^2} \tag{22}$$

where n is the number of values in the dataset and σ^2 is the kriging variance for the location ‘s’.

Root mean squared and the root mean square standardized prediction errors are computed using the squared difference of the error terms. Equations (23) and (24) represent the computation of root mean square error (RMSE) and root mean square standardized error (RMSSE).

$$Root\ Mean\ Square\ Error\ (RMSE) = \sqrt{\frac{1}{n}\sum_{S=1}^{n}\left[Z_S - {Z^*}_S\right]^2} \qquad (23)$$

$$Root\ Mean\ Square\ Standardized\ Error\ (RMSSE) = \sqrt{\frac{1}{n}\sum_{S=1}^{n}\left(\frac{{Z^*}_S - Z_S}{\sigma_S^2}\right)^2} \qquad (24)$$

where Z^*_S is the estimated AADT, Z_S is the count-based AADT, n is the number of values in the dataset and σ^2_S is the variance for the traffic count station 's'.

Figure 13 shows the settings of the Kriging model in ArcGIS Pro. A sample semi-variogram using the exponential model is also shown in Figure 13.

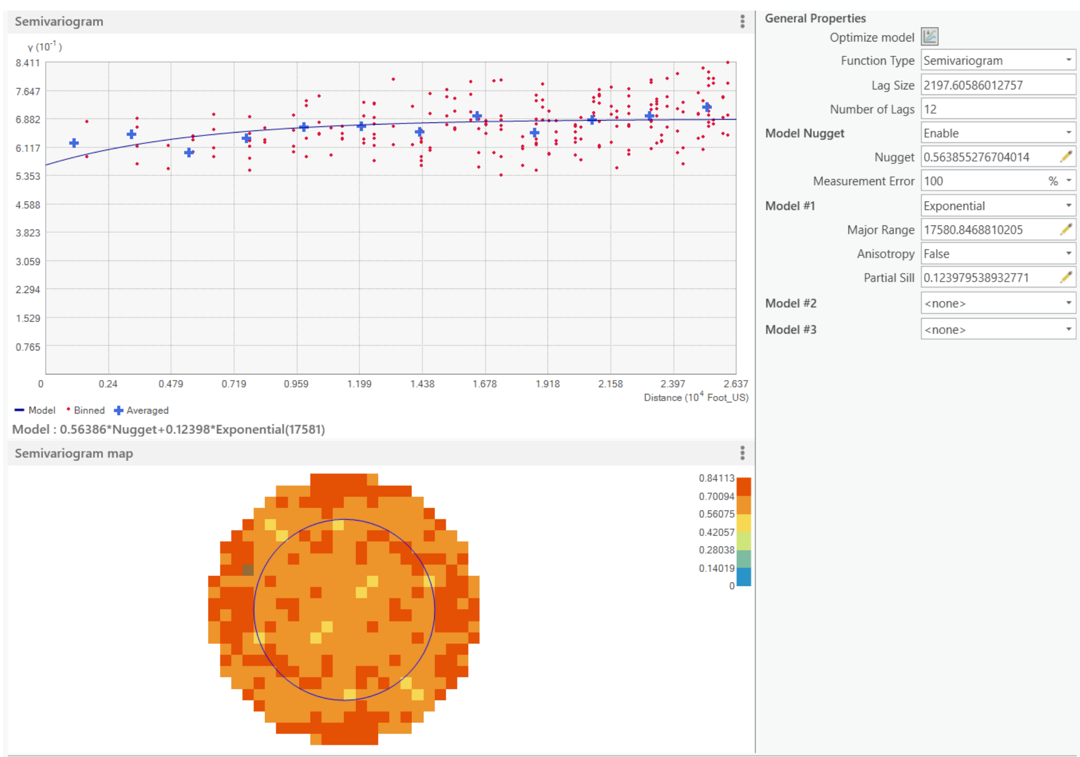


Figure 13 Fitted exponential semi-variogram

4.4.4 Inverse distance weighted (IDW)

The IDW interpolation mechanism works on the assumption that the objects closer are more alike than the ones farther away. In the present research, IDW allocates higher weights to the closer count-based AADT than the farther ones to estimate AADT at a non-covered location. These weights are inversely proportional to the distance values raised to the optimal power 'p'. Equation (25) indicates a general form of the IDW interpolation method (Bartier and Keller, 1996).

$$Z^* = \frac{\sum_{i=1}^{n} w_i \, Z_i}{\sum_{i=1}^{n} w_i} \tag{25}$$

where Z^* is the estimated AADT and w_i indicates weights corresponding to the points Z_i (known count-based AADT). As the distance increases, the weights reduce drastically. The weights for each point are computed as in Equation (26).

$$w^* = \frac{1}{d_i^p} \tag{26}$$

where 'd_i' indicates the distance parameter and 'p' represents the chosen optimal power.

The process of IDW consists of an allocation of two main components, the distance of the vicinity and the optimal value for the power 'p'. Therefore, these two parameters play a significant role in estimating AADT. It is highly important to allocate optimal values for higher accuracy in local road AADT estimation.

The selection of optimal distance of the vicinity also comprises the shape of the area (like circular, elliptical, etc.). Furthermore, the vicinity to be considered also consists of selecting the number of available count-based AADT in the area for interpolation. IDW also gives the flexibility to divide the area into up to eight sectors with minimum and maximum number of count-based AADT for consideration. Similarly, to select the optimal power 'p' for a given data, root mean square error (RMSE) from the cross-validation is used.

4.4.5 Natural neighbor interpolation

Natural neighbor interpolation refers to spatial interpolation that works on the assumption that two objects are related to each other if they are located close to one another (Bobach, 2009). In this research, every traffic count station "claims" to be a neighbor to a traffic count station in the near vicinity. Therefore, the natural neighbor interpolation method considers a local phenomenon (dependence of points based on their location).

Unlike other methods of spatial interpolation, natural neighbor uses the inclusion of a "Thiessen polygon" or "Voronoi diagram" which is defined as the polygon generated around each point (local road traffic count station) representing its area of influence. The boundaries of these polygons are generated such that the edges are equidistant from the points in the adjacent polygons. Therefore, the inclusion of a non-covered location results in the overlap of its surrounding Voronoi diagrams. Based on the polygon generated for the non-covered location, the weighted average of the existing count-based AADT is computed by taking the area of overlap. The general form of the natural neighbor interpolation technique is shown in Equation (27) (ESRI, 2018).

$$Z(S_O) = \sum_{i=1}^{n} W_i \, Z(S_i) \tag{27}$$

where $Z(S_O)$ is the natural neighbor estimation at (So) and n is the number of nearest neighbors (traffic count stations) used for interpolation. The interpolation is carried out using the count-based AADT $Z(S_i)$ and a weight of W_i associated with that.

Even though the method uses a similar mechanism, i.e., the weighted average, natural neighbor interpolation differs from other techniques as the weights vary for each point based on its area of overlap. Therefore, based on the spatial distribution of the count-based AADT, the interpolation technique is carried out using the Voronoi diagrams.

4.5 Validate the Models

Count-based AADT data for selected local functionally classified public road links (~25% of the sample) were set aside for validation purposes. These links were randomly selected while ensuring that they represent a geographically/spatially distributed sample across North Carolina. Each of the developed models was validated using Root Mean Square Error (RMSE), the Mean Absolute Percentage Error (MAPE), and the Mean Percentage Error (MSE). The general equations for estimating these indicators are shown in equations (28), (29), and (30).

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(Count\text{−}based\ AADT_i - Estimated\ AADT_i)^2}{n}} \tag{28}$$

$$MAPE = \frac{1}{n}\sum_{t=1}^{n}\left|\frac{Count\text{−}based\ AADT_i - Estimated\ AADT_i}{Count\text{−}based\ AADT_i}\right| \tag{29}$$

$$MPE = \frac{1}{n}\sum_{t=1}^{n}\left(\frac{Count\text{−}based\ AADT_i - Estimated\ AADT_i}{Count\text{−}based\ AADT_i}\right) \tag{30}$$

4.6 Predicting Local Road AADT at Non-Coverage Locations

The best-fitting model was adopted for estimating AADT at the non-covered locations (locations with no traffic counts). There are nearly 700,000 such locations in the state of North Carolina. The estimated AADT and length of each local road link is multiplied to estimate VMT for each link and summed to compute statewide local road VMT.

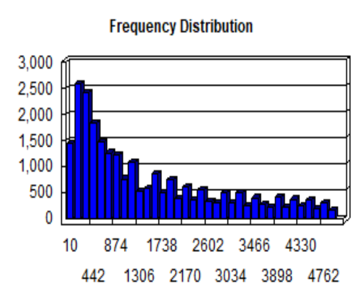
CHAPTER 5: DESCRIPTIVE ANALYSIS

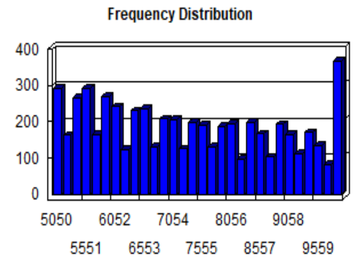
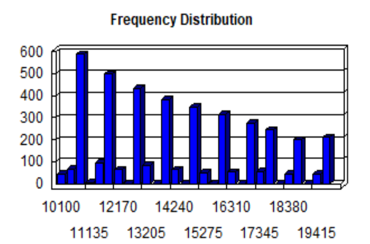
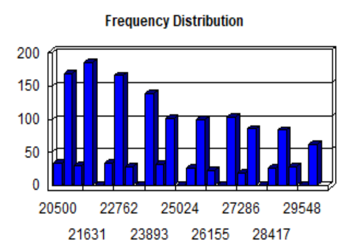
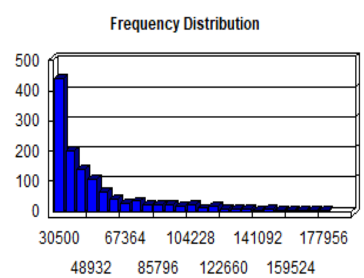
This chapter covers descriptive analysis to understand the relationship between count-based local road AADT data and selected explanatory variables. The analysis was performed based on different AADT ranges, functional classification type, speed limit, population density, employment density, road density, and local travel characteristics.

5.1 AADT Ranges

NCDOT's Traffic Survey Group collects traffic data statewide. Count-based AADT is available at 26,192 traffic count stations for the year 2015. As the local road traffic counts are collected biennially, the average of 2014 and 2016 count-based AADT are also considered in the modeling and assessment process. The final database includes count-based AADT for 36,957 traffic count stations in 100 counties. The descriptive statistics by the AADT range are summarized in Table 2.

Table 2 Descriptive statistics by AADT range

AADT range	# of samples	Min.	Median	Mean	Max.	Std. dev.	Frequency Distribution
<5,000	24,444	10	1,000	1,518	5,000	1,342	

5,000-10,000	5,641	5,050	7,200	7,373	10,000	1,502	
10,000–20,000	4,167	10,100	14,000	14,468	20,000	2,813	
20,000-30,000	1,466	20,500	24,000	24,594	30,000	2,791	
>30,000	1,239	30,500	42,000	53,430	182,000	28,850	

From Table 2, the count-based AADT ranges from 10 to 182,000 in the state of North Carolina. Around 67% of the count-based AADT values are lower than 5,000. The skewness in data distribution can be observed from the distribution plots in Table 2. Therefore, it is recommended to use the median as the measure of central tendency. Further, the count-based AADT for local roads were segregated from the database. The distribution of count-based AADT data for the local roads is shown in Figure 14.

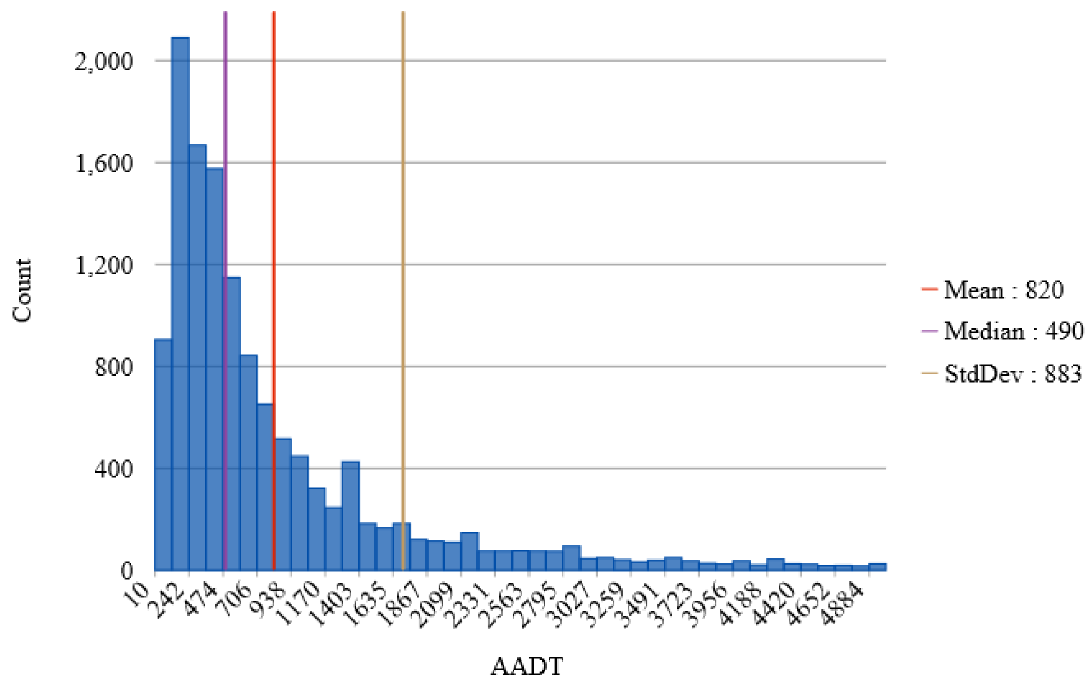


Figure 14 Frequency distribution of count-based local road AADT

5.2 Functional Classification Type

The descriptive statistics of count-based local road AADT by the functional classification type are summarized in Table 3.

Table 3 Count-based local road AADT by functional classification type

Functional classification type	# of samples	Minimum	Median	Mean	Maximum	Standard deviation
Urban	3,035	40	1,200	1,504	5,000	1,201
Rural	9,864	10	413	609	5, 000	623

The functional classification of most of the local road traffic count station is rural. They account for about 76% of the total local road traffic count stations. The median count-based AADT is 1,200 and 413 for urban and rural local roads, respectively. A higher standard deviation is observed in the case of rural local road count-based AADT.

5.3 Speed Limit

The count-based local road AADT data were classified based on the speed limit and are summarized in Table 4. From the road database, most of the rural local roads have a speed limit of 55 mph. However, the speed limit of local urban roads, where there is higher count-based AADT, has a speed limit of 35 mph. Approximately, 70% of the local road links have a speed limit of 55 mph.

To better understand the relationships, the speed limit-based dataset was subdivided into urban and rural local roads. The results for urban and rural local roads by the speed limit are summarized in tables 5 and 6.

Table 4 Count-based local road AADT by the speed limit

Speed limit (mph)	# of samples	Minimum	Median	Mean	Maximum	Standard deviation
<=25	357	40	630	984	4,800	996
30 or 35	2,279	40	910	1,285	5,000	1,125
40 or 45	1,878	75	1,000	1,382	5,000	1,105
50 or 55	8,385	10	380	560	5,000	584

Table 5 Count-based urban local roads AADT by the speed limit

Speed limit (mph)	# of samples	Minimum	Median	Mean	Maximum	Standard deviation
<=25	204	80	662	940	4,300	881
30 or 35	1,217	60	1,300	1,648	4,950	1,237
40 or 45	763	75	1600	1,905	5,000	1,207
50 or 55	851	40	1,400	1,075	5,000	1,017

Table 6 Count-based rural local roads AADT by the speed limit

Speed limit (mph)	# of samples	Minimum	Median	Mean	Maximum	Standard deviation
<=25	153	40	620	1,044	4,800	1,135
30 or 35	1062	40	605	870	5,000	803
40 or 45	1,115	80	730	1,024	5,000	864
50 or 55	7,534	10	360	502	4,900	479

The urban local road links with a speed limit of 25 mph have the lowest median count-based AADT. Contrarily, the rural local road links with a speed limit of 55 mph have the lowest median count-based AADT. The standard deviation was observed to be the highest for rural local roads links with a speed limit of less than or equal to 25 mph.

5.4 Population Density

The descriptive statistics based on population density are summarized in Table 7. The population density was estimated based on TAZ-level data for the year 2015. Approximately, 67% of count-based local road AADT are in areas with a population density of fewer than 200 people per square mile. The count-based local road AADT was observed to increase with an increase in population density.

Table 7 Count-based local road AADT by population density

Population density (people/square mile)	# of samples	Minimum	Median	Mean	Maximum	Standard deviation
<200	8,638	10	390	577	5,000	608
200 – 400	2,251	30	800	1,085	5,000	930
400 – 600	923	40	1,000	1,404	5,000	1,183
600 – 800	423	70	1,300	1,600	5,000	1,205
800 – 1,000	227	80	1,400	1,639	4,900	1,229
1,000 – 1,200	121	60	890	1,352	4,900	1,176
1,200 – 1,400	136	70	1,400	1,806	4,900	1,396
1,400 – 1,600	64	320	1,825	2,313	4,900	1,491
1,600 – 2,000	51	105	2,100	2,207	4900	1,338
>2,000	65	70	1,700	1,975	4800	1,344

5.5 Employment Density

Table 8 shows the count-based local road AADT statistics based on employment density. The TAZ-level total employment information was used to estimate employment density. The majority of local road traffic count stations are in areas with low employment density. The median AADT is 432 at locations with an employment density of 100 employees per square mile.

Table 8 Count-based local road AADT by employment density

Employment density (employment/square mile)	# of samples	Minimum	Median	Mean	Maximum	Standard deviation
<100	10,104	10	430	694	5,000	694
100 - 200	1,254	40	822	1,219	5,000	1,094
200 - 300	552	40	962	1,258	5,000	1,053
300 – 400	282	75	1,200	1,511	4,900	1,204
400 – 500	167	80	1,200	1,622	4,900	1,286
500 - 600	132	70	1,200	1,700	4,900	1,360
600 - 700	78	105	1,100	1,521	4,900	1,309
700 - 800	52	170	1,950	2,051	4,900	1,462
800 - 900	54	190	1,425	1,736	4,700	1,133
900 - 1000	47	90	1,600	1,680	4,000	1,248
>1000	177	70	1,800	2,080	4,950	1,475

5.6 Road Density

As land use data could not be explored statewide, the road density was computed and used as an indicator of development. The road density is defined as the mileage of roads within a preset distance (for example, 1-mile) from a traffic count station. The descriptive statistics based on the road density are summarized in Table 9.

Table 9 Count-based local road AADT by road density

Road density (mileage of road/ 1-mile buffer)	# of samples	Minimum	Median	Mean	Maximum	Standard deviation
< 10	5,670	10	340	456	5,000	421
10 – 20	4,724	40	610	893	5,000	842
20 – 30	1,760	40	992	1,375	5,000	1164
30 – 40	615	60	1,500	1,762	4,900	1,273
> = 40	130	120	1,725	2,022	4,900	1,444

5.7 Local Travel Characteristics

In the case of local roads, most travel is oriented from the land being accessed to the nearest nonlocal road. Also, local roads support through traffic from other local roads. Therefore, it is important to investigate the beginning and ending route characteristics of each link before modeling. For example, one of the most common scenarios is dead-end links as shown in Figure 15.

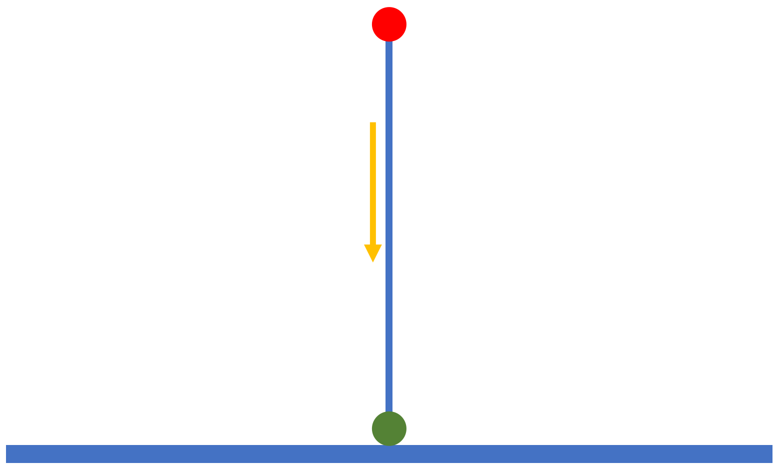


Figure 15 AADT at a dead-end link

The local travel characteristics vary at locations connecting two nonlocal roads. The nonlocal roads with higher AADT typically have a higher level of interaction with local roads. Therefore, the descriptive statistics were developed based on beginning feature and ending feature characteristics and are summarized in Table 10.

Table 10 Count-based local road AADT by beginning and ending feature characteristics

Beginning feature – ending feature	# of samples	Minimum	Median	Mean	Maximum	Standard deviation
Dead-ends (F7)	47	40	130	292	2,450	478
F7 – F7	7,133	10	520	853	5,000	902
F7 – F6/F5	1,346	30	435	695	5,000	747
F7 – F4/F3	690	50	627	1,026	5,000	1,035
F7 – F1/F2	43	200	1,095	1462	4,550	1,139
F6/F5 – F6/F5	81	30	360	695	4,550	870
F6/F5 – F4, F3, F2, F1	45	80	460	884	4,250	964
F1, F2, F3, F4 – F1, F2.F3, F4	25	60	740	1,018	4,400	1,037

Note: F1: Interstate; F2: Principal arterial – other freeways and expressways; F3: Principal arterial; F4: Minor arterial; F5: Major collector; F6: Minor collector; F7: local road.

CHAPTER 6: STATEWIDE LOCAL ROAD AADT MODELING

This chapter covers statewide local road AADT model development and validation details. A Pearson correlation coefficient matrix was developed to evaluate the correlation between explanatory variables. Further, different models were developed based on count-based AADT data, functional classification type, speed limit, and population density. The subset feature in ArcGIS Pro was used to randomly select 75% of the data for modeling and 25% of the data for validation in all modeling scenarios.

6.1 Identifying Potential Explanatory Variables

The potential explanatory variables were identified based on the literature review and surveying other DOTs. The descriptive statistics for all the selected variables are summarized in Table 11.

6.2 Pearson Correlation Coefficient Analysis

In this research, the coefficient analysis was performed by computing Pearson correlation coefficients. The correlation analysis was carried out separately for all data, functional classification type, and speed limit ranges.

6.2.1 All data

Table 12 summarizes the Pearson correlation coefficients between count-based local road AADT and road characteristics. The results indicate that road density, functional classification type, and the nearest AADT nonlocal road have a positive correlation with

Table 11 Descriptive statistics - selected explanatory variables

Variables	Minimum	Maximum	Mean	Median	Std. Dev.
Count-based AADT	10	5,000	820	490	883
# of lanes	1	4	2	2	-
Speed limit (mph)	20	55	49	55	9
Dead-end	0	1	0.004	0	-
Surface type indicator (unpaved)	0	1	0.007	0	-
Surface type indicator (Bitumen)	0	1	0.861	0	-
Surface type indicator (Concrete)	0	1	0.129	1	-
Population	0.21	219.65	8.78	4.43	11.83
# of households	0.11	68.97	3.48	1.74	4.68
Workers	0	79.52	4.15	2.03	5.72
Industrial workers	0	46.20	0.60	0.11	1.99
Heavy industrial Workers	0	23.48	0.38	0.11	1.07
Retail workers	0	54.72	0.41	0.07	1.50
High retail employees	0	60.86	0.36	0.05	1.15
Office employees	0	112.26	0.57	0.08	2.50
Service employees	0	72.63	1.11	0.23	2.94
Government employees	0	64.38	0.30	0.04	1.81
Educational employees	0	298.46	0.34	0.07	2.80
Urban local road	0	1	0.23	0	-
Rural local road	0	1	0.76	0	-
Population density	2.37	5,798.79	231.86	116.95	312.17
Employment density	0	14,347.69	106.86	28.27	311.01
Road density (1-mile)	2.00	74.00	13.70	11.10	8.40
Distance to the nearest nonlocal road (miles) (Dis-nonlocal)	0.010	9.48	0.54	0.21	0.77
AADT at the nearest nonlocal road (AADT-nonlocal)	240	119,000	7,000	4,400	7,908

Note: Socioeconomic variables were extracted using a 100 feet flat buffer

Count-based local road AADT. In general, local roads are designated for land access. Most travel is oriented from land access to the nearest nonlocal road. Hence, nonlocal roads with higher AADT typically have a higher level of interaction with local roads.

Moreover, local functionally classified roads within the vicinity of higher functionally classified roads will have a higher count-based AADT. The positive correlation between count-based local road AADT and nearby nonlocal road AADT and the negative correlation between the distance to the nearest higher functionally classified road and count-based local road AADT substantiate the same.

Contrarily, there is a negative correlation between local road AADT and speed limit. From the road database, most rural local roads have a speed limit of 50 mph or 55mph. However, urban local roads with a lower speed limit have a higher AADT. The negative correlation between local road AADT and speed limit can be attributed to this factor. The presence of dead-ends also has a negative correlation with local road AADT.

The correlation analysis was carried out for explanatory variables extracted using 50 feet, 330 feet, 660 feet, and 1,320 feet buffer widths. Smaller buffer widths were found to be adequate to capture the socioeconomic variables within the vicinity of a local road. Hence, a 100 feet buffer width was considered acceptable for model development and validation. Table 13 summarizes the Pearson correlation coefficients between count-based local road AADT, and socioeconomic variables extracted using the 100 feet buffer width.

The population, workers, service employees, population density, and employment density were observed to have a statistically significant relationship with count-based local road AADT. Similarly, a high positive correlation (multicollinearity) between population density and other employment categories led to the exclusion of some of these explanatory variables in the final model development. The backward elimination approach was adopted to identify the best-suited variables for modeling.

6.2.2 Functional classification type

The speed limit and distance to the nearest nonlocal road have a low negative correlation with urban count-based local road AADT. Explanatory variables such as road density, population density, employment density, count-based AADT at the nearest nonlocal road, and employment categories have a low positive correlation with count-based urban local road AADT. Multicollinearity between employment categories and population density was observed from the analysis.

The road density and population density have a medium positive correlation with rural count-based local road AADT, whereas the distance to the nearest nonlocal road and speed limit has a low negative correlation with count-based rural local road AADT. The results are shown in Table 14.

6.2.3 Speed limit

The available count-based AADT data was divided into four categories based on speed limit ranges. In the case of local roads with speed limits less than or equal to 25 mph, road density, distance to the nearest nonlocal road, and the number of service employees were observed to have a significant effect on count-based local road AADT.

In the case of local roads with a speed limit greater than 25 mph and less than or equal to 35 mph, road density, population density, and employment density have a medium positive correlation with count-based local road AADT. The distance to the nearest nonlocal road has a negative effect on count-based local road AADT for the same category.

For other speed limit ranges, road density, count-based AADT at the nearest nonlocal road, and employment categories such as office and service have a significant correlation with count-based local road AADT. The results are shown in Table 14.

6.2.4 Population density

The count-based AADT database was divided into five categories based on population density. In the case of population density less than 200 people/square mile, road density, employment density, different employment categories have a positive correlation with count-based local road AADT. However, the distance to the nearest nonlocal road has a negative correlation with count-based local road AADT. The results are summarized in Table 14.

The Pearson correlation coefficient matrices related to functional classification type, speed limit, and population density are shown in Appendix A.

Table 12 Correlation between count-based AADT and road characteristics

Attributes	Local road AADT	Speed limit	# of lanes	Func. class. type	Unpaved	Bitumen	Concrete	Road density	Dis-nonlocal	AADT-nonlocal
Speed limit	MN									
# of lanes	LP	LN								
Func. class. type	MP	MN	LP							
Unpaved	LN	LP								
Bitumen	LP	LN	LP	LP	LN					
Concrete	LN	LP	LN	LN	LN	HN				
Road density	MP	HN	LP	HP		LP	LN			
Dis-nonlocal	LN	LP		LN		LN	LP	LN		
AADT-nonlocal	MP	LN	LP	MP	LP	LP	LN	MP	LN	
Dead-end	LN									

Note 1: Dis-nonlocal: Distance to the nearest higher functional class road (miles)
Note 2: AADT-nonlocal: Count-based AADT at the nearest nonlocal road
Note 3: HP, MP, LP, HN, MN, and LN are high positive, moderate positive, low positive, high negative, moderate negative, and low negative, respectively.

Table 13 Correlation between local road AADT and socioeconomic variables – 100 feet buffer width

Attributes	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)
AADT (1)														
Population (2015) (2)	MP													
# of Households (3)	MP	HP												
Workers (4)	MP	HP	HP											
Industrial (5)	LP	MP	MP	LP										
High industrial (6)	LP	MP	MP	MP	MP									
Retail (7)	LP	MP	HP	MP	MP	MP								
High retail (8)	LP	HP	HP	HP	MP	MP	HP							
Office (9)	LP	MP	MP	MP	MP	HP	HP	HP						
Service (10)	MP	HP	MP	HP	MP	HP	HP	HP	HP					
Government (11)	LP	MP	MP	LP	LP	LP	LP	HP	MP	HP				
Education (12)	LP	MP	LP	LP	LP	LP	LP	HP	MP	LP	LP			
Population density (13)	MP	HP	HP	HP	MP	MP	MP	HP	MP	HP	MP	MP		
Employment density (14)	LP	HP	HP	HP	HP	HP	HP	HP	HP	HP	HP	HP	HP	

Note 1: Socioeconomic variables were extracted using 100 feet flat buffer
Note 2: HP, MP, LP, HN, MN, and LN are high positive, moderate positive, low positive, high negative, moderate negative, and low negative, respectively.

Table 14 Correlation analysis summary for Functional classification type, speed limit, and population density

Parameters	Functional classification type		Speed limit (mph)				Population density (people/square mile)				
	Urban	Rural	<= 25	30 or 35	40 or 45	50 or 55	<200	200 - 400	400 - 600	600 - 800	>800
Speed Limit	LN	LN	MN	LP	LP	LN	LN	LN	LN	LN	-
# of Lanes	LP	LN	-	LP	LN	LN	LP	-	-	-	LP
Area type			-	MP	MP	LP	LP	LP	LP	LP	-
Road density	LP	MP	LP	MP	MP	MP	MP	LP	LP	LP	LP
Dis-nonlocal	LN	LN	LN	LN	LN	LN	LN	LN	LN	-	-
AADT-nonlocal	LP	LP	-	LP	LP	LP	LP	LP	LP	LP	LP
Population (2015)	LP	MP	-	MP	MP	MP	LP	LP	LP	-	LP
# of Households	LP	LP	-	MP	MP	MP	LP	LP	LP	-	LP
Workers	LP	MP	-	MP	MP	MP	LP	LP	LP	-	LP
Industrial	LP	LP	-	LP	LP	LP	LP	LP	LP	-	-
High industrial	LP	LP	-	LP	LP	LP	LP	LP	LP	LP	-
Retail	LP	LP	-	LP	LP	LP	LP	LP	LP	-	-
High retail	LP	LP	-	MP	LP	LP	LP	LP	LP	LP	LP
Office	LP	LP	-	LP	LP	LP	LP	LP	LP	-	-
Service	LP	LP	LP	MP	LP	LP	LP	LP	LP	-	LP
Government	LP	LP	-	LP	LP	LP	LP	-	-	-	-
Education	LP	LP	-	LP	LP	LP	LP	LP	LP	-	-
Population density	LP	MP	-	MP	MP	MP	LP	LP	LP	-	LP
Employment density	LP	LP	-	MP	LP	LP	LP	LP	LP	-	LP

Note 2: HP, MP, LP, HN, MN, and LN are high positive, moderate positive, low positive, high negative, moderate negative, and low negative, respectively

6.3 Model Development

OLS regression and geospatial methods such as GWR, Kriging, IDW, and natural neighbor interpolation methods were explored to estimate AADT on local roads. The geospatial methods assume that locations with AADT counts close to one another are alike. The level of correlation reduces with an increase in the distance between these locations. The predictability of the geospatial methods depends on the density and spatial distribution of data points. A comparison of the OLS regression model and selected geospatial techniques was performed initially using all data. One statistical model and one geospatial model was selected from the preliminary analysis. Models were then developed by functional classification type, speed limit, and population density ranges.

6.3.1 Ordinary least square (OLS) regression model

The OLS regression model was used as the base model for all the geospatial models developed in this research. It helps to identify spatial patterns or spatial relationships. The backward elimination approach was used to exclude statistically insignificant explanatory variables when developing the best model. Akaike information criterion (AIC) and R-square were used to test the goodness-of-fit. The best-fitted model details are summarized in Table 15. The results indicate that speed limit, distance to the nearest nonlocal road, office, government, and dead-ends have a negative influence on count-based local road AADT. Similarly, road density, AADT at the nearest nonlocal road, industrial employees, and population density have a positive influence on count-based local road AADT.

The validation was carried out using 25% of the data. The MAPE, MPE, and RMSE for the validation dataset are 86.1, -44.2, and 771, respectively based on the best fitted OLS regression model.

6.3.2 Geographically weighted regression

The significant explanatory variables from the OLS regression model were used to develop the GWR model. The GWR builds a local regression equation for each feature in the dataset. However, when the values of an explanatory variable cluster spatially, problems of multicollinearity may arise in the GWR model. The dummy variables were removed from the model as there is a higher chance of local model failure with binary explanatory variables. Table 16 summarizes the results from the GWR model. The optimum bandwidth is identified by minimizing the AIC value. The optimized AIC is 6658. Similarly, the estimated R-square is 0.44 while the estimated MAPE, MPE, and RMSE for the validation dataset are 82.1, -42.1, and 730, respectively.

Table 15 Statewide OLS model

Parameters	Coefficient	Standard error	p-value
Intercept	2.727	0.031	<0.05
Speed limit	-0.005	<0.001	<0.05
Road density	0.011	<0.001	<0.05
Dis-Nonlocal	-0.049	<0.001	<0.05
AADT- Nonlocal	8*10-6	<0.001	<0.05
Industrial	0.009	<0.001	<0.05
Office	-0.009	<0.001	0.051
Government	-0.004	<0.001	<0.05
Population density	2.2*10^-4	<0.001	<0.05
Dead-end	-0.58733	0.056	<0.05
R-square	0.26		
AIC	7691		
MAPE	86.1%		
MPE	-44.2%		
RMSE	771		

Table 16 Statewide GWR model

Parameters	Minimum	Median	Mean	Maximum	Standard deviation
Intercept	1.061	2.724	2.708	3.9	0.43
Speed limit	-0.022	-0.005	-0.005	0.026	0.007
Road density	-0.014	0.014	0.014	0.053	0.01
Dis-Nonlocal	-0.333	-0.04	-0.044	0.132	0.058
AADT-Nonlocal	-2.4*10-5	7.22*10-6	7.92*106	6.69*10-5	8.67*10-6
Industrial	-1.355	0.009	0.003	1.049	0.117
Office	-1.298	-0.008	-0.027	0.739	0.15
Government	-1.472	-0.004	-0.022	0.71	0.153
Population density	-2.3*10-3	2.4*10-4	4.15*10-4	8.6*10-3	7.2*10-3
R-square	0.44				
AIC	6658				
# of neighbors	254				
MAPE	82.1				
MPE	-42.1				
RMSE	730				

The spatial variation in the coefficients for the entire study area is shown in Figure 16. The influence of each selected explanatory variable differs throughout the state. The coefficient of the intercept varies from 1.061 to 3.9 for the study area.

(a) Intercept

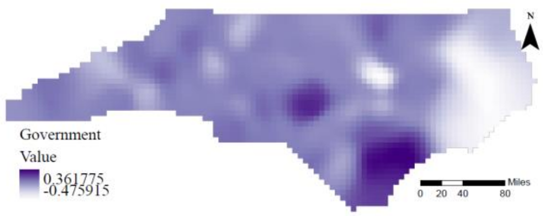
(b) Speed limit

(c) Road Density

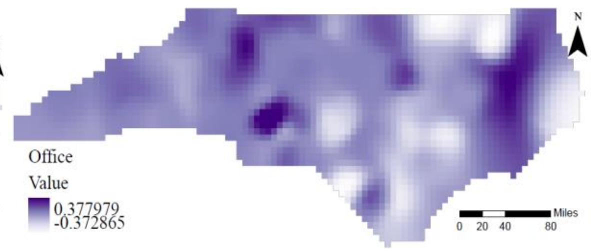
(d) AADT-nonlocal

(e) Dis-nonlocal

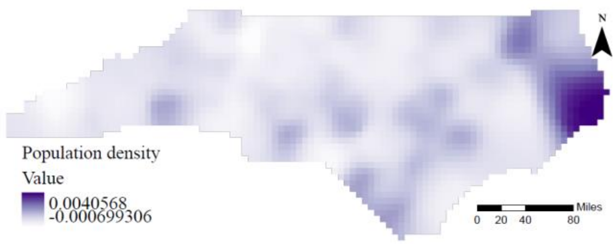
(f) Industrial employees



(g) Governemnt



(h) office



(i) Population density

Figure 16 Spatial variations in coefficients

6.3.3 Kriging

The cross-validation approach identifies the best Kriging model by minimizing the measures of prediction error. The Simple Kriging, Ordinary Kriging, Universal Kriging, and Empirical Bayesian Kriging with different semi-variogram models have been assessed to identify the best Kriging model. Geostatistical Wizard in the ArcGIS Pro was used for the modeling process. The criteria mentioned in Asa et al. (2012) was adopted to find the best model. According to their research, the best Kriging model will have the following properties:

1. A mean prediction error near to zero
2. A standardized mean (SM) prediction error close to zero
3. A small RMSE
4. Standardized root means square error (RMSES) close to one and close to the average standard error (ASE) (Robinson and Metternicht, 2006)

The Empirical Bayesian Kriging with power semi-variogram model was selected as the final model. The cross-validation results are summarized in Table 17.

The raster output from the Empirical Bayesian Kriging model is shown in Figure 17. The raster image is converted into the point dataset. The non-covered location details are spatially joined to the point dataset to estimate local road AADT. The MAPE, MPE, and RMSE for the validation dataset are 84.1%, -44.2%, and 733, respectively (Table 18).

Table 17 Cross-validation results

Measure	Mean	Mean	RMSE	SM	RMSES	ASE
Universal Kriging	Exponential	26.36	721.66	16.67	476.22	1.53
	K-Bessel	12.63	726.64	8.13	499.76	1.46
	Spherical	13.26	726.78	8.52	498.71	1.47
Simple Kriging	Exponential	27.12	722.56	0.06	0.79	964.23
	K-Bessel	33.46	724.44	0.06	0.77	1007.19
	Spherical	19.67	732.32	0.06	0.79	949.9
Ordinary Kriging	Exponential	26.36	721.66	0.05	0.77	1013.45
	K-Bessel	12.63	726.64	0.04	0.79	985.88
	Spherical	13.26	726.78	0.04	0.79	988.45
Empirical Bayesian Kriging	Power	13.05	714.81	0.02	0.95	739.13
	Linear	13.37	720.69	0.02	0.95	743.33

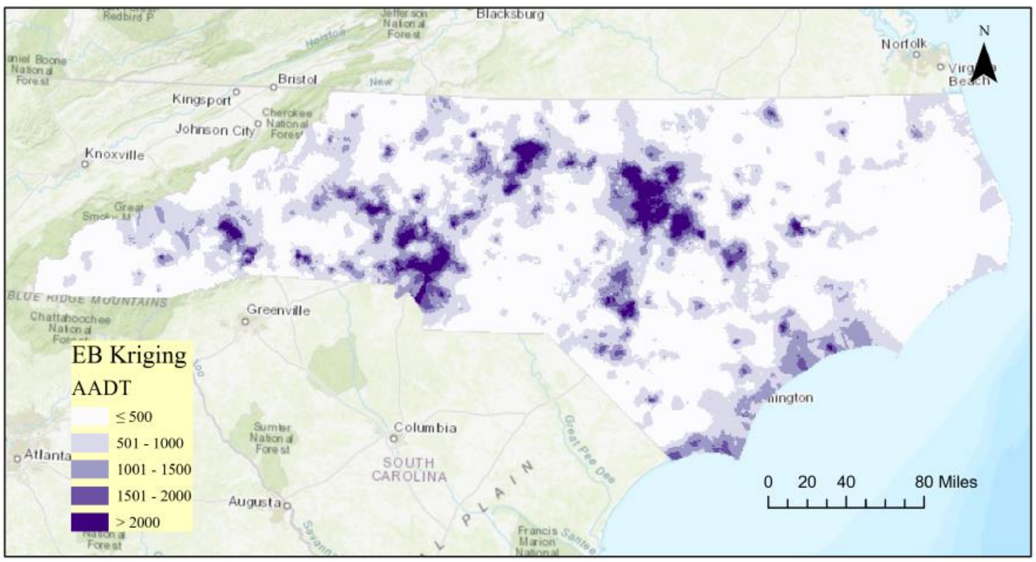


Figure 17 Raster output from Empirical Bayesian Kriging model

6.3.4 Inverse distance weighted (IDW)

To estimate AADT at any non-covered location, IDW uses the count-based AADT values surrounding the prediction location. The count-based AADT closest to the prediction location have more influence on the estimated value than those farther away.

The distance weights are assigned by the second-order power function. The raster image used for estimating local road AADT using the IDW method is shown in Figure 18. The MAPE, MPE, and RMSE are 120.9%, -96.8%, and 726, respectively (Table 18).

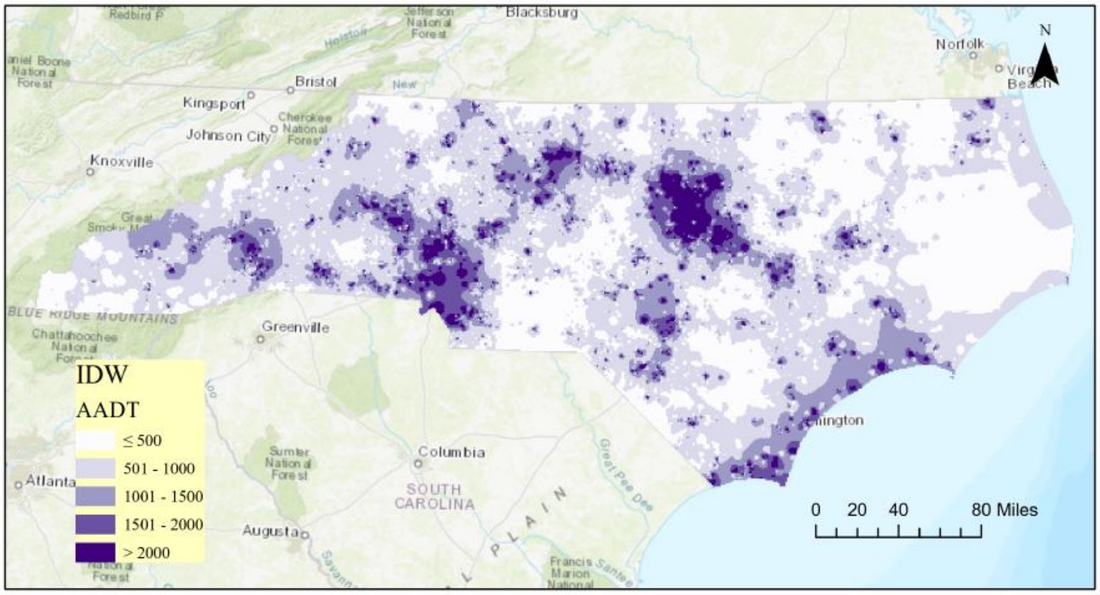


Figure 18 Raster output from IDW model

6.3.5 Natural neighbor interpolation

This method also interpolates a raster surface from traffic count stations using a natural neighbor technique. The raster output form natural neighbor interpolation modeling is shown in Figure 19. The validation results are shown in Table 18. The MAPE, MPE, and RMSE are 89.2%, -47.2%, and 743, respectively (Table 18).

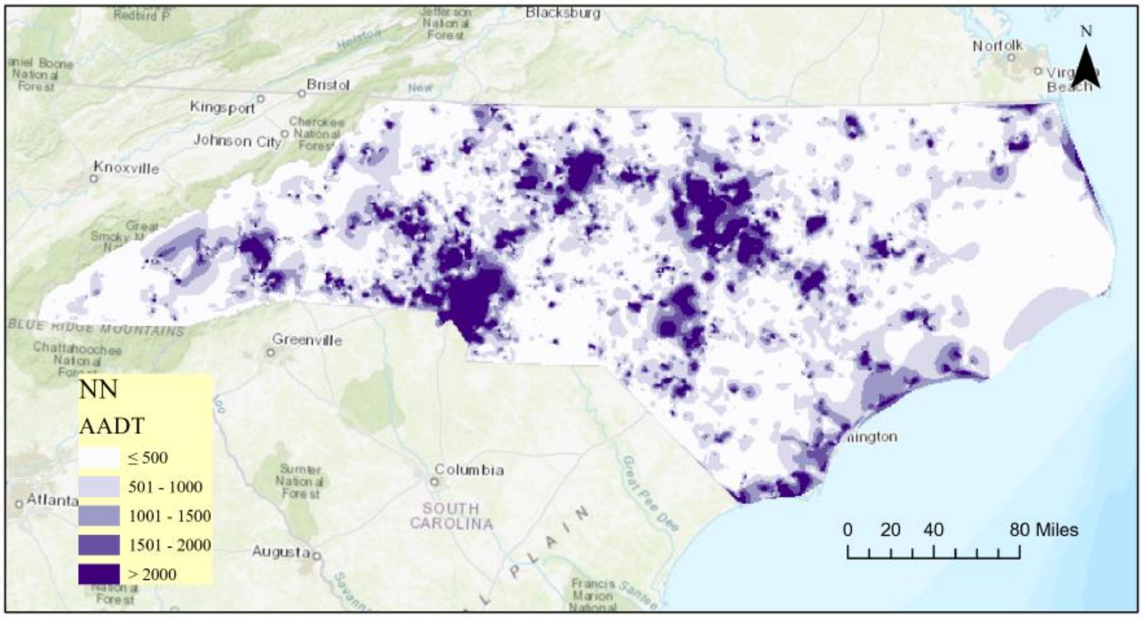


Figure 19 Raster output from natural neighbor interpolation model

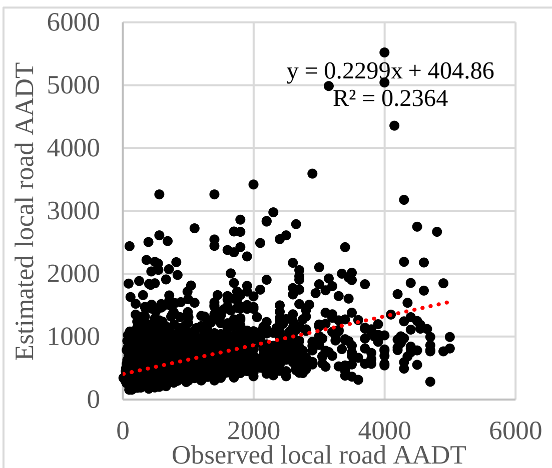
6.3.6 Comparison of models to estimate local road AADT

The validation results for all the selected models are summarized in Table 18 for easy comparison. When comparing OLS regression and geospatial methods, GWR performed better in terms of AIC, R-square, MAPE, MPE, and RMSE values. It indicates that the geospatial methods such as GWR can accommodate the spatial variation in data better than OLS regression model.

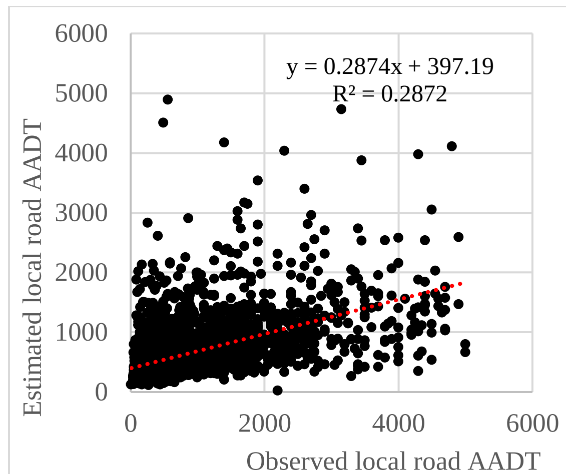
Table 18 Validation results for statewide modeling

Measure	OLS	GWR	Kriging	IDW	NN
MAPE (%)	86.1%	82.1	84.1	120.9	89.2
MPE (%)	-44.2%	-42.1	-44.2	-96.8	-47.2
RMSE	771	733	733	726	743

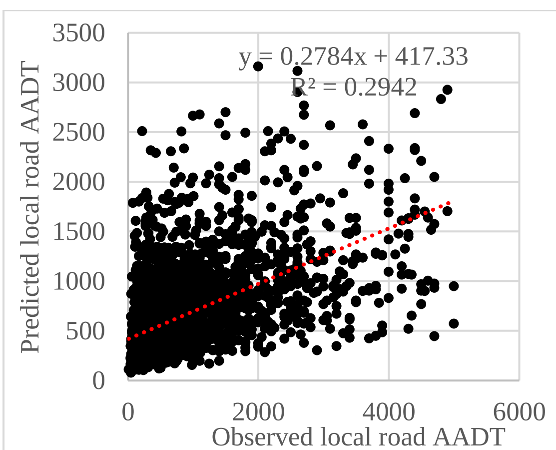
In other words, the GWR is a local regression model in which a certain number of count-based AADT values around the non-covered location where AADT to be calculated are used to fit the model, and the distance between the count-based AADT station and the non-covered location to be calculated is used as the weight. The GWR model is more suitable for estimating the local road AADT than the OLS regression model in a statewide modeling approach. Similarly, the Empirical Bayesian Kriging method outperformed IDW and NN when considering all three validation parameters. While comparing GWR and Empirical Bayesian Kriging methods, both the methods performed similarly in estimating local road AADT. Figure 20 shows the relationship between the observed and estimated local road AADT for each modeling approach.



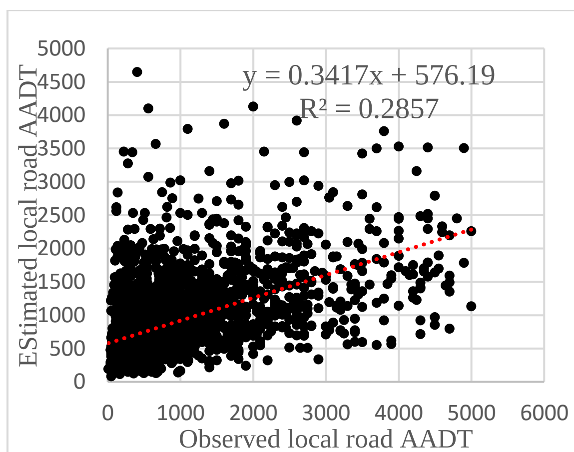
(a) OLS



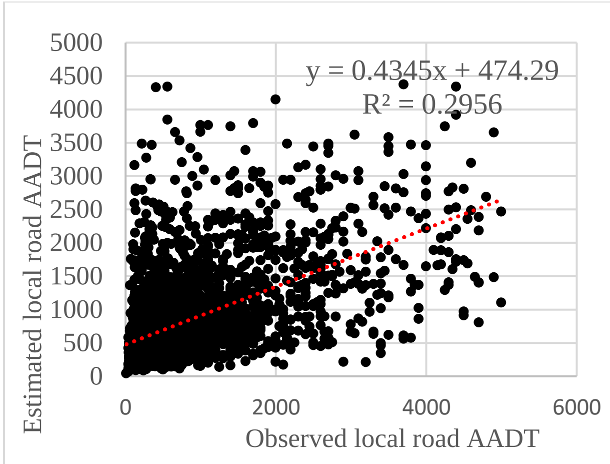
(b) GWR



(c) Empirical Bayesian Kriging



(d) IDW



(e) Natural neighbor

Figure 20 Relationship between observed and estimated AADT

The interpolation models are based on the autocorrelation of the local road AADT, while the OLS regression model is based on the correlation of local road AADT with other factors. Moreover, while looking into all the non-covered locations in North Carolina, it is essential to consider the factors/variables to include in the model to make logical predictions. For example, the roads which are nearby with different speed limits will have different characteristics and different local road AADT. Hence, it is essential to consider such variables in the estimation process rather than only relying on spatial autocorrelation. The disaggregate-level model in this research is further performed using GWR. The OLS regression models are also developed to identify the statistically significant variables (also used for developing GWR models) influencing local road AADT.

6.4 Disaggregate Level Modeling

The models developed based on functional classification type (urban/rural local road), speed limit, and population density are summarized in the following subsections. Explanatory variables selected to develop models by functional classification type, speed limit, and population density are summarized in Table 19.

6.4.1 Functional classification type

Explanatory variables such as road density, distance to nearest nonlocal road, AADT at the nearest nonlocal road, service, and population density influence urban local road AADT at a 95% confidence level (p-value <0.05). Similarly, speed limit, distance to nearest nonlocal road, AADT at the nearest nonlocal road, office, industrial, government, and population density influence rural local road AADT at a 95% confidence level. The results from the model validation are summarized in Table 20. They indicate that the predictability of rural local roads AADT model performs better than the urban local roads AADT model. The range of urban local roads AADT is lower than the range of rural local roads AADT. As observed previously, the GWR models can incorporate the effect of spatial attributes by geographic location better than OLS regression models.

Table 19 Explanatory variables selected to model by functional classification type, speed limit, and population density

Parameters	Functional classification type		Speed limit (mph)				Population density (people/square mile)				
	Urban	Rural	<= 25	30 or 35	40 or 45	50 or 55	<200	200 - 400	400 - 600	600 - 800	>800
Speed Limit	√	√					√	√	√		
# of Lanes											
Area type								√	√		
Road density	√	√	√	√	√	√	√	√	√	√	√
Dis-nonlocal	√	√	√	√		√	√	√	√		
AADT-nonlocal	√	√		√	√	√	√	√	√		√
Population (2015)								√	√		
# of Households											
Workers								√	√		
Industrial	√	√					√				
High industrial										√	
Retail										√	
High retail	√										
Office		√			√	√					
Service		√		√		√					
Government	√										
Education							√	√	√		
Population density	√	√									
Employment density					√						

Table 20 Validation results for models based on functional classification type

Functional classification type	OLS			GWR		
	MAPE (%)	MPE (%)	RMSE	MAPE (%)	MPE (%)	RMSE
Urban	119.1	-65.1	1359	110	-64.2	1154
Rural	73.1	-28.33	636	72.1	-27.3	596

6.4.2 Speed limit

The database was divided into four categories based on speed limit: the speed limit is less than or equal to 25 mph, the speed limit is equal to 30 or 35 mph, the speed limit is equal to 40 or 45 mph, the speed limit is equal to 50 or 55 mph. The OLS regression and GWR models were developed and compared for each speed limit category. The results from the model validation are summarized in Table 21. They indicate that local roads with a speed limit equal to 50 or 55 mph performed better than other speed limit categories.

Table 21 Validation results for models based on the speed limit

Speed limit (mph)	OLS			GWR		
	MAPE (%)	MPE (%)	RMSE	MAPE (%)	MPE (%)	RMSE
<25	91.32	-34.77	1071	92.40	-38.31	1057
30 or 35	106.61	-64.43	1167	107.23	-67.25	1135
40 or 45	78.30	-39.33	960	82.23	-46.42	936
50 or 55	82.71	-40.18	674	80.73	-40.09	574

6.4.3 Population density

The database was divided into four categories based on the population density. The OLS regression and GWR models were developed and compared for each population density category. The results obtained from the OLS regression model and GWR model validation are summarized in Table 22. The models for population density in areas with less than 200 people per square mile performed better than other selected categories.

Table 22 Validation results for models based on population density

Population density (population / square mile	OLS			GWR		
	MAPE (%)	MPE (%)	RMSE	MAPE (%)	MPE (%)	RMSE
<200	80.81	-38.06	627	75.18	-34.63	579
200 - 400	95.66	-48.29	944	97.58	-53.08	907
400 - 600	84.95	-43.74	829	85.19	-45.66	795
600 - 800	112.10	-53.56	1461	120.12	-64.37	1392
800-1000	126.2	-108.39	1418	132.68	-124.27	1366

CHAPTER 7: COUNTY-LEVEL LOCAL ROAD AADT MODELING

This chapter presents the results from the county-level statistical and geospatial models. The process involved identifying variables, performing Pearson correlation coefficient analysis, developing models, and validation is explained by selecting Duplin County (rural) and Wake County (urban) as examples.

7.1 Descriptive Statistics

Ten counties were considered for modeling based on the quality of land use data, population density, road density, and the number of local road traffic count stations available in the county. These counties are spatially distributed in the state of North Carolina. They represent all three regions in the state- coastal plain, piedmont, and mountains.

The raw dataset consists of several land use categories. As the count-based local road AADT data was considered for the year 2015, land use developments up until the year 2015 were considered for the model development. The selected counties and their characteristics for county-level modeling are summarized in Table 23.

The population density in the selected counties varied from 72.21 to 1,894.45 people/square mile. The number of local road AADT counts available for modeling ranges from a low of 55 in Mecklenburg County to a high of 295 in Wake County (Table 24). As an example, the spatial distribution of local road AADT count locations in Duplin County and Wake County are shown in Figures 21 and 22. The descriptive statistics such as minimum, median, mean, maximum, and standard deviation of local road AADT are summarized in Table 24.

Table 23 Selected counties for county-level modeling

County	Area (Square miles)	Road length (Miles)	Road density (Length / Square Miles)	Population (2015)	Population density – (2015)
Buncombe	659.67	3,450.11	5.27	253,178	383.79
Columbus	953.16	1706.46	1.79	56,694	59.48
Dare	1248.63	857.23	0.69	35,663	28.56
Davidson	567.52	2833.21	4.99	164,622	290.07
Duplin	819.27	1650.16	2.01	59,159	72.21
Iredell	596.71	2,515.19	4.22	169,866	284.67
Mecklenburg	545.84	5,221.07	9.57	1,034,070	1894.45
Randolph	790.11	2,452.30	3.10	142,799	180.73
Wake	856.24	6,445.37	7.53	1,024,198	1196.15
Wayne	556.98	1,771.31	3.18	124,132	222.86

Table 24 Descriptive statistics – selected counties

County	# of local road traffic count stations	Minimum local road AADT	Median local road AADT	Mean local road AADT	Maximum local road AADT	Std. deviation of local road AADT
Buncombe	217	910	160	1,273	4,400	1,025
Columbus	203	40	430	580	3,700	551
Dare	59	60	560	807	4,300	823
Davidson	204	60	672	922	4,500	846
Duplin	235	90	470	608	2,750	456
Iredell	266	60	590	1061	4900	1118
Mecklenburg	55	60	1,450	1,547	4,350	1,200
Randolph	280	25	565	823	4,200	782
Wake	295	50	1,300	1,725	5,000	1,288
Wayne	192	60	697	1,002	4,900	907

7.2 Identifying the Explanatory Variables

The explanatory variables were extracted by generating 100 feet buffers along each subject local road link, as mentioned in the "Methodology" chapter. The descriptive statistics for the selected explanatory variables of Duplin County and Wake County are shown in Table 25 and Table 26, respectively.

7.3 Correlation Assessment

The correlation analysis was performed by computing Pearson correlation coefficients. The computed Pearson correlation coefficient matrices for Duplin County and Wake County are shown in Table 27 and Table 28, respectively.

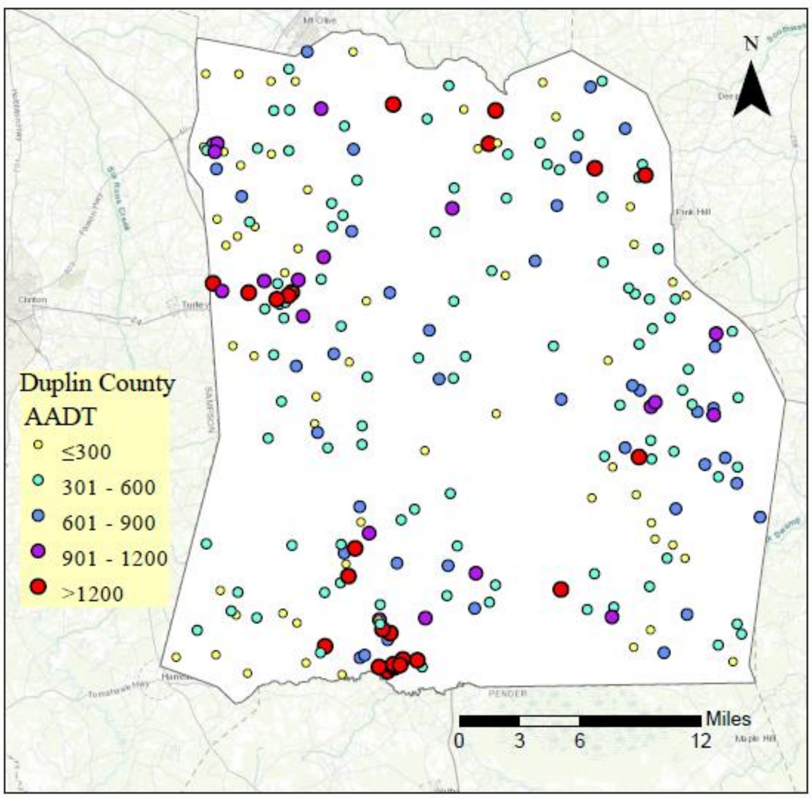


Figure 21 Spatial distribution of local road traffic count stations in Duplin County

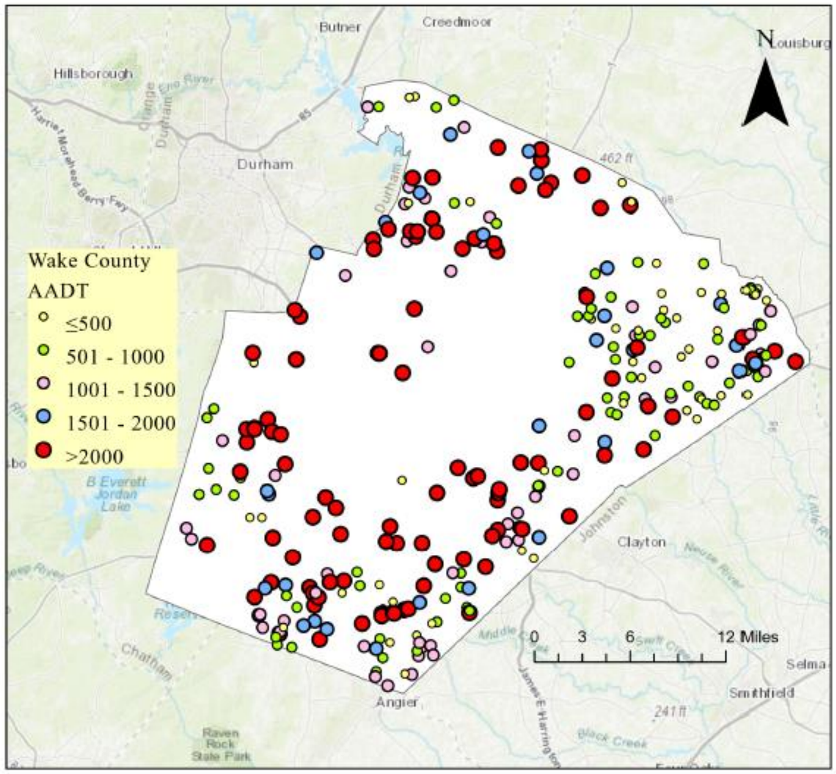


Figure 22 Spatial distribution of local road traffic count stations in Wake County

In the case of Duplin County, road density, AADT at the nearest nonlocal road, population, and different employment categories have a positive correlation with count-based local road AADT. The speed limit and distance to the nearest nonlocal road have a negative correlation with count-based local road AADT. The Pearson correlation coefficient for the number of single-family residential units with count-based local road AADT is relatively small (compared to road density and the speed limit) even though it is significant at a 95% confidence level. Similarly, commercial land use has a positive correlation with count-based local road AADT.

In the case of Wake County, agricultural land use and single-family residential land use have a positive correlation with count-based local road AADT. However, the road characteristics were found to have a significantly higher influence on count-based local road AADT. For example, agriculture and single-family residential units were only found to be the significant land use variables of the fourteen land uses considered for modeling.

Table 25 Descriptive statistics of the explanatory variables – Duplin County

Variables	Minimum	Median	Mean	Maximum	Std. Deviation
Speed limit (mph)	20	55	52.46	55.00	6.70
Functional class type	0	0	0.0043	1	-
Road density	2.00	7.48	10.17	37.68	7.14
Dis-nonlocal (miles)	0.02	0.35	0.87	5.15	1.05
AADT-nonlocal	390	2,700	3,659	23,000	3,365
Socioeconomic variables					
Population	1.50	2.86	3.14	11.33	1.44
# of households	0.60	1.18	1.23	4.51	0.56
Workers	0.65	1.42	1.39	4.78	0.61
Industrial	0	0.13	0.47	2.84	0.63
Hi-industrial	0	0.11	0.11	0.57	0.11
Retail	0	0.06	0.08	0.56	0.08
Hi-retail	0	0.10	0.11	0.30	0.09
Office	0	0.13	0.22	1.12	0.24
Service	0	0.19	0.30	1.56	0.31
Government	0	0	0.06	0.68	0.13
Education	0	0.05	0.10	0.70	0.14
Population density	39.63	75.61	82.92	299.33	38.25
Employment density	3.71	25.85	38.89	116.74	33.06
Land use					
# of multi-family units	0	1	3	32	4
# of single-family units	0	9	12	68	12
Commercial area	0	0	381.16	753.85	988.45
Vacant area	0	404.94	404.58	746.14	170.92

Note: Land use categories' areas are expressed in per 1,000 square feet

Table 26 Descriptive statistics of the explanatory variables – Wake County

Variables	Minimum	Median	Mean	Maximum	Std. Deviation
Speed limit (mph)	20	45	45.73	55	8.17
Functional class type	0	1	0.89	1	0.31
Road density	3.73	18.21	20.27	50.58	8.99
Dis-nonlocal (miles)	0.01	0.10	0.30	2.67	0.45
AADT-nonlocal	430	7,000	11,471	151,000	14,152
Socioeconomic variables					
Population	1.00	29.75	31.11	115.75	19.91
# of households	1.00	10.93	11.73	51.59	7.75
Workers	0.06	14.80	16.38	70.36	10.81
Industrial	0	0.09	0.65	22.72	2.19
Hi-industrial	0	0.35	1.16	23.04	2.95
Retail	0	0.50	1.33	35.37	3.34
Hi-retail	0	0.28	1.00	16.12	1.88
Office	0	0.89	2.04	64.21	6.16
Service	0	1.60	3.59	72.63	7.65
Government	0	0.09	0.40	9.10	1.13
Education	0	0.47	0.87	5.77	1.10
Population density	2.83	785.64	837.20	3,055.99	526.55
Employment density	4.23	133.99	299.45	7,582.65	683.96
Land use					
# of single-family units	0	19	26	125	23
# of multi-family units	0	0	2	62	8
Agricultural area	0	0	125.12	731.72	190.60
Commercial area	0	0	0.38	0.75	0.99
Industrial area	0	0	20.17	555.99	72.59
Institutional area	0	0	15.26	343.09	54.58
Office area	0	0	14.76	740.08	73.97
Resource area	0	0	0.63	76.26	5.85
Retail area	0	0	13.44	342.13	48.79
School area	0	0	3.59	304.49	25.21
Vacant area	0	81.28	122.04	578.68	135.93

Table 27 Pearson correlation coefficient matrix - Duplin County

Variable	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)	(17)	(18)	(19)	(20)
AADT (1)																				
Speed limit (2)	MN																			
Func. class type (3)		LN																		
Road density (4)	MP	HN																		
Dis-nonlocal (miles) (5)	LN	LP		MN																
AADT-nonlocal (6)	LP	LN		MP	LN															
Population (7)	LP			MP	LN	LP														
# of households (8)	LP	LP		MP	LN	LP	HP													
Workers (9)	LP			MP	LN	LP	HP	HP												
Industrial (10)								LP												
Hi-Industrial (11)	LP			LP	LN	LP	HP	HP	HP	LP										
Retail (12)	LP	LN		LP			LP	MP	LP	MP	MP									
Hi-Retail (13)	LP	LN		LP	LN	MP	MP	MP	LP	MP	MP	HP								
Office (14)	LP	LN		LP	LN	LP	HP	HP	HP	MP	HP	HP	HP							
Service (15)				LP	LN	LP	MP	MP	MP	MP	HP	MP	HP	HP						
Government (16)					LN		MP	MP	MP		MP	MP	LP	HP	HP					
Education (17)				LP	LN	LP	HP	HP	HP	LN	MP	LP	MP	HP	HP	HP				
Population density (18)	LP			MP	LN	LP	HP	HP	HP		HP	LP	MP	HP	MP	MP	HP			
Employment density (19)	LP			MP	LN	LP	HP	HP	MP	HP	HP	HP	HP	HP	HP	HP	MP	HP		
# of multi-family units (20)			HP																	
# of single-family units (21)	LP	MN	LP	MP	LN															LP
Commercial area (22)	MP	MN		HP	LN	LP	MP	MP	MP		MP	LP	LP	LP	LP		LP	MP	LP	

Note: HP, MP, LP, HN, MN, and LN are high positive, moderate positive, low positive, high negative, moderate negative, and low negative, respectively

Table 28 Pearson correlation coefficient matrix - Wake County

Variable	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)	(17)	(18)	(19)
AADT (1)		LN																	
Speed limit (2)																			
Func. class type (3)	LP	LN																	
Road density (4)	MP	MN	MP																
Dis-nonlocal (miles) (5)	LN			LP															
AADT-nonlocal (6)	LP		LP	LP	LP														
Population (7)	LP	LN	MP	HP		LP													
# of Households (8)	LP	LN	MP	HP	LN	LP	HP												
Population density (9)	LP	LN	MP	HP		LP	HP	HP											
Employment density (10)	LP	LN	LP	MP		MP	LP	MP	LP										
# of multi-family units (11)	LP	LN	LP	LP			LP	LP	LP										
# of single-family units (12)	LP	LN		MP			LP	LP	LP										
Agricultural area (13)	MN	MP		MN		LN	LN	LN	LN	LN	MN	LN							
Commercial area (14)	LP	LN		LP						LP	LN								
Industrial area (15)		LN		LP						LP	LN		LN						
Institutional area (16)				LP			LP	MP	LP										
Office area (17)		LN			LP					LP	LN								
Resource area (18)																			
Retail area (19)	LP	LN		LP	LN		LP	LP	LP				LN						
School area (20)	LP																		

Note: HP, MP, LP, HN, MN, and LN are high positive, moderate positive, low positive, high negative, moderate negative, and low negative, respectively

7.4 Model Development and Validation

Based on the calibration and validation results from the statewide modeling, the OLS and GWR models were selected to estimate local road AADT for the selected counties. In the case of Duplin County, speed limit, road density, distance to the nearest nonlocal road, single-family residential units, AADT at the nearest nonlocal road, commercial area, and vacant area (parcels) are the significant explanatory variables at a 95% confidence level. In the case of Wake County, road density, agricultural land use, and single-family land use are the significant explanatory variables at a 95% confidence level. The predictability of these county-level models is summarized in Table 29.

Table 29 County-level model validation

County	OLS			GWR		
	MAPE (%)	MPE (%)	RMSE	MAPE (%)	MPE (%)	RMSE
Duplin	52.6	-22.2	452	50.1	-19.8	374
Wake	120.0	-88.3	993	120.1	-86.2	962

7.5 Comparison between Statewide Model and County-Level Model

The spatial distribution of count-based local road AADT, descriptive statistics of explanatory variables, and Pearson correlation coefficient matrices for other selected counties are shown in Appendix B. A comparative assessment was carried out between the statewide and county-level model estimates. The MAPE, MPE, and RMSE were computed using the validation datasets and compared for the statewide estimates and the county-level estimates. The results are summarized in Table 30. In most of the cases, the county-level model was observed to estimate local road AADT more accurately than the statewide model.

Table 30 Comparison between statewide and county-level model validation results

County	GWR						OLS					
	Statewide			County-level			Statewide			County-level		
	MAPE	MPE	RMSE	MAPE	MPE	RMSE	MAPE	MPE	RMSE	MAPE	MPE	RMSE
Buncombe	46.2	-1.5	908	68.1	-36.2	822	48.2	- 4.4	936	72.8	-35.8	919
Columbus	74.2	-38.4	374	78.34	-25.2	368	70.1	-38.2	289	79.11	-35.6	431
Dare	73.1	-22.3	808	91.9	-76.2	641	73.1	-21.2	1154	94.6	-68.61	752
Davidson	92.1	-59.1	641	79.26	-30.9	867	81.1	-42.7	833	85.6	-34.1	892
Duplin	57.1	-19.2	478	60.1	-19.8	399	51.2	-4.2	478	52.6	-20.2	452
Iredell	91.9	-34.2	1011	92.9	-32.1	888	98.4	-48.5	1,370	95.2	-46.4	883
Mecklenburg	47.4	-1.20	1,224	60.1	-19.2	954	38.3	-16.5	1370	98.2	-46.4	1,111
Randolph	68.2	-18.8	813	92.5	-32.1	792	63.5	-12.8	772	111.9	-81.2	868
Wake	120.1	-84.1	1055	120.1	-86.2	962	88.6	-32.5	1,254	120.0	-88.3	993
Wayne	83.1	-28.2	713	108	-71.1	820	77.8	2.54	868	85.9	-55.82	852

The land use parcel descriptions are very different in many of the selected counties. Also, there are 4,744 unique land use descriptions when all counties in the state of North Carolina are considered. Hence, developing a land use-based model for the entire state needs statewide parcel data with a standardized land use variable list and descriptions for each county.

7.6 Prediction at Non-Covered Locations

The developed county-level models were used for estimating AADT at non-covered locations in each county. The sample estimations made for non-covered locations in Duplin County and Wake County are shown in figures 23 to 28.

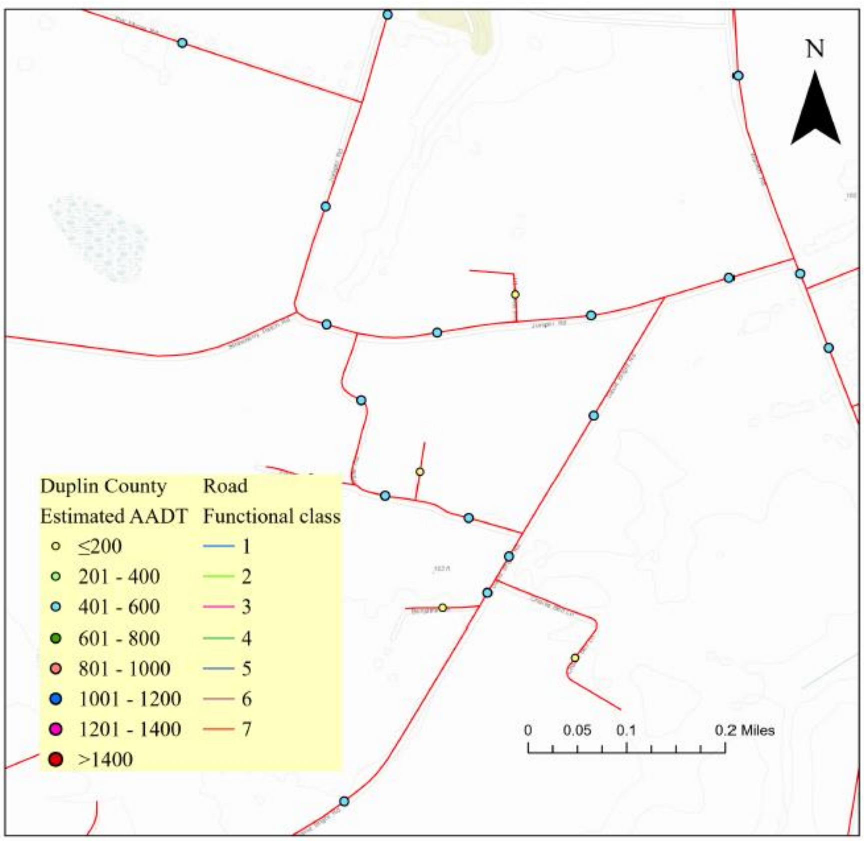


Figure 23 Estimated AADT at non-covered locations in Duplin County – low density

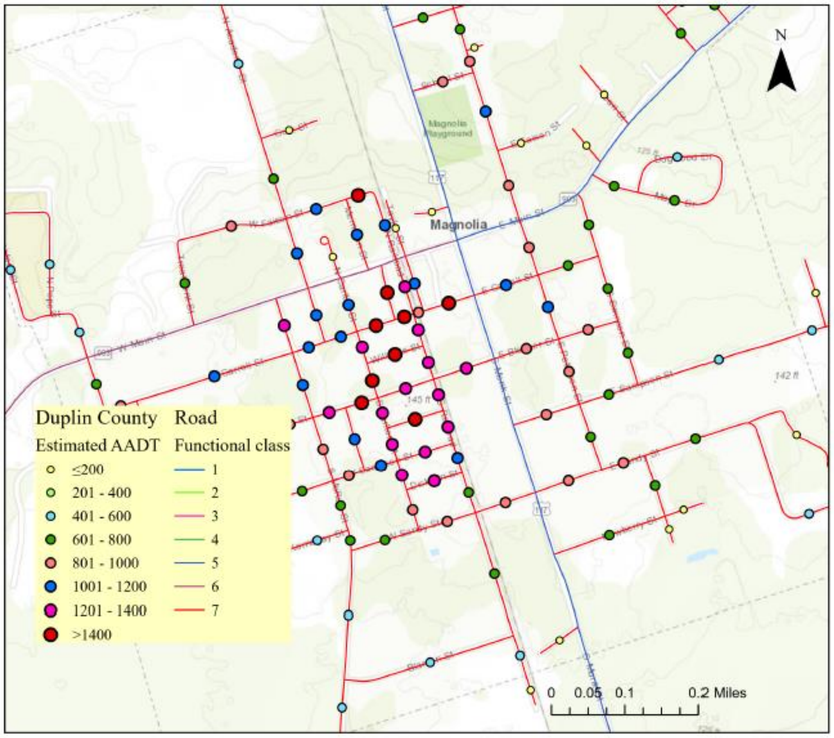


Figure 24 Estimated AADT at non-covered locations in Duplin County – low density

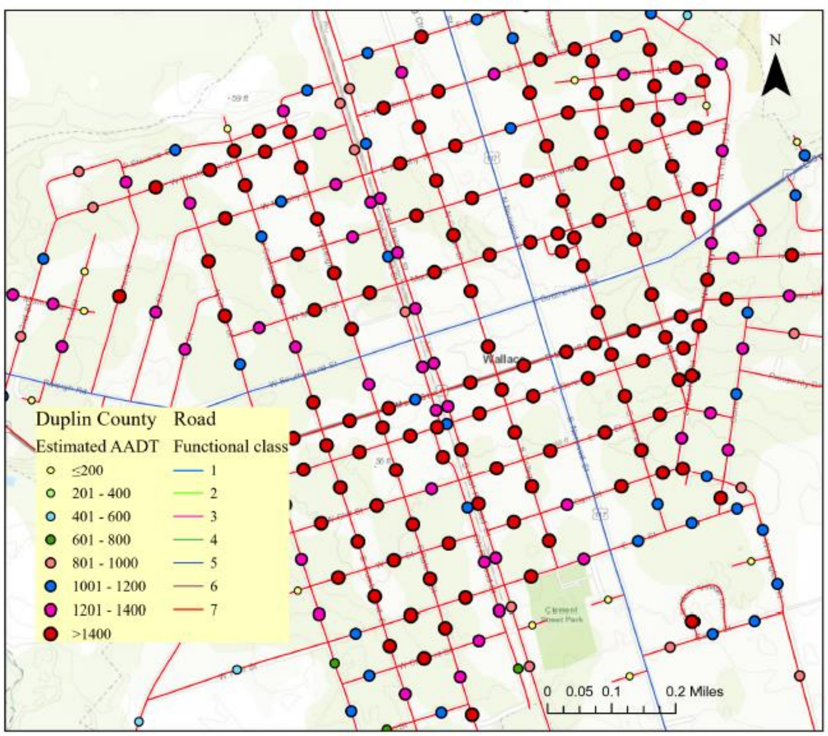


Figure 25 Estimated AADT at noncovered locations in Duplin County – high density

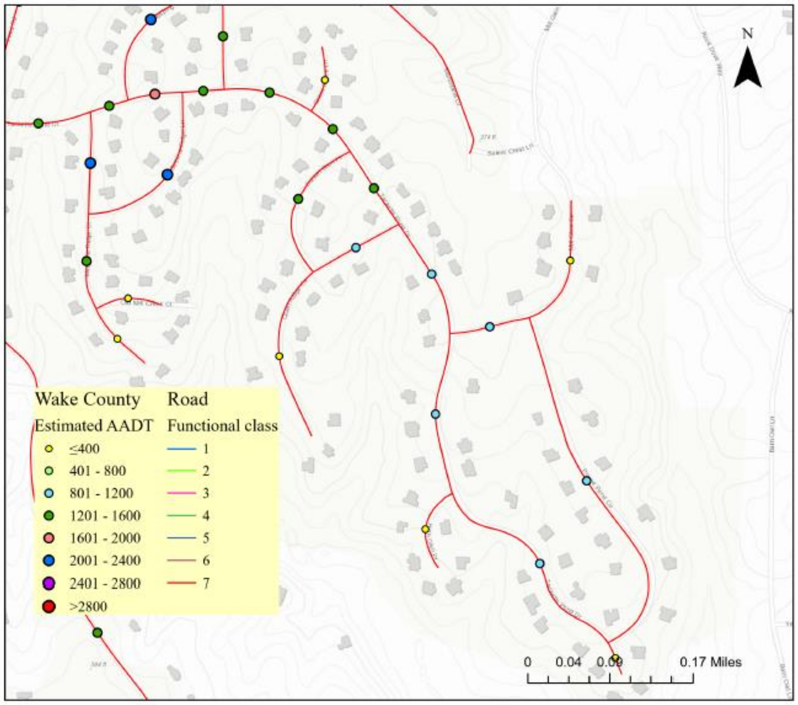


Figure 26 Estimated AADT at noncovered locations in Wake County – low density

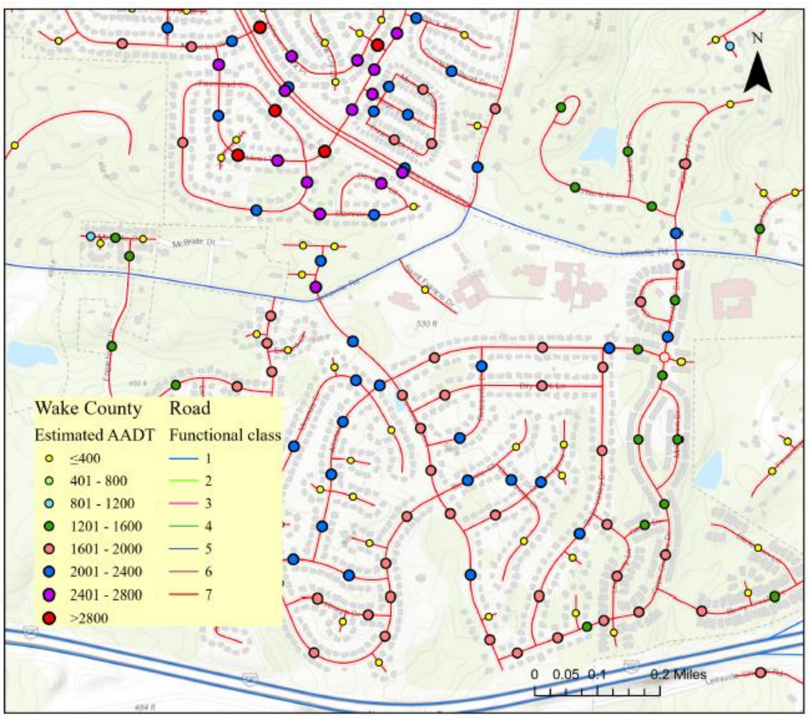


Figure 27 Estimated AADT at noncovered locations in Wake County – high density

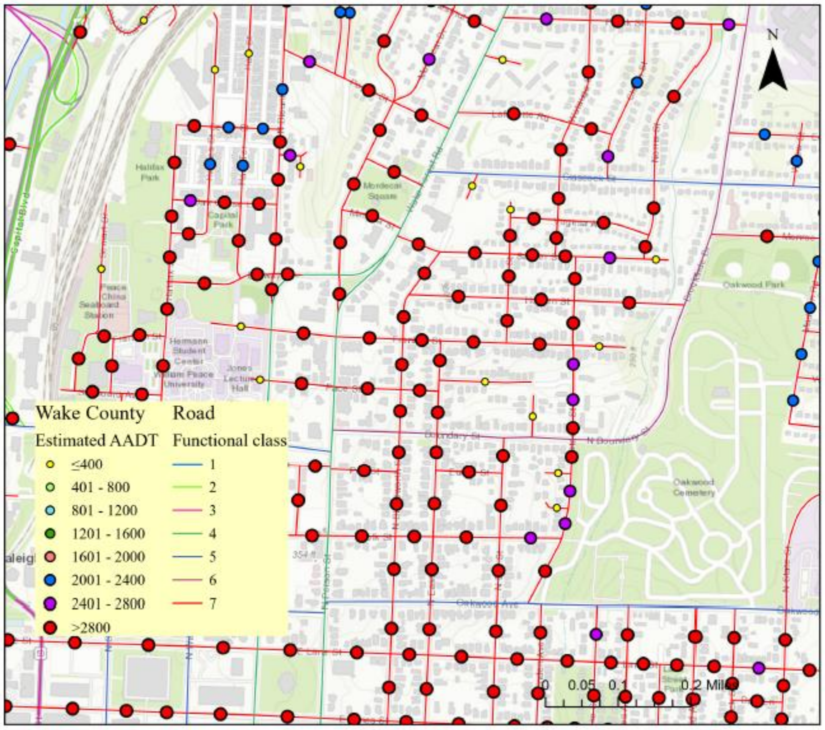


Figure 28 Estimated AADT at noncovered locations in Wake County – high density

From figures 23-28, local road AADT is higher at locations with high road density. Similarly, local road AADT is lower at locations that are far from a nonlocal road. At many locations, the predictions are found to be logical. However, predictions are overestimated at locations like dead-ends and where the local road connects to nonlocal roads. Hence, it is essential to look into the sampling requirements

CHAPTER 8: ERROR ANALYSIS AND SAMPLING REQUIREMENTS

Geospatial variations in error estimate based on road characteristics (including speed limit, accessibility, and connectivity), functional classification type (urban and rural local road), etc. need to be examined (statistical correlations) to assess where local road AADT estimates are less reliable. Solutions or what additional data need to be captured to achieve a higher acceptable level of reliability can be recommended from this analysis. Therefore, this chapter compares the median prediction error associated with the developed statewide and county-level models and investigates the sampling requirements.

8.1 Statewide Model Error Analysis

The statewide GWR method performed better than the statewide OLS method. Therefore, the error analysis was carried out using the results from the GWR model. The Pearson correlation coefficient analysis was carried out to identify the locations with a higher prediction error. The correlation between the prediction error and count-based local road AADT, speed limit, functional class type, road density, dis-nonlocal, AADT-nonlocal, population density, employment density, and the number of dead-end links was examined. The results from the Pearson correlation coefficient analysis between the median prediction error and selected explanatory variables from the statewide model is summarized in Table 31.

Table 31 Correlation analysis between prediction error and explanatory variables

Variable	Pearson correlation
Count-based local road AADT	HP
Speed limit	MN
Functional class type	MP
Road density	MP
Dis-nonlocal	LN
AADT-nonlocal	MP
Population density	MP
Employment density	LP
Dead-ends	LP

Note: LN, MN, LP, MP, and HP are low negative, moderate negative, low positive, moderate positive, and high positive correlations, respectively.

The prediction error was observed to increase with an increase in the count-based local road AADT. It is logical as there are a smaller number of counts with higher local road AADT in the database. Similarly, there is a positive correlation between the median prediction error and the functional class type. It indicates that there are unknown parameters that influence the local road AADT at locations with higher local road AADT. Therefore, it is important to investigate local roads with high AADT and identify associated factors. As seen in the disaggregate-level regression, the model performance was low for urban local roads compared to rural local roads (Table 20). The road density, which was also considered as a variable indicating development in an area, has a positive correlation with the median prediction error. Likewise, the links with higher speed limits have a low median prediction error. The frequency distribution of errors is similar to the statewide AADT distribution. As mentioned earlier, the median prediction error is considered to be the measure of central tendency. The distribution of median prediction errors in each county from the statewide model is shown in Figure 29.

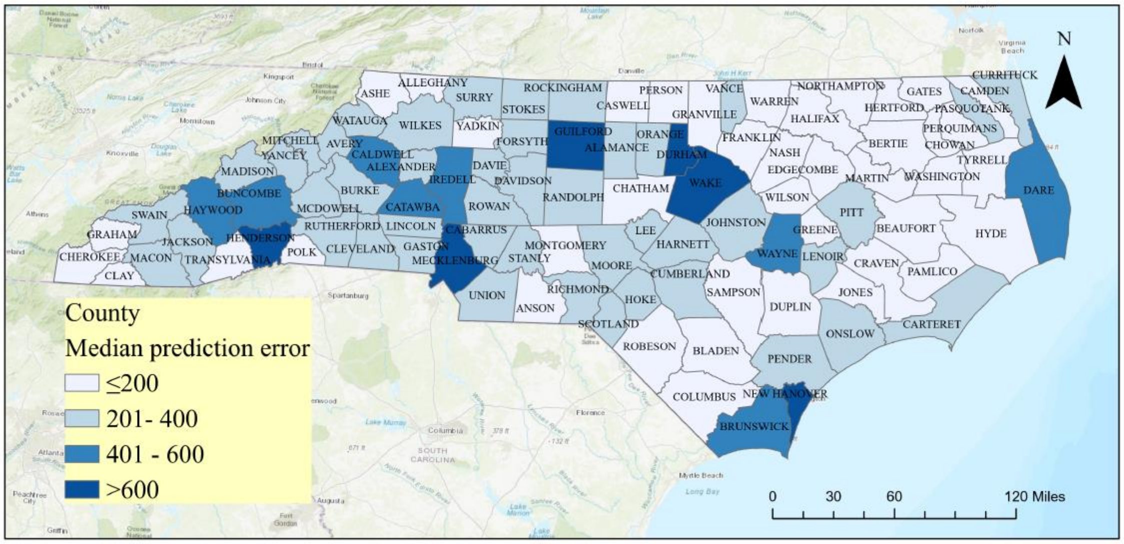


Figure 29 Median prediction error distribution by county

From Figure 29, Durham County, Guilford County, Henderson County, New Hanover County, Mecklenburg County, and Wake County have high median prediction errors when compared to other counties. Figure 30 and Figure 31 show the median prediction error by county for rural and urban local roads, respectively.

From Figure 30, most of the counties have a lower median prediction error. Urban counties like Wake County and Durham County, in addition to Brunswick County, have a comparatively higher median prediction error than other counties. The median prediction error is higher for counties in the mountains region but relatively lower for counties in the piedmont and coastal plain regions. Contrarily, from Figure 31, the median prediction error is relatively higher for counties in the piedmont and coastal plain regions. The maximum median prediction error was observed for Pender County and Stanly County. The high median prediction error could be attributed to the lower number of AADT counts for some counties.

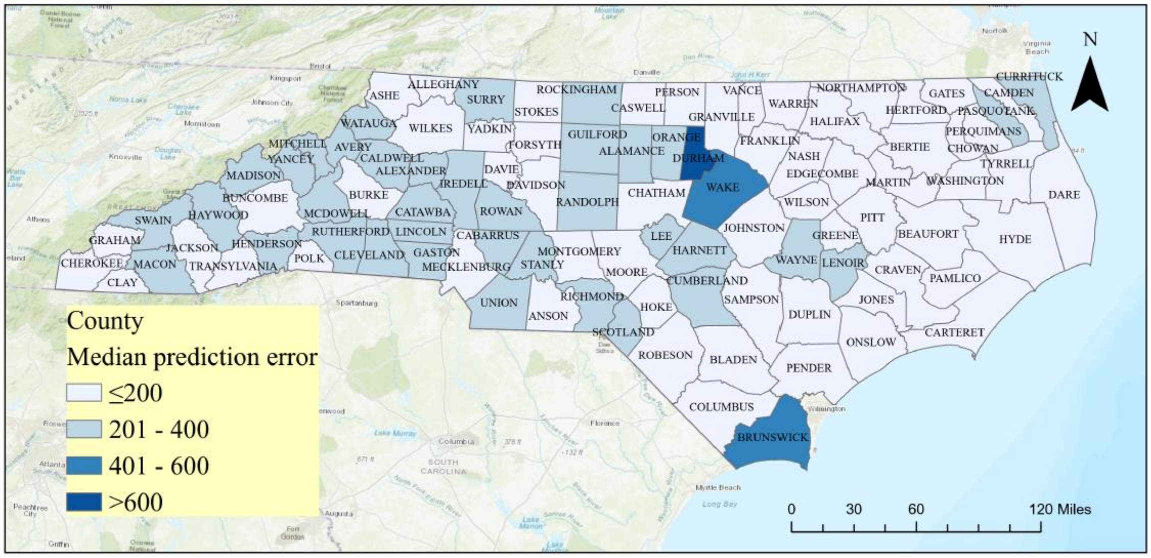


Figure 30 Median prediction error by county - rural

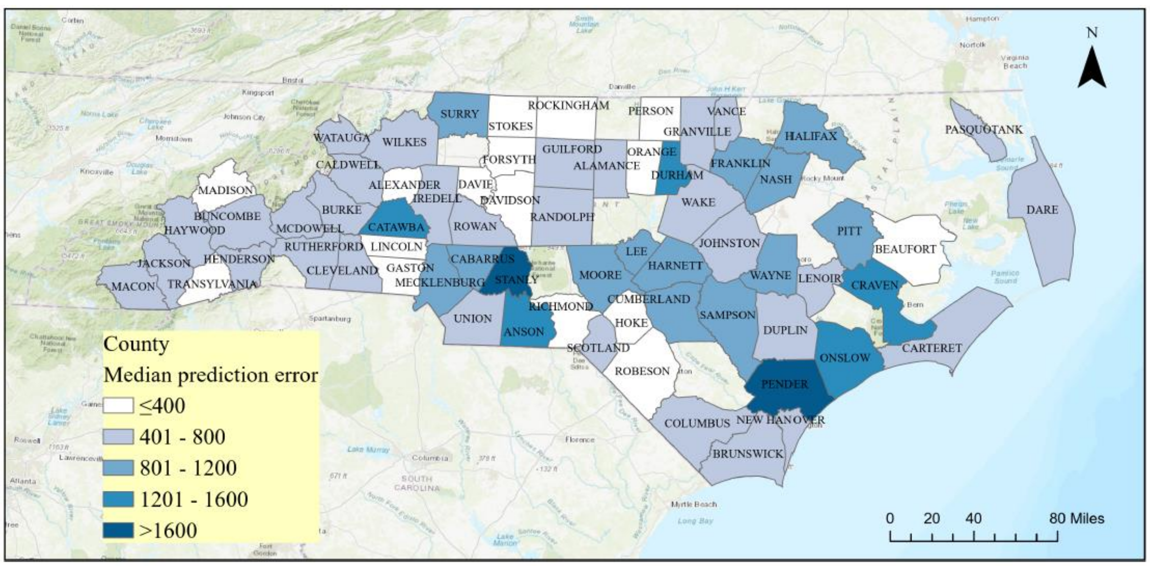


Figure 31 Median prediction error by county – urban

Figures 32 to 35 show the median prediction error by county based on the speed limit category. The median prediction error was found to be less for links with a speed limit of less than or equal to 25 mph. Most of the counties have a lower median prediction error for links with speed limits equal to 50 or 55 mph. Henderson County and Currituck County have a higher median prediction error, possibly because there are less than ten AADT counts available for local roads with a speed limit of 50 or 55 mph.

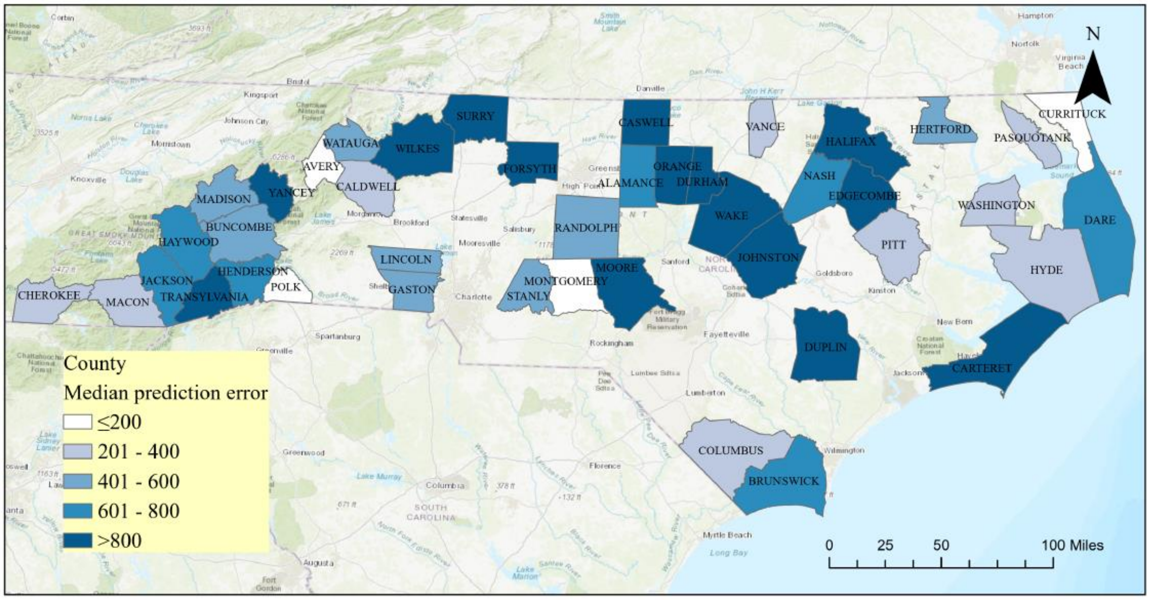


Figure 32 Median prediction error by county - speed limit ≤ 25 mph

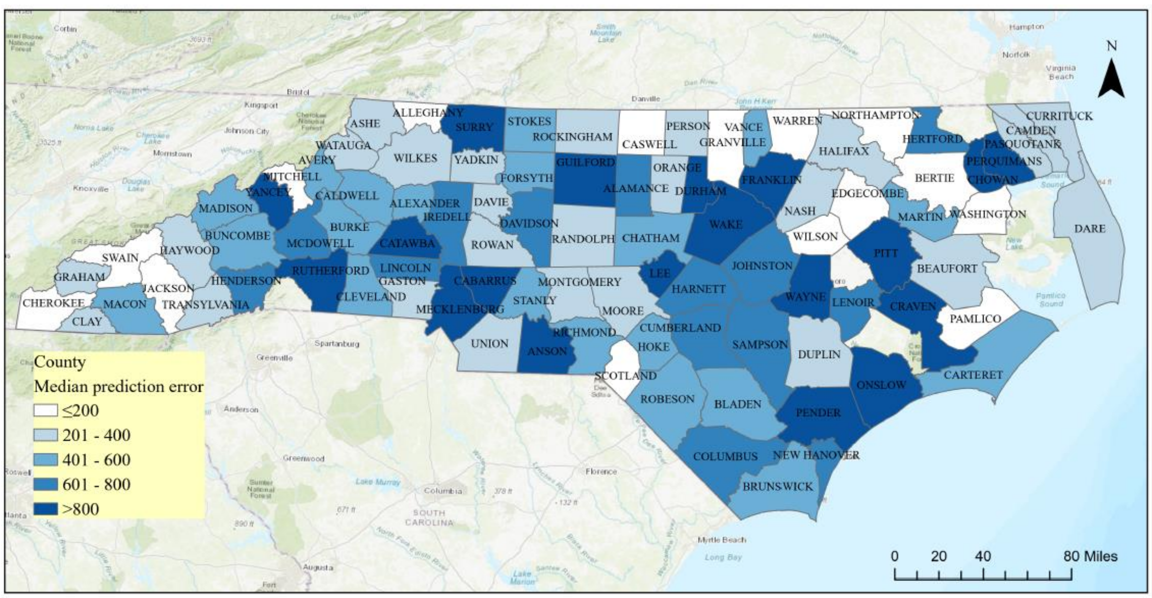


Figure 33 Median prediction error by county - speed limit = 30 or 35 mph

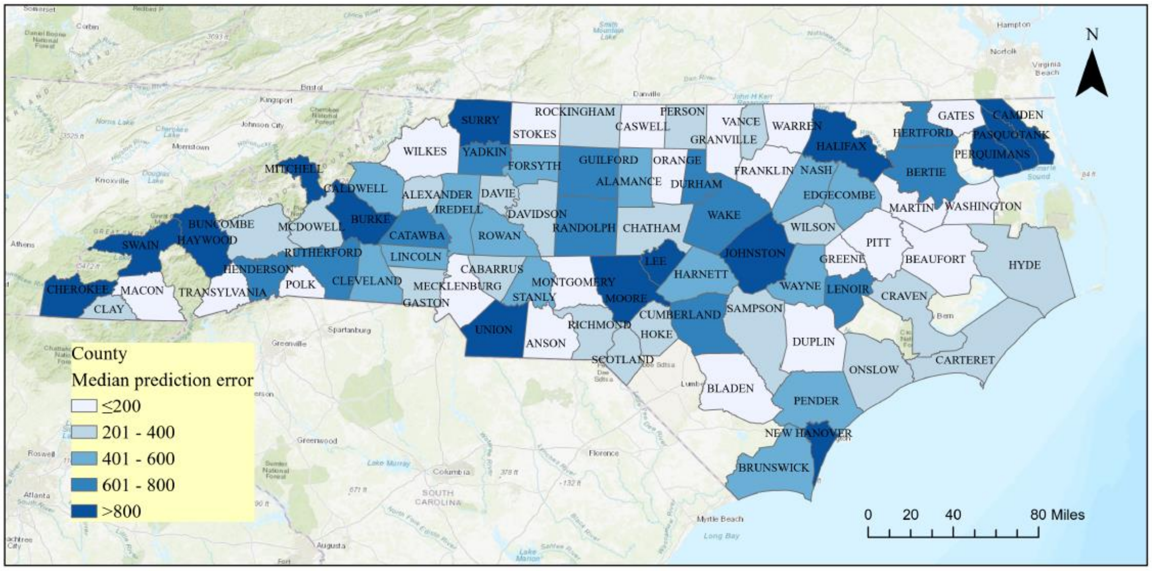


Figure 34 Median prediction error by county - speed limit = 40 or 45 mph

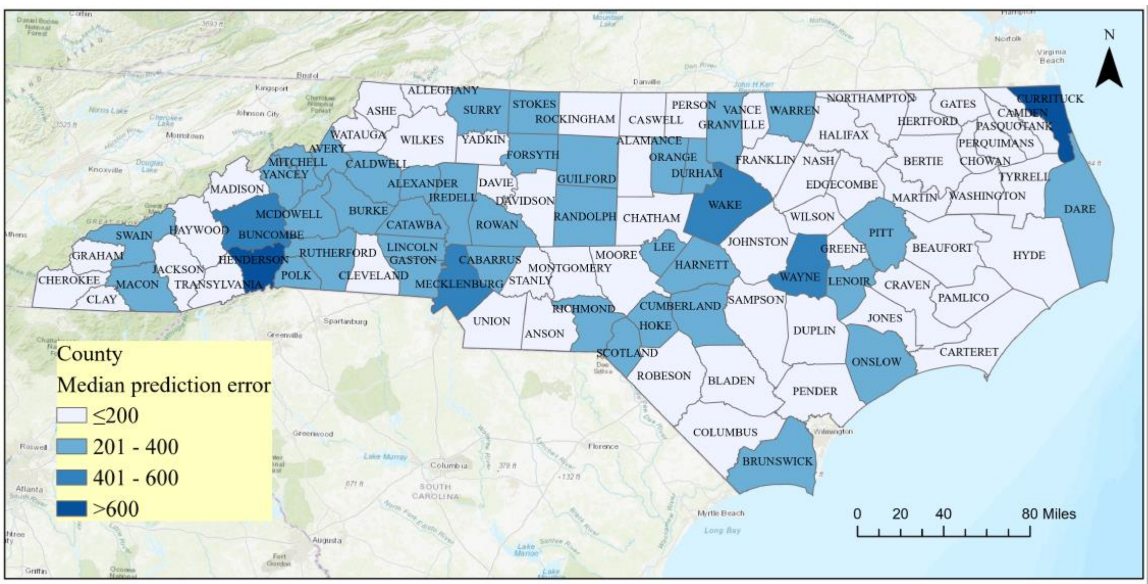


Figure 35 Median prediction error by county - speed limit = 50 or 55 mph

8.2 County-Level Model Error Analysis

The performance of county-level models is better than statewide models in most of the analytical scenarios. Also, the county-level GWR models performed better than the county-level OLS models. Hence, the prediction error analysis has been performed based on results from the county-level GWR models. The prediction error distribution for Duplin

County is shown in Figure 36. The median prediction error is found to be 217 for the county.

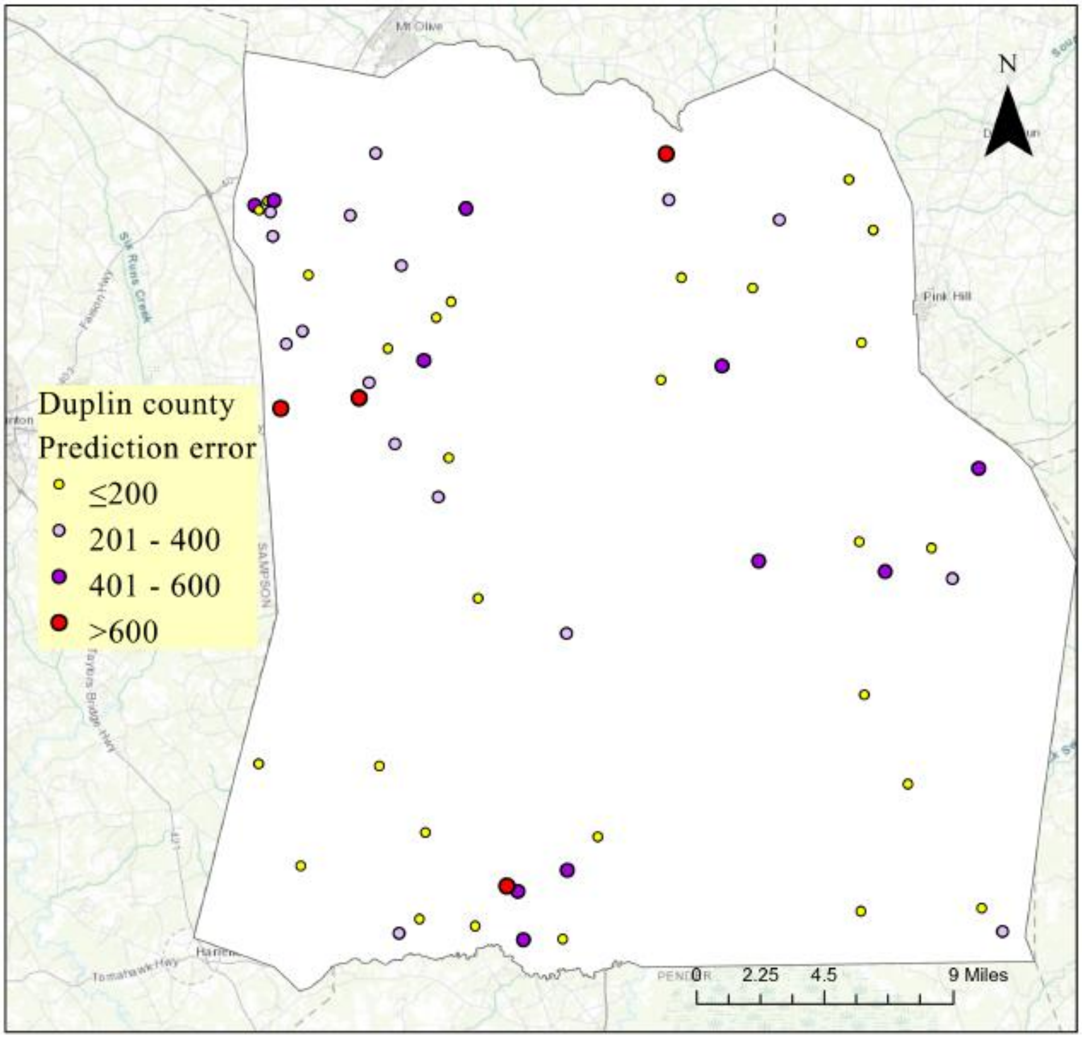


Figure 36 Prediction error distribution in Duplin county

Similarly, the prediction error distribution for Wake County is shown in Figure 37. As indicated in the modeling section, the prediction error is high for Wake county. The median prediction error is 594.

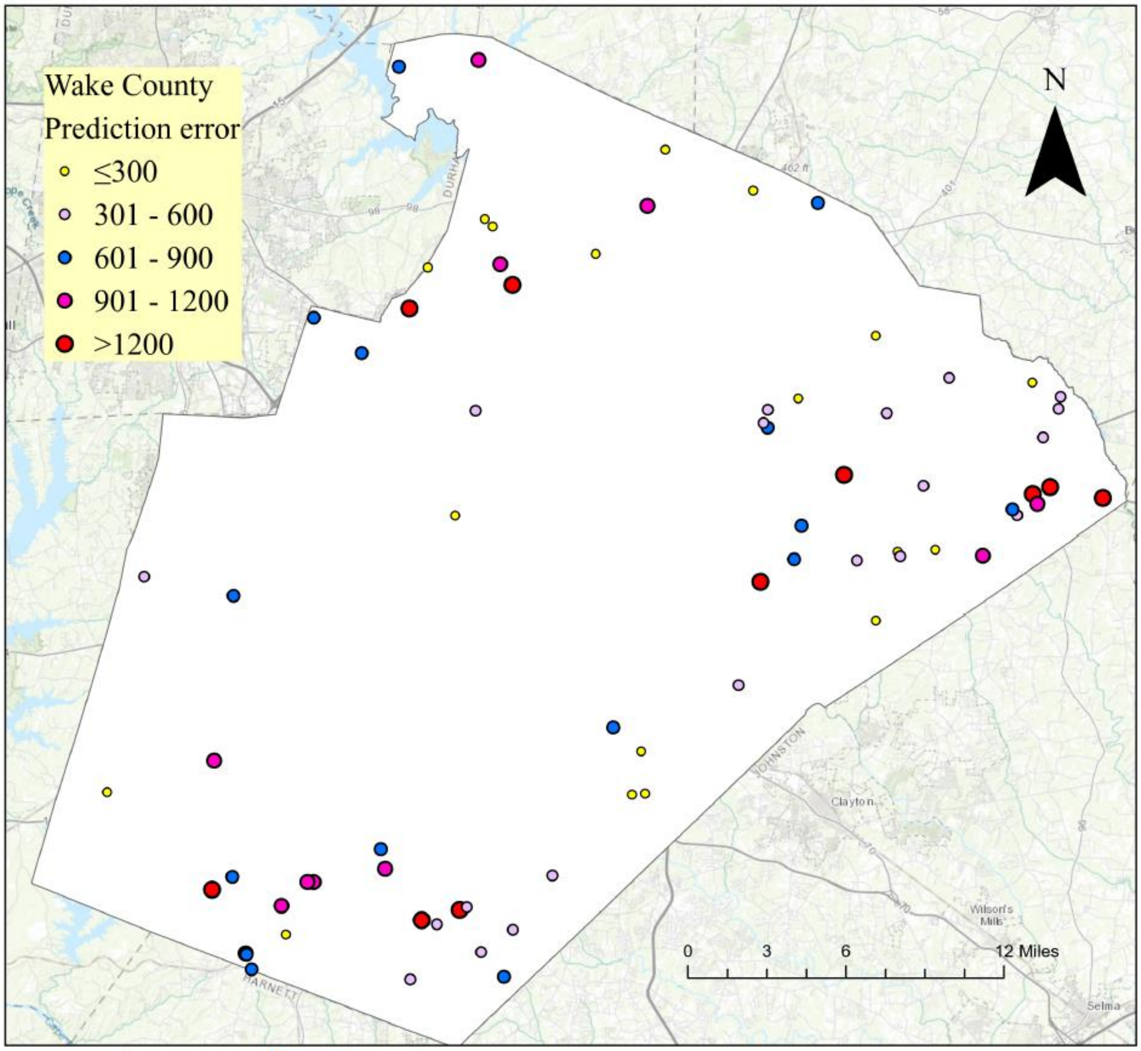


Figure 37 Prediction error distribution in Duplin county

Likewise, an assessment of prediction errors was carried out for ten selected counties in North Carolina. The assessment was conducted by functional class type and speed limit. Table 32 summarizes the median prediction error for the ten selected counties. The number of available local AADT counts is shown in parenthesis. A relatively higher prediction error was observed for Buncombe County, Mecklenburg County, Wake County, and Wayne County when all data were considered for assessment

Table 32 Median prediction error for selected counties

County	Median error						
	All data	Functional class type		Speed limit			
		Urban	Rural	<=25mph	30 mph or 35mph	40 mph or 45 mph	50 mph or 55 mph
Buncombe	494 (217)	534 (184)	204 (36)	(14)	535 (106)	493 (37)	835 (60)
Columbus	220 (203)	163 (8)	225 (195)	NA (0)	244 (31)	120 (3)	225 (168)
Dare	214 (59)	204 (24)	210 (35)	210 (16)	181 (27)	NA (0)	217 (16)
Davidson	292 (204)	869 (78)	226 (126)	NA (1)	272 (26)	812 (33)	207 (144)
Duplin	217 (235)	NA (1)	217 (234)	NA (1)	505 (26)	640 (4)	211 (204)
Iredell	298 (266)	794 (82)	242 (184)	264 (3)	774 (36)	623 (79)	228 (148)
Mecklenburg	777 (55)	786 (44)	357 (11)	NA (0)	1,060 (21)	484 (20)	223 (14)
Randolph	320 (280)	566 (53)	258 (227)	277 (2)	1,181 (47)	746 (28)	196 (204)
Wake	594 (295)	599 (264)	530 (31)	NA (8)	836 (62)	513 (144)	533 (101)
Wayne	447 (192)	1,085 (42)	322 (150)	NA (2)	750 (15)	795 (30)	375 (145)

Except for Columbus County and Dare County, the median prediction errors are higher for urban local roads in other counties. The median prediction errors for rural local roads are relatively low. It is highest for Wake County, followed by Mecklenburg County and Wayne County.

The median prediction errors are higher for local roads with a speed limit greater than 25 mph and less than 50 mph. In most of the cases, the median prediction error seems to depend on the number of available local road traffic count stations and county characteristics. Figure 38 shows the relationship between median error and the number of counts for the selected counties for modeling.

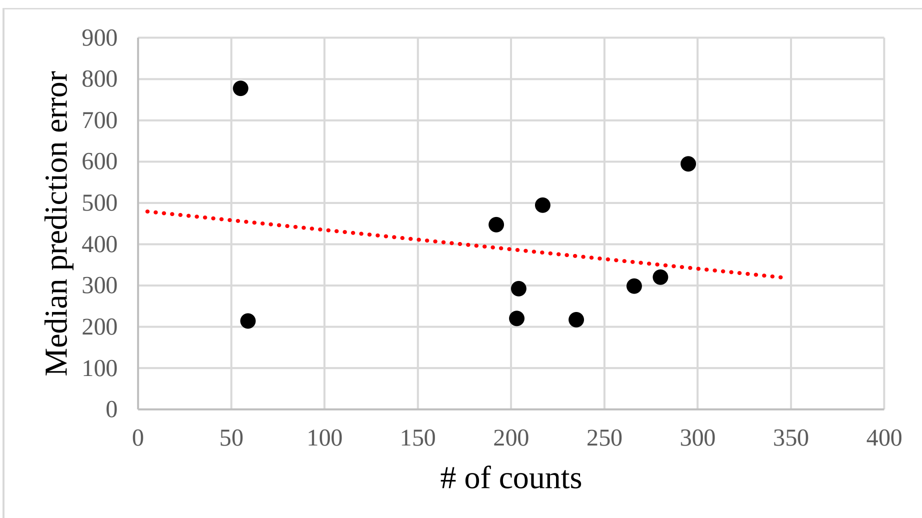


Figure 38 Relationship between number of local road traffic count stations and median prediction error

8.3 Local Road AADT Counts and Sampling Size

The results from the statewide GWR model indicate that counties with a low number of local road local road traffic count stations, the number of urban local road traffic count stations, links with a speed limit greater than or equal to 30 mph but less than 50 mph, population density more than 400 per square mile, the locations with high road density, and high employment density are locations where the median prediction error is higher. Hence, there is a need to collect more samples from such areas. A comparison of non-covered locations, available count-based local road AADT, and percent covered by selected characteristics, statewide, are summarized in Table 33.

Table 33 Comparison of non-covered locations and available local road AADT counts

Characteristic	Category	Non-covered locations	Available local road traffic count stations	% covered
Functional class type	Urban	418,449	3,035	0.72
	Rural	328,180	9,864	3.00
Speed limit (mph)	<=25	23,775	357	1.50
	30 or 35	340,599	2,279	0.67
	40 or 45	22,501	1,878	8.30
	50 or 55	359,804	8,385	2.33
Population density	<200	272,262	8,638	3.17
	200 - 400	121,861	2,251	1.78
	400 - 600	61,991	923	1.48
	600 - 800	47,278	423	0.89
	800-1000	28,848	227	0.79
	1000 - 1200	23,279	121	0.52
	1200 - 1400	25,594	136	0.53
	>1400	152,620	180	0.12
Employment density	<200	440,445	11,358	2.51
	200 - 400	87,100	834	0.96
	400 - 600	56,019	299	0.53
	600 - 800	55,889	130	0.23
	800-1000	18,063	101	0.56
	>1000	102,216	177	0.17
Local travel characteristics	Dead-end	218,043	49	0.02
	Local (F7) to local (F7)	430,510	7,186	1.67
	Local to nonlocal	89,734	4,905	5.48
	Nonlocal to nonlocal	8,394	179	2.13
Total		759,578	746,679	12,899

Note*: Local travel characteristic information is not available for some links*

From Table 33, local road traffic count stations are available for only 0.02% of dead-ends. Likewise, only 0.67% of local roads with a speed limit equal to 30 mph or 35 mph are covered. The percent of local road traffic count stations are also lower in high population density areas and high employment density areas.

The findings from the county-level models indicate that land use characteristics such as single-family residential units, multi-family residential units, and commercial areas influence local road AADT. The prediction error is relatively low for local road AADT counts locations in these land use areas. This could be attributed to the fairly good number of local road AADT count locations in the selected counties near these land use areas. Contrarily, the prediction error is high at local road AADT count locations near schools, institutions, government, office, and industrial land uses. Not enough number of local road AADT count locations are near these land use areas. This should be considered when identifying new locations for the data collection on local roads in the future.

As the county-level models have better prediction than statewide models, the sample size requirement was assessed based on non-covered locations and local road traffic count stations in each county. The non-covered locations were further divided into different categories based on functional class type and speed limit ranges. This ensured collecting a spatially distributed sample size based on key characteristics.

Typically, the population of a dataset is well defined by its sample size. This value is computed using the statistically acceptable range of “margin of error”. Equations 31 and 32 (FHWA, 2018) are used to compute the required number of local road count locations to improve the accuracy of local road AADT estimations.

$$ss = \frac{Z^2 \times C^2}{p^2} \tag{31}$$

$$N = \frac{ss}{1 + \frac{ss-1}{Pop}} \tag{32}$$

where Z = Z-statistic for a predefined confidence level, c = coefficient of variation (standard deviation divided by the mean), p is the desired prediction error rate, ss = sample size, Pop = population (total number of local road links), and N = final sample size.

The HPMS recommends using a higher confidence level and a lower prediction error rate when sampling for higher functionally classified roads. It ensures a higher level of prediction in the AADT estimates. However, the variability in traffic volumes and factors that influence the traffic volumes on local roads is significantly higher than the higher functionally classified roads. To account for such a variability in traffic volumes, a 70% confidence level and 15% prediction error rate were considered acceptable for local roads and used to estimate the sample sizes.

The total number of traffic count stations and non-covered locations are used as the population. They were identified from the road characteristics shapefile obtained from NCDOT. For example, the total number of traffic count stations and non-covered locations in Mecklenburg County is 43,045. These include 320 non-covered locations with speed limit equal to 25 mph, 38,883 non-covered locations with speed limit equal to 30 or 35 mph, 521 non-covered locations with speed limit equal to 40 or 45 mph, and 3,321 non-covered locations with speed limit equal to 50 or 55 mph in the Mecklenburg County. There are 58 local road traffic count stations currently available for modeling. If the desired prediction error rate is 0.15, coefficient of variation is 0.76 (based on all traffic count stations in the county as there are no traffic count stations on local roads with speed limit equal to 25 mph), 0.80, 0.51, and 1.19 for 25 mph, 30 or 35 mph, 40 or 45 mph, and 50 or 55 mph speed limit groups, respectively, and Z = 1.036 (at a 70% confidence level), the final sample size obtained using equations (32) and (33) is 135 (for Mecklenburg County).

Any sample size greater than 135 will increase the model prediction accuracy for Mecklenburg County at a 70% or higher confidence level.

The sample size requirement was also checked by the speed limit category, and link connectivity. They were also computed for the state of North Carolina. The results at a 70% confidence level is summarized in Table 34.

Table 34 Available local road traffic count stations and minimum recommended sample size by county based on the speed limit at a 70% confidence level

County	Speed limit (mph)								Total	
	≤ 25		30 or 35		40 or 45		50 or 55			
	Avail.	Min. reco.	Avail.	Min. reco.	Avail.	Min. reco.	Avail.	Min. reco.	Avail.	Min. reco.
Alamance	3	44	26	28	44	20	95	36	168	128
Alexander	4	12	17	20	37	28	76	46	134	106
Alleghany	0	3	7	38	2	1	69	32	78	75
Anson	0	0	16	35	9	35	134	41	159	111
Ashe	1	19	10	43	2	2	87	32	100	96
Avery	6	24	10	31	0	2	32	46	48	102
Beaufort	3	3	3	35	10	35	101	26	117	99
Bertie	2	10	11	11	5	4	85	20	103	44
Bladen	1	21	16	41	3	28	111	25	131	115
Brunswick	2	4	29	21	22	16	79	46	132	88
Buncombe	14	38	106	27	37	25	61	33	218	122
Burke	2	39	41	32	20	15	23	38	86	125
Cabarrus	0	31	8	30	25	20	26	45	59	127
Caldwell	3	11	33	22	14	21	45	27	95	82
Camden	1	38	4	15	9	13	34	66	48	133
Carteret	6	30	23	25	7	30	24	48	60	133
Caswell	5	7	10	41	18	16	58	32	91	95
Catawba	1	24	43	15	86	18	76	34	206	91
Chatham	1	36	10	29	7	28	115	30	133	123
Cherokee	7	48	37	30	5	26	36	107	85	212
Chowan	0	53	2	37	2	9	42	21	46	120
Clay	1	17	12	27	9	9	20	31	42	84
Cleveland	0	39	45	32	58	29	105	30	208	130
Columbus	1	6	32	27	3	28	169	35	205	96
Craven	1	54	18	37	28	31	64	72	111	194
Cumberland	0	48	14	41	22	16	171	58	207	163
Currituck	7	33	17	26	3	25	20	34	47	117
Dare	17	37	27	37	0	20	17	23	61	117
Davidson	1	36	26	32	34	19	151	30	212	117
Davie	1	17	7	52	10	17	112	39	130	125
Duplin	3	14	26	21	4	14	208	21	241	70
Durham	4	13	30	9	38	17	19	39	91	78
Edgecombe	1	20	14	16	6	15	97	35	118	87
Forsyth	4	26	56	43	43	26	103	76	206	171
Franklin	0	41	12	27	23	46	62	40	97	154
Gaston	4	16	45	28	55	28	64	23	168	94
Gates	1	7	2	6	7	2	73	17	83	32
Graham	2	22	13	34	3	14	12	25	30	95
Granville	0	36	8	55	15	27	68	27	91	144
Greene	0	6	4	19	3	5	101	20	108	49
Guilford	0	30	55	22	71	30	45	33	171	114
Halifax	5	19	15	51	9	22	104	83	133	175
Harnett	0	32	12	36	6	15	114	34	132	117
Haywood	19	33	49	26	11	15	16	73	95	148
Henderson	39	47	77	19	38	21	39	41	193	129
Hertford	3	37	13	54	14	14	68	37	98	142
Hoke	0	60	1	67	7	21	72	84	80	232
Hyde	3	51	4	7	4	10	28	20	39	88
Iredell	3	37	37	43	80	27	150	46	270	153

County	Speed limit (mph)								Total	
	≤ 25		30 or 35		40 or 45		50 or 55			
	Avail.	Min. reco.	Avail.	Min. reco.	Avail.	Min. reco.	Avail.	Min. reco.	Avail.	Min. reco.
Jackson	14	52	40	42	3	8	28	53	85	155
Johnston	7	35	36	24	19	15	172	50	234	124
Jones	0	11	2	1	2	8	47	32	51	51
Lee	3	5	25	49	20	30	81	44	129	127
Lenoir	1	23	17	17	8	26	121	22	147	88
Lincoln	3	23	14	32	50	23	66	47	133	125
Macon	16	62	55	42	14	47	62	55	147	206
Madison	3	19	7	19	2	0	34	33	46	71
Martin	0	21	19	48	5	3	108	31	132	103
McDowell	1	42	27	43	6	5	39	26	73	116
Mecklenburg	0	26	24	30	20	12	14	66	58	135
Mitchell	5	20	14	22	5	22	27	25	51	89
Montgomery	4	19	19	27	8	33	133	56	164	134
Moore	3	10	39	50	16	23	172	49	230	131
Nash	2	12	30	32	19	42	146	29	197	115
New Hanover	0	31	18	35	8	15	6	68	32	149
Northampton	1	24	14	23	1	18	93	34	109	100
Onslow	2	0	17	23	41	22	53	38	113	83
Orange	6	33	14	22	30	16	65	44	115	114
Pamlico	1	12	15	22	2	14	31	22	49	70
Pasquotank	9	7	7	26	5	14	39	83	60	132
Pender	0	29	15	23	14	29	109	30	138	112
Perquimans	0	47	2	18	6	13	53	37	61	115
Person	0	15	11	21	15	33	85	18	111	86
Pitt	9	16	28	36	13	20	187	69	237	139
Polk	6	30	12	33	20	13	44	37	82	113
Randolph	2	6	49	36	29	19	218	44	298	106
Richmond	2	7	26	29	6	12	139	68	173	117
Robeson	1	11	26	28	13	43	222	46	262	128
Rockingham	2	2	42	38	49	31	91	34	184	104
Rowan	0	30	29	30	32	19	136	24	197	103
Rutherford	1	37	55	24	42	32	129	57	227	150
Sampson	0	19	25	31	7	20	221	28	253	99
Scotland	0	27	13	25	5	12	91	29	109	93
Stanly	1	28	34	37	46	32	127	36	208	133
Stokes	0	22	12	32	8	21	137	44	157	119
Surry	6	39	46	25	14	20	106	24	172	107
Swain	7	43	23	20	2	2	21	53	53	118
Transylvania	14	37	25	33	11	2	15	29	65	102
Tyrrell	0	7	2	18	1	5	36	20	39	51
Union	4	9	26	56	42	28	134	49	206	142
Vance	4	1	14	60	27	49	45	29	90	139
Wake	8	15	63	20	129	21	105	38	305	94
Warren	4	23	5	13	9	14	97	28	115	78
Washington	5	15	9	9	7	4	38	20	59	49
Watauga	5	27	20	36	3	1	42	19	70	83
Wayne	2	47	15	29	30	21	148	34	195	131
Wilkes	6	32	27	33	19	24	149	39	201	128
Wilson	2	9	17	44	21	21	123	62	163	136
Yadkin	0	14	10	16	5	15	87	40	102	86
Yancey	3	17	12	67	1	11	32	75	48	170
North Carolina	357	2,477	2,279	3,051	1,878	1,938	8,385	4,026	12,899	11,492

Table 35 Available local road traffic count stations and minimum recommended sample size by county based on link connectivity at a 70% confidence level

County	Link connectivity (beginning and ending features)																			
	F7 - F7		F7 - F6/F5		F7 - F4/F3		F7 - F1/F2		F6/F5 – F6/F5		F6/F5 – F1/F2/F3/F4		F1/F2/F3/F4 – F1/F2/F3/F4		Unknown		Dead-end		Total	
	Avail.	Min. reco.	Avail.	Min. reco.	Avail.	Min. reco.	Avail.	Min. reco.	Avail.	Min. reco.	Avail.	Min. reco.	Avail.	Min. reco.	Avail.	Min. reco.	Avail.	Min. reco.	Avail.	min. reco.
Alamance	93	46	38	33	35	37		5	1	16		22		18	1	39	0	42	168	258
Alexander	70	54	40	41	21	22		0		9		0	1	4	2	0	1	49	134	180
Alleghany	48	67	23	52	6	32		0		2		1		4	1	55	1	70	78	283
Anson	95	53	42	100	12	51		0	2	4	1	4		27	7	29	0	81	159	348
Ashe	47	85	23	54	13	17		0		2		4		2	17	19	1	68	100	250
Avery	27	47	11	13	10	22		0		6		0		5	0	32	0	51	48	177
Beaufort	61	34	30	21	17	10		5	1	6		6		10	8	6	0	32	117	131
Bertie	52	17	33	27	4	4		0	2	3		3		9	12	9	0	19	103	91
Bladen	51	51	56	83	13	19		0	2	1		6		15	9	14	1	66	131	254
Brunswick	94	32	13	19	15	41	1	13	2	7		0		19	7	33	0	33	132	198
Buncombe	152	32	39	26	17	21	7	11		12	1	12		24	2	7	0	31	218	175
Burke	43	25	16	29	23	24	2	21		7	2	10		14	0	27	0	31	86	189
Cabarrus	37	30	10	32	9	23		10		13	1	13		25	2	23	0	34	59	202
Caldwell	57	29	16	14	11	20		0	1	11		11		15	10	18	0	30	95	147
Camden	25	66	13	40	6	12		0	2	4		0		6	2	3	0	47	48	179
Carteret	28	40	9	16	17	51		0		6		5		24	6	13	0	42	60	196
Caswell	36	104	34	21	13	24	1	7	1	7		0	1	3	5	15	0	84	91	265
Catawba	101	28	66	21	32	17	3	2	2	7		11		20	2	31	0	24	206	161
Chatham	82	55	38	28	6	15	1	5		15	1	9		30	5	24	0	42	133	223
Cherokee	57	57	22	37	5	27		0		6	1	7		36	0	33	1	54	85	256
Chowan	27	20	11	12	7	50		11		1	1	4		6	0	8	0	60	46	172
Clay	24	25	6	26	11	16		0		2		1		2	1	18	0	24	42	112
Cleveland	123	45	50	29	31	34		11		12		10	2	16	2	13	0	46	208	215
Columbus	83	46	72	35	27	20	3	10	3	3	1	19		18	16	25	0	42	205	216
Craven	74	54	23	90	11	73	1	14		10		4		41	2	11	0	62	111	360
Cumberland	125	44	30	40	36	41	2	6	1	10	1	16		43	12	21	1	52	207	272
Currituck	28	44	12	16	6	20		0		4		1		3	1	16	1	37	47	140
Dare	36	41	13	21	11	43		0		3	1	15		2	0	16	2	48	61	189
Davidson	120	35	54	27	32	58	1	28	1	25		14	2	15	2	2	1	41	212	245
Davie	80	42	14	28	32	25	2	1		3		3		9	2	2	0	40	130	153
Duplin	110	40	87	36	5	5	1	14	3	14		0		13	35	30	0	36	241	188
Durham	50	27	14	21	21	21	2	1		14		9		20	4	12	0	23	91	147
Edgecombe	46	25	48	51	13	28		11	5	13	2	0		15	4	11	0	38	118	191
Forsyth	120	52	54	41	22	77	4	29		27	1	26		41	5	46	2	54	206	393
Franklin	44	64	17	34	32	31		0		4		9		20	4	0	0	52	97	213
Gaston	104	35	19	23	39	27	1	18		7		17	1	25	4	7	2	32	168	191
Gates	46	15	21	15	11	10		0	2	3	1	0		2	2	9	1	15	83	69
Graham	18	55	2	11	8	7		0	1	2	1	5		13	0	11	1	52	30	156

County	Link connectivity (beginning and ending features)																			
	F7 - F7		F7 - F6/F5		F7 - F4/F3		F7 - F1/F2		F6/F5 – F6/F5		F6/F5 – F1/F2/F3/F4		F1/F2/F3/F4 – F1/F2/F3/F4		Unknown		Dead-end		Total	
	Avail.	Min. reco.	Avail.	Min. reco.	Avail.	Min. reco.	Av ail.	Min. reco.	Avail.	Min. reco.	Avail.	Min. reco.	Avail.	Min. reco.	Avail.	Min. reco.	Avail.	Min. reco.	Avail.	min. reco.
Granville	50	50	15	11	20	42		6		5	1	9	1	10	4	12	2	53	91	196
Greene	44	17	48	26	7	13	1	4	2	8		0		6	6	5	0	22	108	101
Guilford	90	29	46	27	28	27	2	22	1	18	2	15	2	8	0	29	0	31	171	207
Halifax	48	66	42	88	22	40	1	14		5		7		15	20	38	1	71	133	346
Harnett	74	37	26	10	23	37		13		10	2	9	1	22	6	22	1	38	132	199
Haywood	68	27	12	62	12	32	1	22		9		5		24	2	11	0	32	95	225
Henderson	118	31	30	31	35	24	1	4		13	1	13		18	8	31	0	29	193	192
Hertford	59	76	29	29	2	16		0	3	8	3	3		1	2	16	0	56	98	204
Hoke	38	39	24	44	13	55		0	1	10	1	6		17	3	10	0	73	80	253
Hyde	6	22	12	34	8	14		0	5	5	2	4		0	6	10	0	32	39	121
Iredell	171	55	52	47	38	39	3	17		16		3	1	28	5	28	0	52	270	285
Jackson	63	58	12	43	9	35		0		8		1		36	1	41	0	57	85	279
Johnston	112	42	61	47	39	41	2	0	2	0	4	11		31	14	26	2	46	234	245
Jones	25	36	16	23	7	6		1		1		1	1	7	2	8	0	28	51	111
Lee	69	56	37	52	16	33	3	10		6		8	2	18	2	24	3	53	129	260
Lenoir	82	32	41	28	14	12	1	16	4	11		0		18	5	7	0	31	147	154
Lincoln	76	38	32	24	23	30		8	1	7		4		13	1	31	0	39	133	194
Macon	110	62	19	22	14	39		0		9		2		25	4	17	1	65	147	241
Madison	34	30	7	48	4	49		20		13		9		21	1	27	0	33	46	250
Martin	73	44	32	20	14	25		0	1	6		4		21	12	11	1	40	132	169
McDowell	46	53	23	55	2	2		15		9		8		11	2	43	0	56	73	253
Mecklenburg	45	22	2	66	8	33	2	34		20	1	22		26	0	24	2	28	58	275
Mitchell	33	30	12	20	4	7		0		4		0		3	2	13	1	31	51	109
Montgomery	96	60	42	69	7	21	1	18		6		2		1	18	39	1	60	164	277
Moore	127	55	34	58	51	42		5		10		10		34	18	51	1	53	230	319
Nash	90	46	85	20	7	52	2	28	5	17		15		32	8	19	1	46	197	274
New Hanover	20	35		31	12	33		7		14		16		28	0	22	0	33	32	218
Northampton	49	41	34	22	15	19		5	2	6	1	4		8	8	6	1	31	109	141
Onslow	79	28	12	44	16	32		8		9		12		24	6	24	0	32	113	212
Orange	72	27	27	54	12	28	3	33		5		7		22	1	32	0	35	115	243
Pamlico	34	22	13	26		0		0	1	4		0		0	1	13	0	23	49	89
Pasquotank	31	47	19	50	7	41		8	1	5	1	4		19	1	22	0	60	60	255
Pender	65	50	26	37	21	35		10	1	3	0	1		18	25	20	0	45	138	217
Perquimans	30	47	24	74		18		0		1		1		10	7	14	0	63	61	229
Person	53	39	33	23	19	26		0		6		10		11	6	13	2	34	111	162
Pitt	117	60	54	50	43	55	1	29	1	18	3	14		39	18	38	1	60	237	362

County	Link connectivity (beginning and ending features)																			
	F7 - F7		F7 - F6/F5		F7 - F4/F3		F7 - F1/F2		F6/F5 – F6/F5		F6/F5 – F1/F2/F3/F4		F1/F2/F3/F4 – F1/F2/F3/F4		Unknown		Dead-end		Total	
	Avail.	Min. reco.	Avail.	Min. reco.	Avail.	Min. reco.	Avail.	Min. reco.	Avail.	Min. reco.	Avail.	Min. reco.	Avail.	Min. reco.	Avail.	Min. reco.	Avail.	Min. reco.	Avail.	min. reco.
Polk	51	42	23	23		0	1	16		6		3		11	7	14	1	37	82	153
Randolph	157	36	70	45	45	44	5	27	4	5		13	3	16	14	16	1	43	298	245
Richmond	103	60	21	36	31	54		15		5		9	1	32	17	33	2	60	173	304
Robeson	145	46	83	63	19	57		30		30	3	16		18	12	15	0	52	262	327
Rockingham	111	41	46	42	22	44	1	8	1	11	1	11		25	2	11	0	43	184	235
Rowan	124	37	40	40	27	35	1	17		16		5		24	5	45	0	37	197	255
Rutherford	146	49	42	45	31	42	1	12	1	9		1		25	6	13	0	51	227	246
Sampson	116	42	64	28	39	26	3	5	1	12	6	9		16	24	11	0	40	253	188
Scotland	58	47	23	29	14	16	1	16	3	9	2	11	1	23	7	17	0	46	109	214
Stanly	119	49	71	54	9	39		0	2	8		8		5	7	15	2	52	208	229
Stokes	84	52	46	55	20	32		5	1	9		1		6	6	8	0	53	157	220
Surry	97	47	33	64	17	21	4	5		14	3	4		25	18	29	0	49	172	260
Swain	43	42	8	26	1	15	1	10		4		10		12	0	21	0	36	53	175
Transylvania	36	47	9	10	18	37		0		5		6		12	2	6	0	42	65	164
Tyrrell	20	45	12	18	2	7		5		4		3		3	5	12	0	30	39	127
Union	95	63	58	44	33	37		0	2	9	6	15	1	35	11	27	0	57	206	286
Vance	37	48	27	51	19	66		6	3	4	2	7		15	2	36	1	69	90	301
Wake	173	26	49	32	71	25	1	23	1	19	1	19	3	21	6	22	1	27	305	213
Warren	71	29	28	17	3	13	1	6		3	0	1		6	12	32	0	25	115	132
Washington	36	26	20	18		1	2	18		7		2		10	1	12	0	22	59	117
Watauga	48	34	4	50	11	24		0		0		2		16	7	25	1	33	70	185
Wayne	99	34	56	31	22	31		17	3	3	1	10	1	28	13	8	0	38	195	200
Wilkes	132	66	34	29	20	31		10	1	12		9		21	14	26	2	61	201	266
Wilson	90	67	51	32	16	69		19	1	12	1	15		38	4	31	1	64	163	348
Yadkin	55	51	25	16	13	42		12		10	2	3		11	7	13	0	47	102	205
Yancey	29	85	8	30	9	36		0	1	6		0		11	1	30	0	82	48	280
North Carolina	7,186	4,382	3,103	3,604	1,724	2,970	78	868	88	836	66	705	25	1,703	629	2,012	49	4,448	12,899	21,527

Note1: F1: Interstate; F2: Principal arterial – other freeways and expressways; F3: Principal arterial; F4: Minor arterial; F5: Major collector; F6: Minor collector; F7: local road

Note 2: Shaded cells indicate that the minimum number of recommended local road traffic count stations are more than the available number of local road traffic count stations

8.4 Growth Factor Analysis and AADT Estimation at Non-covered Locations

The growth factor analysis is critical in the case of local roads, as most of the local road AADT is not available. There are currently 759,578 non-covered locations in the state of North Carolina while counts are collected at 12,899 local road locations. Even at stations where local road count-based AADT is available, they are not collected annually. The AADT at all these locations can be either based on the counts collected during the reporting year or the growth factor estimates from the previous year or the GWR model developed for prediction.

Currently, the local roads are counted in alternating years. In general, data are collected at 50% of available local road traffic count stations in odd years while data are collected at the other 50% of available local road traffic count stations in even years. Hence, a growth factor is computed using count-based AADT for the reporting year and count-based AADT estimated two years ago, for each local road with available count-based AADT. It was then divided by two to represent the annual growth factor for the reporting year, for the location. The analysis was carried out using the data from the year 2006 to 2018. For example, the year 2015 growth factor for a location is based on 2015 count-based AADT and 2013 count-based AADT. A one-year growth factor is generated. The 2015 growth factors are suitable for estimating 2015 AADT estimates at locations where count-based AADT estimate is available for the year 2014. The median and mean growth factors are nearly the same in all the analysis years. The past 5-year and all year average growth factors are estimated as 1.02 and 1.01 for North Carolina. On average, the count-based local road AADT does not seem to change significantly from year to year while considering the average growth factors for the state.

The improved performance of county-level AADT estimation models in the validation section substantiates that county-level growth factors are appropriate for the local roads. As an example, the mean growth factors for the year 2015 for all the counties are spatially depicted in Figure 39. The mean growth factor for the year 2015 from the statewide data was 1.03 while the county-level growth factor estimates varied from 0.93 for Tyrell County to 1.13 for Perquimans County. The mean growth factor estimates for each county, by year, are summarized in Table 35.

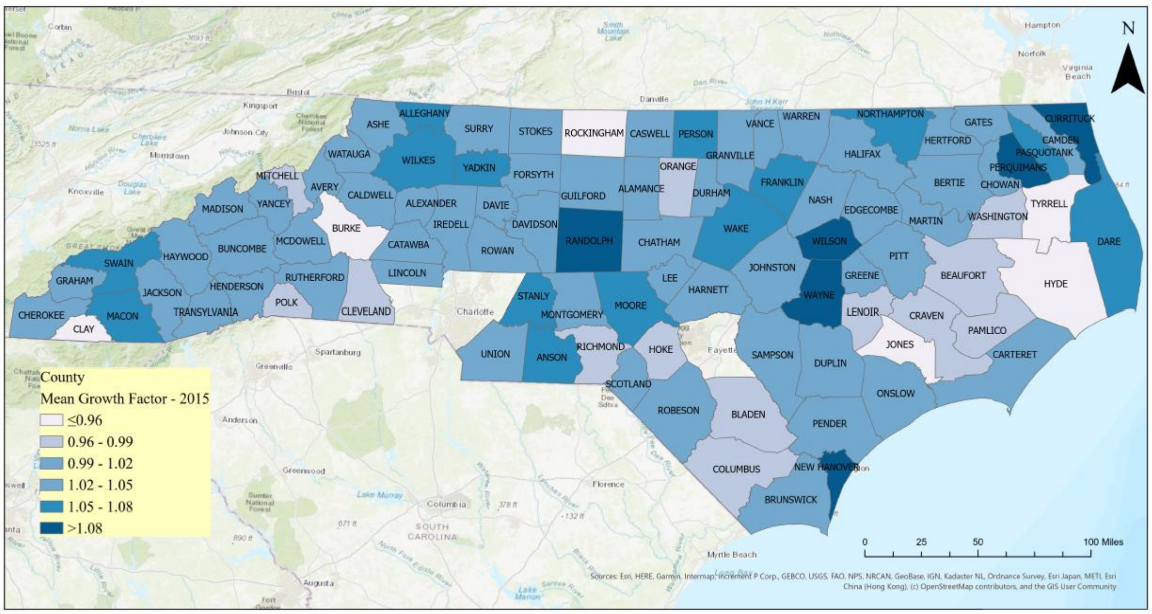


Figure 39 County-level mean growth factor estimates for the year 2015

Table 36 Mean growth factor estimates for each county

County	Mean growth factor										Average	
	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	5-Year	All
Alamance	1.05	0.99	1.03	1	0.95	1	1.03	1	0.97	1.01	1.00	1.00
Alexander	0.99	0.98	1.03	1	0.94	1.01	1.04	1.03	1	1	1.02	1.00
Alleghany	1.06	0.95	1.01	1	0.97	1.01	1.02	1.06	1	1.07	1.03	1.02
Anson	1.02	1.13	1.09	1.02	0.97	0.95	1.01	1.06	1.03	1.05	1.02	1.03
Ashe	0.99	0.96	1.01	1.02	0.99	0.98	0.93	1.03	1.1		1.01	1.00
Avery	1.1	0.98	0.98	1	0.98	1.01	1.08	1.02	1.02	1.1	1.05	1.03
Beaufort	0.96	1.03	1.06	1	0.99	1.01	1	0.98	0.94	0.95	0.98	0.99
Bertie	0.97	0.99	0.98	0.99	0.98	0.98	1	1	1.05	1.02	1.01	1.00
Bladen	0.94	1.04	1.04	1.04	0.99	0.97	0.99	0.99	1	1.03	1.00	1.00
Brunswick	0.98	0.96	1	0.99	1.02	1.01	1	1.03	1.04	1.08	1.03	1.01
Buncombe	1.01	1	0.98	0.98	1.01	0.97	0.99	1.05	1.04		1.01	1.00
Burke	1.01	0.98	0.98	0.99	0.94	0.99	1.01	0.96	1.04	1.02	1.00	0.99
Cabarrus	1.02		1		0.97		1.02				1.02	1.00
Caldwell	1.02	0.99	1.02	1.03	0.94	0.98	0.97	1.03	1.02	0.99	1.00	1.00
Camden	1	0.92	1.04	1.1	1	0.99	0.94	1.03	1.1	0.97	1.01	1.01
Carteret	0.98	1	1.03	1	0.99	0.98	1.03	1.05	1.04	1.01	1.02	1.01
Caswell	0.99	0.93	1	1.07	0.98	0.97	0.99	1.02	1.07	1	1.01	1.00
Catawba		1.01		0.99		1		1.03		1.02	1.02	1.01
Chatham	0.98	0.96	1.01	1.02	1.05	1.02	0.99	1.02	1	0.93	0.99	1.00
Cherokee	1	1	1.03	0.99	0.92	0.97	1.02	1.03	1.06	1.03	1.02	1.01
Chowan	0.99	1.01	1.01	0.96	0.97	1.02	1.03	1.03	1.01	0.89	1.00	0.99
Clay	1.04	1.07	0.94	0.91	1.01	1.04	0.98	0.96	1.07	1	1.01	1.00
Cleveland	1	0.96	0.99	0.98	1.03	1.08	1.02	0.98	1.01	1	1.02	1.01
Columbus	0.97	0.99	1.03	0.99	0.99	1	0.99	0.99	1.03	1.05	1.01	1.00
Craven	1.02	1	1.04	0.99	1.01	1	0.97	0.97	1.03	1	0.99	1.00
Cumberland	0.98		1.03		0.98		1.01		1.05		1.03	1.01
Currituck	1.02	0.98	0.97	1.01	1.03	0.93	1.02	1.12	0.99	0.95	1.00	1.00
Dare	0.96	0.9	0.95	1.03	0.99	1.04	1.08	1.06	1	1.02	1.04	1.00
Davidson	1.03	0.97	0.95	0.99	1	0.97	0.99	1.04	1.02	1.03	1.01	1.00
Davie	0.98	0.97	0.98	0.98	0.99	0.96	1.03	1	0.95	1.08	1.00	0.99
Duplin	1.11	0.98	0.96	1.01	1	0.97	1.01	1.04	1	0.99	1.00	1.01
Durham		1		1		1.17		1.04		0.99	1.07	1.04
Edgecombe	1	0.99	0.98	0.98	1	0.99	0.99	1.03	1.1	1.01	1.02	1.01
Forsyth		0.99		1		1		1.02		1.05	1.02	1.01
Franklin	0.95	0.99	1.03	1.01	1.06	0.96	0.96	1.06	1.04	1.03	1.01	1.01
Gaston	1.01		0.97		0.99		1.03		1		1.02	1.00
Gates	0.99	1.02	0.93	0.96	1.02	1.01	1	1.03	1.02	1.08	1.03	1.01
Graham	1.03	0.95	0.98	0.99	0.92	0.91	1.04	1.05	1.01	0.95	0.99	0.98
Granville	0.99	0.99	0.98	1.01	1.07	0.98	0.96	1.04	1.02	1	1.00	1.00
Greene	0.96	1.02	1.04	1.03	0.97	0.95	1.01	1.05	1	1.01	1.00	1.00
Guilford		1.01		0.99		1		1.03		1.05	1.03	1.02
Halifax	0.94	1.01	1	0.97	1.03	0.95	1.02	1.04	1.02	1.04	1.01	1.00
Harnett	1.04	0.98	0.98	0.95	1.04	1.04	0.93	1.01	1.04	0.94	0.99	1.00
Haywood	0.99	1.07	1	0.92	1.02	1	0.97	1.02	1.04	1	1.01	1.00
Henderson	1.02	1	0.97	0.98	1.01	0.99	1.06	1.04	0.99	1.05	1.03	1.01
Hertford	1.03	1	0.96	1.03	1.03	0.99	0.95	1.03	1.05	0.96	1.00	1.00
Hoke	0.99	1.05	0.99	1.03	1.05	1.05	0.93	0.99	1.08	1.06	1.02	1.02
Hyde	1.04	0.93	1	0.98	0.9	1.06	0.97	0.95	1.12	1.02	1.02	1.00
Iredell	1.04	0.98	0.98	1	1.01	1.02	1	1	1	1	1.00	1.00
Jackson	1.07	1.02	1.04	1.15	0.96	1.01	1.15	1.05	0.96	1.03	1.04	1.04
Johnston	1	0.95	1.03	1.01	1.05	1.06	1	1.04	1.02	0.98	1.02	1.01
Jones	0.97	0.99	1.01	0.97	1.01	1.01	0.88	0.96	1.09	1.04	1.00	0.99
Lee	1.03	0.98	0.97	0.96	1.06	1.04	0.97	1.02	1.04	0.99	1.01	1.01
Lenoir	1.03	0.99	0.95	1	1.03	0.98	0.98	0.98	1.05	1.06	1.01	1.01
Lincoln	1.01	0.99	0.95	0.94	1.05	1	1.03	1.03	0.98	1.09	1.03	1.01

County	Mean growth factor										Average	
	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	5-Year	All
Macon	1.01	1.06	1.01	1	0.96	0.99	1.02	1.06	1.06	1.03	1.03	1.02
Madison	1	0.99	1.04	0.99	0.96	1.01	0.98	1.03	1.1	1.01	1.03	1.01
Martin	1.01	1	1	1	0.98	0.99	1	1.01	0.97	0.99	0.99	1.00
McDowell	1	1	0.98	0.96	0.99	1.03	1.06	1.01	0.99	0.99	1.02	1.00
Mecklenburg	1.03		0.96		1.03		1.12		1.04		1.08	1.04
Mitchell	1	1	1	0.98	1.01	1.05	0.95	0.98	0.98	1.03	1.00	1.00
Montgomery	0.95	0.97	1.07	0.96	0.94	1.04	1.07	1	0.99	1.01	1.02	1.00
Moore	1.03	1.02	0.97	0.99	1.05	0.98	0.96	1.06	1.03	1.01	1.01	1.01
Nash	1.03	0.99	0.99	0.99	0.96	0.97	1.02	1.01	1.06	1.03	1.02	1.01
New Hanover		0.97		1.01		0.98		1.09		0.98	1.02	1.01
Northampton	0.95	1	1.02	1.03	1.02	0.92	0.95	1.06	1	0.82	0.95	0.98
Onslow	1.05	1	1.06	1.07	1.04	0.99	0.97	1.01	1.04	1.04	1.01	1.03
Orange	1	0.99	1.06	1	1	1.01	0.95	0.99	1.08	1.02	1.01	1.01
Pamlico	1.03	0.98	1	1.01	0.98	0.99	0.96	0.98	1	0.95	0.98	0.99
Pasquotank	0.97	0.98	0.96	0.99	1.02	1	1.01	1.07	1.06	0.81	0.99	0.99
Pender	0.99	0.96	1.01	1.03	1	0.97	0.96	1.04	1.08	1.01	1.01	1.01
Perquimans	0.95	1.02	0.97	0.93	1.06	1	1.02	1.13	1.02	1.01	1.04	1.01
Person	1	0.98	1	1.02	1.01	0.97	1.01	1.06	1	1.04	1.02	1.01
Pitt	0.99	1.02	1.02	1.03	1	0.96	0.99	1.04	1	1.04	1.01	1.01
Polk	1.01	0.97	0.95	1.03	1.02	1.09	1.08	0.98	0.96	1.01	1.02	1.01
Randolph	0.94	0.99	1.02	0.97	1.05	0.97	0.93	1.1	1.04	0.97	1.00	1.00
Richmond	1.01	1.05	1.05	0.99	0.95	0.99	1.05	0.98	1.02	1.02	1.01	1.01
Robeson	0.99	0.99	1.01	0.98	0.99	1.01	1	1.02	1.06	1.01	1.02	1.01
Rockingham	1.01	1.02	1.03	1.02	0.99	0.99	0.95	0.96	1.03	1.07	1.00	1.01
Rowan	1.01	1.01	0.97	0.97	1	0.99	1	1.01	1	1.19	1.04	1.02
Rutherford	1.04	0.97	1.01	1.02	1	1.08	1.02	1.02	1	0.99	1.02	1.02
Sampson	1	1	1.03	1.04	1.02	0.95	0.98	1.03	1	1.09	1.01	1.01
Scotland	0.95	1	1.03	1.04	1.02	0.96	0.96	1.04	1.08	1.1	1.03	1.02
Stanly	0.98	1.04	1.02	0.94	1.01	0.99	0.94	1.07	1.03	1.01	1.01	1.00
Stokes	0.99	1.02	1.01	1	0.95	0.99	1.01	1.02	0.98	1.01	1.00	1.00
Surry	0.98	1.02	0.98	0.97	1	1.01	1	1.04	1	0.99	1.01	1.00
Swain	1.06	1.01	0.98	1.06	1.03	0.88	0.98	1.06	1	1	0.98	1.01
Transylvania	0.99	0.99	0.99	1	1	1	1.06	1.04	0.96	1.01	1.01	1.00
Tyrrell	1.04	0.99	0.98	0.98	1.05	1.08	0.94	0.93	1.01	1.16	1.02	1.02
Union	1.01	1.02	1.06	1.04	0.97	0.99	1.06	1.02	0.99	1.08	1.03	1.02
Vance	0.95	1.02	0.98	1.01	1.05	0.93	0.96	1.04	1.01	0.98	0.98	0.99
Wake		0.99		1.01		1.05		1.06		1.04	1.05	1.03
Warren	0.99	1	1.04	0.98	0.97	0.98	1.03	1.04	0.96	1.09	1.02	1.01
Washington	0.99	0.99	1.05	1.01	0.99	0.99	1.03	0.97	0.95	0.95	0.98	0.99
Watauga	1.01	1	1	0.97	0.93	1.05	1.09	1	1	1.01	1.03	1.01
Wayne	0.97	1.02	1.05	0.97	1.01	1.01	1	1.12	1.02	0.98	1.03	1.02
Wilkes	0.96	0.97	1	0.95	1.03	0.97	1.02	1.08	0.96	0.98	1.00	0.99
Wilson	1.01	0.98	1	0.99	1.02	0.96	0.97	1.09	1.04	0.94	1.00	1.00
Yadkin	1.02	1.02	0.96	0.96	0.97	0.99	0.99	1.06	1.01	0.97	1.00	1.00
Yancey	1.04	0.95	1	0.98	1.01	1.01	0.96	1	1.04	0.97	1.00	1.00
North Carolina	1	0.99	1	1	1	1	1.01	1.03	1.03	1.02	1.02	1.01

The following flowchart (Figure 40) illustrates the applicability of growth factors to estimate local road AADT at non-covered and available traffic count stations. As the data considered for modeling is the year 2015, it is considered as the base year.

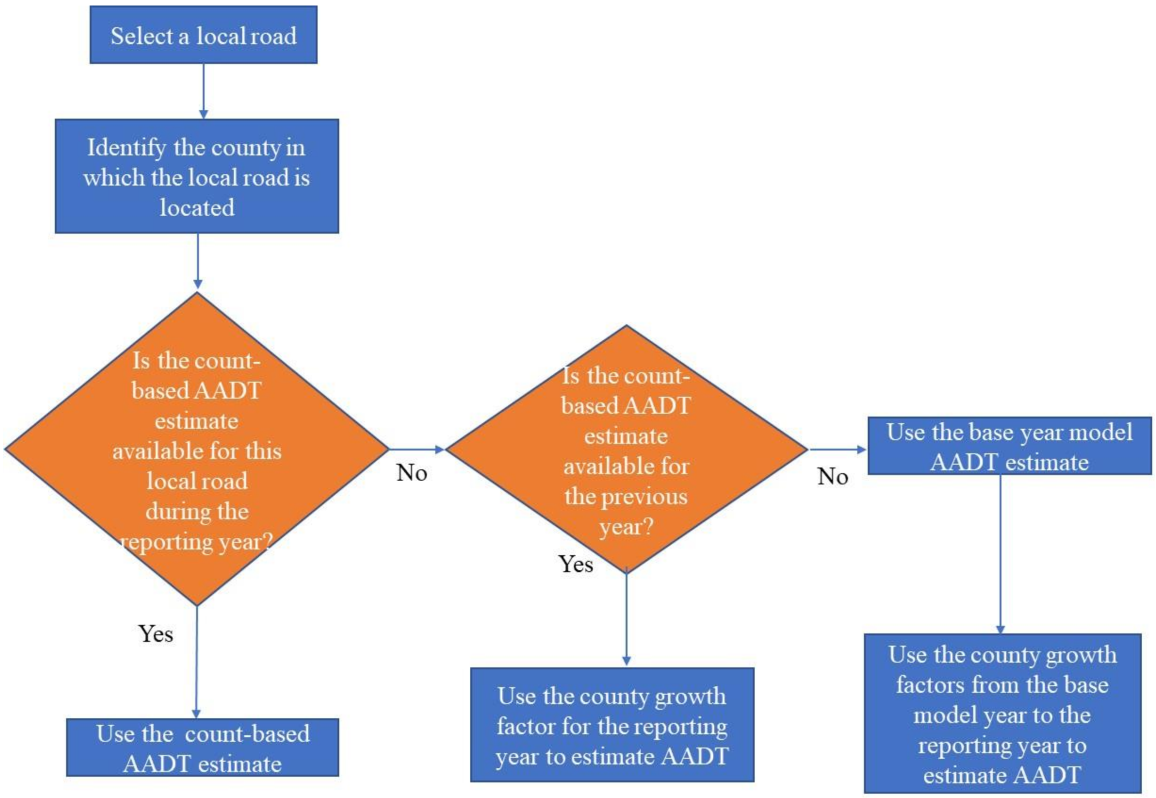


Figure 40 Application of growth factors to estimate local road AADT

The local road AADT estimates from the traffic count stations are reported directly. If traffic count data was not collected at a local road during the reporting year but is available for the previous year, the growth factor for the county in which the local road is located and the previous year count-based AADT are used to estimate AADT for the reporting year. For example, consider a local road in Columbus County at which count-based AADT =1,500 in the year 2016. Using the year 2017 growth factor for Columbus County (=1.05), the estimated AADT for the reporting year 2017, for this local road, is equal to 1,500 × 1.05 = 1,575. The mean growth factor for North Carolina or the 5-year average growth factor may be used if a growth factor could not be computed due to lack of an adequate number of local road traffic count stations for a county for a particular year.

If traffic count data was not collected at a local road during the reporting year or in any of the previous years, the growth factors for the county in which the local road is located and the estimated AADT for the base year are used to estimate AADT for the reporting year. For example, consider a local road in Columbus County at which traffic count data was never collected in the field. The estimated AADT during the base year (2015) for this local road link is 1,500. Using the year 2016 and year 2017 growth factors for Columbus County (1.03 and 1.05, respectively), the estimated AADT for the reporting year 2017, for this local road, is equal to $1{,}500 \times 1.03 \times 1.05 = 1{,}622$. The local road AADT using the recommended modeling method should be estimated every five years (or whenever TAZ-level data or census block-level data are updated and made available) for non-covered locations.

CHAPTER 9: MODEL ACCURACY BASED ON VOLUME RANGE

As per the FHWA definitions, local roads are generally not intended for long-distance travel. They provide direct access to land use developments. Also, through traffic is usually discouraged in the case of local roads. Hence, it is important to look int the count-based AADT ranges while developing the models to estimate AADT at non-covered locations.

9.1 Statewide Model with Count-based AADT<=3000

All local roads with a count-based AADT less than 5,000 were initially considered for model development. This value was selected based on consultations with the staff of NCDOT. However, 96% of the count-based AADT (count-based local road AADT less than 5,000) is less than 3,000. Moreover, the error analysis indicates that locations with a high count-based AADT have higher prediction errors. Therefore, the model accuracy was tested using data for local road count-based AADT less than 3,000. This analysis may help to redefine the count-based AADT ranges for local functionally classified road. The descriptive statistics for all the selected variables are summarized in Table 36.

The Pearson correlation coefficients are summarized in Table 37. The backward elimination method was adopted to develop the best-fitted model using the modified database. The OLS and GWR model results were developed. The validation was carried out using 25% of the data. The MAPE, MPE, and RMSE for the validation dataset are 85.1, -41.2, and 609, respectively based on the best-fitted OLS regression model. While comparing validation results with the statewide model illustrated in Chapter 6, the MAPE and MPE are found to be the same. However, the RMSE value is 609, and which is 20% less than the statewide AADT model.

Table 37 Descriptive statistics - count-based AADT<=3,000

Variables	Minimum	Median	Mean	Maximum	Std. Dev.
Count-based AADT	10	470	694	3000	626
# of lanes	1	2	2	4	
Speed limit (mph)	20	55	49.36	55	8.78
Surface type indicator (unpaved)	0	0	0.008	1	-
Surface type indicator (Bitumen)	0	1	0.859	1	-
Surface type indicator (Concrete)	0	0	0.132	1	-
Population	0	4.15	8.14	219.65	10.91
# of households	0	1.64	3.23	68.97	4.31
Workers	1	1.95	3.84	79.52	5.24
Industrial workers	0	0.09	0.54	46.20	1.83
Heavy industrial Workers	0	0.10	0.34	23.48	0.96
Retail workers	0	0.06	0.36	54.72	1.33
High retail employees	0	0.05	0.31	60.86	1.02
Office employees	0	0.08	0.47	112.26	2.02
Service employees	0	0.21	0.85	72.63	2.19
Government employees	0	0.03	0.26	50.20	1.44
Educational employees	0	0.07	0.31	298.46	2.82
Urban local road	0	0	0.21	1	-
Rural local road	0	1	0.79	1	-
Population density	0.81	109.66	214.95	5798.79	287.99
Employment density	0	26.66	92.78	14347.69	262.43
Road density (1-mile)	2.0	10.7	13.23	74.00	8.02
Distance to the nearest nonlocal road (miles) (Dis-nonlocal)	0.01	0.21	0.55	9.49	0.78
AADT at the nearest nonlocal road (AADT-nonlocal)	50	3200	5209	119000	6368

Table 38 Pearson correlation coefficients - AADT<=3000

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)	(17)	(18)	(19)
AADT (1)																			
Speed limit (2)	MN																		
# of lanes (3)	LP	LN																	
Type area (4)	MP	MN																	
Road density (5)	MP	HN	LP	HP															
Dis-nonlocal (6)	LN	LP		LN	LN														
AADT-nonlocal (7)	LP	LN		MP	LP	LN													
Population (2015) (8)	MP	MN	LP	HP	HP	LN	MP												
# of Households (9)	MP	MN	LP	HP	HP	LN	MP	HP											
Workers (10)	MP	MN	LP	HP	HP	LN	MP	HP	HP										
Industrial (11)	LP	LN	LP	LP	LP	LN	LP	MP	MP	LP									
High industrial (12)	LP	LN	LP	MP	MP	LN	LP	MP	MP	MP	MP								
Retail (13)	LP	LN		LP	MP	LN	LP	HP	HP	MP	MP	MP							
High retail (14)	LP	LN	LP	MP	MP	LN	LP	HP	HP	MP	MP	MP	HP						
Office (15)	LP	LN		LP	MP	LN	LP	HP	HP	MP	MP	MP	HP	HP					
Service (16)	LP	LN	LP	MP	MP	LN	LP	HP	HP	HP	MP	HP	HP	HP	HP				
Government (17)	LP	LN	LP	LP	LP	LN	LP	MP	MP	MP	LP	MP	MP	MP	MP	MP			
Education (18)	LP	LN		LP	LP	LN	LP	MP	LP	LP	LP	LP	MP	HP	MP	LP	MP		
Population density (19)	MP	MN	LP	HP	HP	LN	LP	HP	HP	HP	MP	MP	HP	HP	HP	HP	MP	MP	
Employment density (20)	LP	LN	LP	MP	MP	LN	LP	HP	HP	HP	HP	HP	HP	HP	HP	HP	HP	HP	HP

Note 2: HP, MP, LP, HN, MN, and LN are high positive, moderate positive, low positive, high negative, moderate negative, and low negative, respectively.

Similarly, the MAPE, MPE, and RMSE for the validation dataset are 78.1, -37.4, and 573, respectively based on the best-fitted GWR model. A comparative assessment between statewide AADT estimation model validation results and models developed with count-based AADT less than or equal to 3,000 is shown in Table 38.

Table 39 Comparison between statewide model and model with count-based AADT <=3000

Measure	OLS		GWR	
	Statewide	Count-based AADT<=3000	Statewide	Count-based AADT<=3000
MAPE (%)	86.1	85.1	84.1	78.1
MPE (%)	-44.2	-41.2	-44.2	-37.4
RMSE	771	609	733	573

9.2 County-Level Models with Count-based AADT<=3,000

Based on the availability of samples, Wake county and Iredell County were selected for modeling in this case. Duplin County or other counties were not selected as the number of count-based AADT id greater than 3,000 are zero or very few. The sample size for these counties is almost equal for <=3,000 and <=5,000 datasets. Therefore, a significant change in model and validation results are not observed for those counties.

The descriptive analysis and Pearson correlation coefficients assessment were carried out for Wake county and Iredell County modified datasets. OLS and GWR models were then developed. For the Wake County, the MAPE, MPE, and RMSE for the validation dataset are 80.1, -37.4, and 743, respectively based on the best-fitted OLS model. A comparison between the county-level models developed in Chapter 7 and the model developed using count-based AADT less than or equal to 3000 for Wake County is shown in Table 39. From Table 39, there exists a notable improvement in the accuracy of local road AADT estimates from GWR and OLS models developed using the count-based AADT values less than or equal to 3,000.

Table 40 County-level model using count-based AADT <=3000 – Wake County

Measure	OLS		GWR	
	County-level	Count-based AADT<=3000	County-level	Count-based AADT<=3000
MAPE (%)	120	80.1	120	82.3
MPE (%)	-88.3	-37.4	-86.2	-41.1
RMSE	993	743	962	732

The analysis results for the Iredell County are summarized in Table 40. A notable improvement in prediction accuracy of the models developed based on count-based AADT less than 3,000 was also observed even in the case of Iredell County.

Table 41 County-level model using AADT <=3000 – Iredell County

Measure	OLS		GWR	
	County-level	Count-based AADT<=3000	County-level	Count-based AADT<=3000
MAPE (%)	95.2	80.1	92.8	77.5
MPE (%)	-46.4	-39.5	-32.1	-36.5
RMSE	883	680	888	624

In summary, the model accuracy is better when samples with low count-based AADT values are considered for estimating the local road AADT. Hence, redefining the count-based AADT ranges for local functionally classified roads may improve the model predictability and estimates to a significant extent.

CHAPTER 10: CONCLUSIONS

Collecting traffic data and/or estimating and reporting AADT is important for planning, designing, building, and maintaining the road infrastructure. As local roads account for a major proportion of the road infrastructure in the state of North Carolina, it will also serve as an important variable in the road safety analysis and improvement programs. This research was mainly aimed at developing a sustainable and repeatable method to estimate AADT for all the local functionally classified roads.

A detailed literature review was conducted on AADT and VMT generation methods for functionally classified major, minor, and local roads. The most common methods used for estimating AADT at non-covered locations include statistical methods, geospatial methods, and machine learning methods. The predictability of geospatial methods over traditional statistical methods was illustrated in many of the past studies. This research adopted statistical and geospatial techniques to estimate local roads AADT. A survey was also conducted to gather information on other state DOT’s practices on meeting the HSIP and HPMS requirements. Some DOTs have undertaken (some ongoing) noteworthy research initiatives to estimate AADT at non-covered locations.

The model development was carried out in two levels: the statewide AADT estimation and county-level AADT estimation. This research examined five different modeling methods to estimate local roads AADT. They include traditional OLS regression, GWR, and geospatial interpolation techniques such as Kriging, IDW, and natural neighbor interpolation.

AADT based on traffic counts collected at 12,899 locations on local roads in North Carolina during the years 2014, 2015, and 2016 were considered as the dependent variable. The road, socioeconomic, demographic, and land use characteristics based on data gathered from NCDOT for the year 2015 were considered as the explanatory variables. The explanatory variables were screened by computing and comparing Pearson correlation coefficients. A detailed descriptive analysis was carried out to understand the relationship between count-based local road AADT and selected explanatory variables.

The statewide model development and validation results indicated that the GWR model performed relatively better when compared to other considered statistical and geospatial methods. GWR can incorporate the effect of spatial variations in data, by geographic location, when estimating the local road AADT. The errors in estimated local road AADT are lower for locations with a higher number of nearby traffic count stations.

Local road AADT estimation models were also developed based on functional classification type (urban/rural), speed limit, and population density. The results indicate that models for rural local roads, speed limit equal to 50 or 55 mph, and population density less than 200 people per square mile performed better than models for other categories. It can be concluded that road, socioeconomic, and demographic characteristics influence local road AADT and, hence, the model predictability.

The development of county-level local road AADT estimation models and incorporating land use data for modeling followed this task. Ten counties were considered for modeling based on the quality of land use data, population density, road density, and the number of AADT counts available in the county. A comparative assessment was carried out between the statewide and county-level model estimates. The MAPE, MPE, and RMSE were computed using the validation datasets and compared for the statewide estimates and the county-level estimates. The county-level models were observed to estimate local road AADT relatively better than the statewide models. The inclusion of land use variables in modeling can be mainly attributed to the improved performance of county-level models. The developed county-level models were used for estimating AADT at non-covered locations in each county.

The median prediction errors associated with statewide and county-level models were assessed to recommend future sampling requirements to improve model accuracy. The median prediction errors are higher for urban local roads and local roads with a speed limit greater than 25 mph and less than 50 mph. In most of the cases, the median prediction error seems to depend on the number of traffic count stations, count-based AADT, and county characteristics. The prediction errors were also low at local road AADT count locations near single-family residential units, multi-family residential units, and the commercial area. Contrarily, they are relatively higher at local road traffic count stations near schools, institutions, government, office, and industrial land uses. This could be attributed to differences in the number of local road traffic count stations by land use area type (more the number of count locations, lower the prediction error). A sampling plan based on the number of local road traffic count stations, functional classification type, speed limit ranges, and road connectivity type like dead-ends is recommended. Further, it is recommended to collect traffic counts and estimate spatially distributed count-based local road AADT data at 12,000 (based on the speed limit) to 22,000 (based on link connectivity, beginning and ending features) different stations biennially. This will help develop enhanced local road AADT estimation models.

Developing growth factors is very important as socioeconomic and demographic data at TAZ-level or census block level are updated and made available every 5 or 10 years. This research recommends the use of county-level growth factors based on count-based local road AADT counts for future AADT estimations. Count-based local road AADT and growth factor for the reporting year, for the county in which the local road is located, must be used if count-based AADT available for the previous years. For non-covered locations, the estimated AADT for the base year and growth factors from the base year to the reporting year must be used.

This research assessed the models' predictability by considering a lower count-based AADT range (AADT<=3,000) for model development and validation. The validation results showed a notable improvement in efficiency based on MAPE, MPE, and RMSE values. These results provide useful insights to redefine local road classification in the study area based on AADT ranges.

Overall, the generated models will minimize the costs associated with lapses in traffic count data collection programs and plans. The methodological framework adopted in this research can be adopted by other researchers and practitioners in the same field. The local road AADT estimates will also help the practitioners in planning and prioritizing road infrastructure projects for future improvements and air quality estimates, in addition to HSIP and HPMS reporting

10.1 Limitations and Scope for Future Work

This research can be further extended in several ways. The statewide model was developed using road characteristics and TAZ-level socioeconomic and demographic characteristics for the year 2015. The statewide travel demand model has 2,741 TAZs. This number is lower than the number of TAZs in the regional models developed and maintained by metropolitan planning organizations (MPOs) and rural planning organizations (RPOs). Considering available TAZ-level socioeconomic and demographic data for all MPOs and RPOs in the study area and using for modeling purposes may yield accurate local road AADT estimates.

The census data was not used as it was eight to nine years old at the time of this research. The census data at block-level could also be explored to develop the statewide local road AADT model using GWR.

Land use data were used along with road, socioeconomic and demographic data to develop county-level local road AADT models. These county-level models were observed to yield relatively better local road AADT estimates than the statewide model (for selected counties). However, the applicability of land use (parcel-level) information could not be tested using data for all counties in North Carolina. About 27% of statewide parcels do not have parcel descriptions. There are 26 counties in North Carolina without any land use data. Additionally, there are 4,744 unique land use descriptions of parcels in the county-level land use databases. Developing a land use dataset for the state with standardized attributes may improve the model accuracy to a significant extend.

Geospatial data such as socioeconomic, demographic, and land use characteristics were extracted using a 100 feet flat buffer. Road characteristics are for the subject local road link. While one-way dead-ends are not much affected, traffic on other local roads may be influenced by upstream and/or downstream link characteristics. Accounting for this as well as cross-street link characteristics may increase the predictability of the local road AADT models. However, objectively extracting these details for all the local roads (including non-covered locations) is not an easy task and requires robust tools. This should be explored in the future.

Advanced machine learning/deep learning techniques were not explored in this research as one of the key objectives is to investigate the role of socioeconomic, demographic, land use, and network characteristics on local road AADT. Developing AADT estimation models based on such methods and comparing them with findings from this research merit further research.

Probe data are being explored for travel time and pattern predictions. The number of probes detected on a link could be correlated to the AADT on the link. The possibility of using sampled probe data for AADT prediction or calibration also merits an investigation.

REFERENCES

1. 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(6), 493-506.
2. Asa, E., Mohamed S., Joseph M., & Arun B. (2012). Comparison of linear and nonlinear kriging methods for characterization and interpolation of soil data. Journal of Computing in Civil Engineering 26(1), 11-18.
3. Bartier, P. M., & Keller, C. P. (1996). Multivariate interpolation to incorporate thematic surface data using inverse distance weighting (IDW). Computers & Geosciences, 22(7), 795-799.
4. Bobach, T. A. (2008) Natural neighbor interpolation - Critical assessment and new results. Ph.D. Thesis, TU Kaiserslautern.
5. Brunsdon, C., Fotheringham, A. S., & Charlton, M. E. (1996). Geographically weighted regression: a method for exploring spatial nonstationarity. Geographical Analysis, 28(4), 281-298.
6. Castro, N., Manoel, Y. Jeong, M. K. Jeong, and L. D. Han. (2009). AADT prediction using support vector regression with data-dependent parameters. Expert Systems with Applications, 36 (2), 2979-2986.
7. Chow, L. F., Zhao, F., Liu, X., Li, M. T., & Ubaka, I. (2006). Transit ridership model based on geographically weighted regression. Transportation Research Record: Journal of the Transportation Research Board, 1972, 105-114.

8. Chowdhury, M., Sadek, A., Ma, Y., Kanhere, N., and Bhavsar, P. (2006). Applications of Artificial Intelligence Paradigms to Decision Support in Real-time Traffic Management, Transportation Research Record: Journal of the Transportation Research Board, 1968, 92-98.

9. Das, S., & Tsapakis, I. (2019). Interpretable machine learning approach in estimating traffic volume on low-volume roadways. International Journal of Transportation Science and Technology, in press. DOI: 10.1016/j.ijtst.2019.09.004

10. Du, H., & Mulley, C. (2006). Relationship between transport accessibility and land value: Local model approach with geographically weighted regression. Transportation Research Record: Journal of the Transportation Research Board, 1977, 197-205.

11. Duddu, V. R., & Pulugurtha, S. S. (2013). Principle of demographic gravitation to estimate annual average daily traffic: Comparison of statistical and neural network models. Journal of Transportation Engineering, 139(6), 585-595.

12. Eom, J., Park, M., Heo, T., & Huntsinger, L. (2006). Improving the prediction of annual average daily traffic for nonfreeway facilities by applying a spatial statistical method. Transportation Research Record: Journal of the Transportation Research Board, 1968, 20-29.

13. ESRI 2018. ArcGIS Desktop: Release 10.6.1 Redlands, CA: Environmental Systems Research Institute.

14. Fotheringham, A., Brunsdon, C., & Charlton, M. (2002). Geographically weighted regression: The analysis of spatially varying relationships. West Sussex, England: John Wiley & Sons, LTD.

15. FHWA (Federal Highway Administration). (2013). Highway Functional Classification Concepts, Criteria and Procedures. Washington, DC. Website:https://www.fhwa.dot.gov/planning/processes/statewide/related/highway_functional_classifications/fcauab.pdf, Accessed on 7th September 2019.

16. FHWA (Federal Highway Administration). (2005). Highway performance management system field manual. Washington, DC. Website: http://www.fhwa.dot.gov/policy/ohpi/hpms/fieldmanual. Accessed on 2nd December 2018.

17. 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. Proceedings of Transportation Research Board 87th Annual Meeting, Transportation Research Board, Washington, DC.

18. Gadda, S. C., Kockelman, K. M., & Magoon, A. (2007). Estimates of AADT: Quantifying the uncertainty. Eleventh World Conference on Transportation Research, Berkeley, California.

19. Goel, P. K., McCord, M. R., & Park, C. (2005). Exploiting correlations between link flows to improve estimation of average annual daily traffic on coverage count segments: Methodology and numerical study. Transportation Research Record, Journal of the Transportation Research Board, No. 1917, 100-107.

20. Gribov, A., & Krivoruchko, K. (2020). Empirical Bayesian kriging implementation and usage. Science of the Total Environment, 137290.

21. Gross, F. (2017). Highway Safety Improvement Program (HSIP) Evaluation Guide (No. FHWA-SA-17-039), Washington, DC.

22. Holik, W., Tsapakis, I., Vandervalk, A., Turner, S., & Habermann, J. (2017). Innovative traffic data QA/QC procedures and automating AADT estimation. Southeast Michigan Council of Government Case Study. Federal Highway Administration, Report No: FHWA-SA-17-035.

23. Islam, S. (2016). Estimation of annual average daily traffic (AADT) and missing hourly volume using artificial intelligence, All thesis. 2562, Clemson University, https://tigerprints.clemson.edu/all_theses/2562, Accessed on 28th December 2019.

24. Jayasinghe, A., & Sano, K. (2017). Estimation of annual average daily traffic on road segments: Network centrality-based method for metropolitan areas. Transportation Research Board Annual Meeting Compendium of Papers, No. 17-03141.

25. Jiang, Z., McCord, M. R., & Goel, P. K. (2006). Improved AADT estimation by combining information in image-and ground-based traffic data. Journal of transportation engineering, 132(7), 523-530.

26. Kis, I. M. (2016). Comparison of ordinary and universal Kriging interpolation techniques on a depth variable (a case of linear spatial trend), case study of the Sandrovac Field. Rudarsko-Geolosko-Naftni Zbornik, 31(2), 41.

27. Kusam, P. R., & Pulugurtha, S. S. (2015). Spatial proximity and dependency to model urban travel demand. Journal of Urban Planning and Development, 142(2), 2015, 04015014.

28. Lam W. H. K., Tang Y. F., & Tam, M. L. (2006). Comparison of two non-parametric models for daily traffic forecasting in Hong Kong. Journal of Forecasting, 25, 173-192.

29. Li, M. T., Zhao, F., Wu, Y. (2004). Application of regression analysis for identifying factors that affect seasonal traffic fluctuations in Southeast Florida. Transportation Research Record, Journal of the Transportation Research Board, No. 1870, 153-161.

30. Lingras, P., Sharma, S. C., Liu, G. X., & Xu, F. (2000). Estimation of annual average daily traffic on low-volume roads factor approach versus neural networks. Transportation Research Record, Journal of the Transportation Research Board, No. 1719, 103-111.

31. Ma, Y., Chowdhury, M., Sadek, A., and Jeihani, M. (2012) "Integration of Vehicle Infrastructure Integration (VII) system and Artificial Intelligence (AI) for online travel time prediction," IEEE Transactions on Intelligent Transportation Systems, 13 (3), 1369 – 1382.

32. Manes, A. S., & Pulugurtha, S. S. (2019). Modeling the influence of land use developments on transportation system performance. Mineta Transportation Institute Publications, Website: https://transweb.sjsu.edu/sites/default/files/1702A-Pulugurtha-Effect-Land-Use-Development-Travel-Time.pdf, Accessed on 7th September 2019.

33. McCord, M. R., and Goel, P.K. (2009). Estimating AADT from combined air photos and ground based data: system design, prototyping, and testing. USDOT Region V Regional University Transportation Center, NEXTRANS Project No 002OY01.

34. 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(1), 136-142.

35. Meijer, J. R., Huijbregts, M. A., Schotten, K. C., & Schipper, A. M. (2018). Global patterns of current and future road infrastructure. Environmental Research Letters, 13(6), 064006.

36. Mohamad, D., Sinha, K. C., Kuczek, T., & Scholer, C. F. (1998). Annual average daily traffic prediction model for county roads. Transportation Research Record: Journal of the Transportation Research Board, 1617, 69-77.

37. NCDOT (North Carolina Department of Transportation). (2014). Functional Classification System. Raleigh., https://connect.ncdot.gov/projects/planning/Planning%20Document%20Library/Functional%20Classification%20System.pdf, Accessed on 7th September 2019.

38. Neveu, A. J. (1983). Quick-response procedures to forecast rural traffic. Transportation Research Record, 944, 47-53.

39. Oliver, M. A., & Webster, R. (1990). Kriging: a method of interpolation for geographical information systems. International Journal of Geographical Information System 4(3), 313-332.

40. Park, N. (2004). Estimation of average annual daily traffic (AADT) using geographically weighted regression (GWR) method and geographic information system (GIS). Dissertation report, Florida International University, FL.

41. Pulugurtha, S. S., & Kusam, P. R. (2012). Modeling annual average daily traffic with integrated spatial data from multiple network buffer bandwidths. Transportation Research Record, 2291(1), 53-60.

42. Raja, P., Mehrnaz, D., & Anderson, M. D. (2018). Estimation of average daily traffic on low volume roads in Alabama. International Journal of Traffic and Transportation Engineering, 7(1) 1-6.

43. Robinson, T. P., & Metternicht, G. (2006). Testing the performance of spatial interpolation techniques for mapping soil properties. Computers and Electronics in Agriculture, 50(2), 97-108.

44. Saha, S. K., & Fricker, J. D. (1986). Traffic volume forecasting methods for rural state highways. Transportation Research Record: Journal of the Transportation Research Board, 1203, 1986 pp.10-26.

45. Seaver, W. L., Chatterjee, A., & Seaver, M. L. (2000). "Estimation of traffic volume on rural local roads." Transportation Research Record: Journal of the Transportation Research Board, 1719, 121-128.

46. Selby, B. & K. M. Kockelman. (2013). Spatial prediction of traffic levels in unmeasured locations: applications of universal kriging and geographically weighted regression. Journal of Transport Geography, Vol. 29, 2013, pp. 24-32.

47. Selby, B. F. (2011). Spatial prediction of AADT in unmeasured locations by universal kriging and microsimulation of vehicle holdings and car-market pricing dynamics, Doctoral dissertation, The university of Texas at Austin, https://repositories.lib.utexas.edu/handle/2152/47011, Accessed on 18th August 2019.

48. Selby, B., & Kockelman, M. K. (2011). Spatial prediction of AADT in unmeasured location by universal Kriging. Proceedings of Transportation Research Board 90th Annual Meeting, Transportation Research Board, Washington, DC.

49. 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(1), 04014022.

50. 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(1), 24-31.

51. 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(5), 426-432.

52. Smith, B. L., & Demetsky, M. J. (1997). Traffic flow forecasting: Comparison of modeling approaches. Journal of Transportation Engineering, Vol. 123 (1), 261-266.

53. Smith, B. L., Williams, B. M., & Oswald, R. K. (2002). Comparison of parametric and nonparametric models for traffic flow forecasting. Transportation Research Part: C, Vol. 10, 303-321.

54. Staats, W. N. (2016). Estimation of Annual Average Daily Traffic on Local Roads in Kentucky, Master’s Thesis, Civil Engineering Department, University of Kentucky.

55. Sun, X., & Das, S. (2015). Developing a Method for Estimating AADT on all Louisiana Roads, Louisiana Department of Transportation, Report No: FHWA/LA. 14/548.

56. Tang, Y. F., W. H. K. Lam, & Ng, P. L. P. (2003). Comparison of four modeling techniques for short-term AADT forecasting in Hong Kong. Journal of Transportation Engineering, 129(3), 271–277.

57. Tsapakis, I., Holik, W., Vandervalk, A., Turner, S., & Habermann, J. (2017). Procedures in traffic volume estimation. Virginia Department of Transportation Case Study. Federal Highway Administration, Report No: FHWA-SA-17-036.

58. U. S. Census Bureau, Department of commerce, 2010 Census Traffic Analysis Zone (TAZ) State-based. Website: https://catalog.data.gov/, Accessed on 11th December 2019.

59. United States Environment protection agency, an official website of the United States government, Website: https://www.epa.gov/, accessed on 11th December 2019.

60. Wang, T. (2012). Improved annual average daily traffic (AADT) estimation for local roads using parcel-level travel demand modeling. Dissertation report, Florida International University, FL.

61. Wang, T., Gan, A., & Alluri, P. (2013). Estimating annual average daily traffic for local roads for highway safety analysis. Transportation Research Record: Journal of the Transportation Research Board, 2398, 60-66.

62. Wang, X., & Kockelman, K. (2009). Forecasting network data: Spatial interpolation of traffic counts from Texas data. Transportation Research Record: Journal of the Transportation Research Board, 2105, 100-108.

63. Xia, Q., Zhao, F., Chen, Z., Shen, L., & Ospina, D. (1999). Estimation of annual average daily traffic for nonstate roads in a Florida county. Transportation Research Record: Journal of the Transportation Research Board, 1660, 32-40.

64. Yang, H., Lu, X., Cherry, C., Liu, X., & Li, Y. (2017). Spatial variations in active mode trip volume at intersections: a local analysis utilizing geographically weighted regression. Journal of Transport Geography, 64, 184-194.

65. Zhao, F., & Chung, S. (2001). Contributing factors of annual average daily traffic in a Florida county: exploration with geographic information system and regression models. Transportation Research Record: Journal of the Transportation Research Board, 1769, 113-122.

66. Zhao, F., & Park, N. (2004). Using geographically weighted regression models to estimate annual average daily traffic. Transportation Research Record: Journal of the Transportation Research Board, 1879, 99-107.

67. Zhong, M., & Hanson, B. L. (2009). GIS-based travel demand modeling for estimating traffic on low-class roads. Journal of Transportation Planning and Technology, 32(5), 423-439.

68. Zhong, M., Lingras, P., & Sharma, S. (2004). Estimation of missing traffic counts using factor, genetic, neural, and regression techniques. Transportation Research Part C: Emerging Technologies, 12(2), 139-166.

APPENDIX A: RESULTS FROM STATEWIDE MODELS

The descriptive statistics and Pearson correlation coefficient matrices from statewide models based on local road functional class type, speed limit, and population density are summarized in this Appendix.

Table A1 Descriptive statistics of the explanatory variables – local road functional type (urban)

Variable	Minimum	Median	Mean	Maximum	Std. Dev.
Speed limit (mph)	20	45	42	55	10
# of lanes	1	2	2	4	0
Surface type indicator (unpaved)	0	0	0	1	0
Surface type indicator (Bitumen)	0	1	1	1	0
Surface type indicator (Concrete)	0	0	0	1	0
Population	0.11	16.66	21.75	219.65	17.23
# of households	0.05	6.57	8.63	68.97	6.81
Workers	0.06	7.94	10.43	79.52	8.42
Industrial workers	0	0.34	1.49	46.20	3.48
Heavy industrial workers	0	0.46	1.05	23.48	1.96
Retail workers	0	0.42	1.23	54.72	2.81
High retail employees	0	0.43	1.04	60.86	2.12
Office employees	0	0.56	1.75	112.26	4.85
Service employees	0	1.19	2.85	72.63	5.34
Government employees	0	0.11	0.87	64.38	3.51
Educational employees	0	0.42	0.93	298.46	5.67
Population density	2.83	439.70	574.23	5,798.79	454.69
Employment density	0.27	135.72	304.06	143	573.85
Road density (1-mile)	4.77	22.05	23.28	74.00	8.63
Distance to the nearest nonlocal road (miles) (Dis-nonlocal)	0.00	0.11	0.27	3.63	0.37
AADT at the nearest nonlocal road (AADT-nonlocal)	50	7,300	9,891	151,000	9,913

Table A2 Descriptive statistics of the explanatory variables – local road functional type (rural)

Variable	Minimum	Median	Mean	Maximum	Std. Dev.
Speed limit (mph)	20	55	51	55	8
# of lanes	1	2	2	4	0
Surface type indicator (unpaved)	0	0	0	1	0
Surface type indicator (Bitumen)	0	1	1	1	0
Surface type indicator (Concrete)	0	0	0	1	0
Population	0.03	3.32	4.76	75.31	4.79
# of households	0.01	1.34	1.89	28.34	1.89
Workers	0	1.54	2.20	27.44	2.23
Industrial workers	0	0.06	0.33	32.68	1.05
Heavy industrial workers	0	0.07	0.17	14.56	0.33
Retail workers	0	0.04	0.16	15.78	0.49
High retail employees	0	0.03	0.15	6.80	0.38
Office employees	0	0.05	0.20	16.82	0.58
Service employees	0	0.15	0.42	29.48	1.04
Government employees	0	0.03	0.13	23.65	0.56
Educational employees	0	0.05	0.16	8.95	0.39
Population density	0.81	87.64	125.73	1,988.22	126.40
Employment density	0.00	17.93	45.72	2,557.55	94.11
Road density (1-mile)	2.00	9.22	10.73	45.51	5.67
Distance to the nearest nonlocal road (miles) (Dis-nonlocal)	0.00	0.24	0.62	9.49	0.84
AADT at the nearest nonlocal road (AADT-nonlocal)	70	2,600	4,126	103,000	4,788

Table A3 Descriptive statistics of the explanatory variables – speed limit <= 25 mph

Variable	Minimum	Median	Mean	Maximum	Std. Dev.
# of lanes	2	2	2	2	0
Surface type indicator (unpaved)	0	0	0.01	1	0.11
Surface type indicator (Bitumen)	0	1	0.91	1	0.29
Surface type indicator (Concrete)	0	0	0.08	1.00	0.27
Population	0.08	8.86	15.25	219.65	18.86
# of households	0.04	3.41	6.15	68.97	7.21
Workers	0.04	3.92	6.70	79.52	7.96
Industrial workers	0	0.19	1.17	33.89	2.97
Heavy industrial workers	0	0.24	0.84	23.01	2.03
Retail workers	0	0.33	0.97	32.28	2.31
High retail employees	0	0.27	0.89	60.86	3.50
Office employees	0	0.39	1.51	61.08	4.52
Service employees	0	0.64	2.05	41.09	3.77
Government employees	0	0.11	1.14	38.72	3.70
Educational employees	0	0.20	1.37	298.46	15.93
Urban local road	0	1	0.57	1	-
Rural local road	0	0	0.43	1	-
Population density	2.13	233.94	402.48	5,798.79	497.85
Employment density	1.68	84.20	268.06	14,347.69	837.60
Road density (1-mile)	4.15	23.64	24.80	57.17	10.03
Distance to the nearest nonlocal road (miles) (Dis-nonlocal)	0.01	0.09	0.22	4.07	0.39
AADT at the nearest nonlocal road (AADT-nonlocal)	135	6,200	7,671	36,000	6,032

Table A4 Descriptive statistics of the explanatory variables – speed limit = 30 mph or 35 mph

Variable	Minimum	Median	Mean	Maximum	Std. Dev.
# of lanes	1	2	2	1	2
Surface type indicator (unpaved)	0	0	0	0	-
Surface type indicator (Bitumen)	0	1	0.91	0	-
Surface type indicator (Concrete)	0	1	0.08	0	-
Population	0.03	9.31	15.61	150.72	17.58
# of households	0.01	3.79	6.29	62.40	7.03
Workers	0	4.48	7.29	70.76	8.41
Industrial workers	0	0.26	1.22	36.96	2.88
Heavy industrial workers	0	0.28	0.80	23.48	1.82
Retail workers	0	0.24	1.04	54.72	2.87
High retail employees	0	0.22	0.86	22.49	1.76
Office employees	0	0.31	1.50	112.26	4.97
Service employees	0	0.68	2.44	72.63	5.43
Government employees	0	0.09	0.79	64.35	3.40
Educational employees	0	0.19	0.65	16.98	1.37
Urban local road	0	1	0.54	1	-
Rural local road	0	0	0.46	1	-
Population density	0.81	245.91	411.93	3,979.04	463.85
Employment density	0.01	77.06	253.04	7,582.65	525.38
Road density (1-mile)	3.70	21.69	22.53	74.00	9.44
Distance to the nearest nonlocal road (miles) (Dis-nonlocal)	0.02	0.11	0.27	6.69	0.46
AADT at the nearest nonlocal road (AADT-nonlocal)	150	5,900	8,108	119,000	8,100

Table A5 Descriptive statistics of the explanatory variables – speed limit = 40 mph or 45 mph

Variable	Minimum	Median	Mean	Maximum	Std. Dev.
# of lanes	1	2	2	0	2
Surface type indicator (unpaved)	0	0	0.01	0.07	-
Surface type indicator (Bitumen)	0	1	0.91	0.28	-
Surface type indicator (Concrete)	0	0	0.08	0.28	-
Population	0	9	13	13	109
# of households	0.14	9.09	13.13	109.22	13.08
Workers	0.04	3.60	5.11	51.59	5.04
Industrial workers	0.05	4.50	6.50	70.36	6.75
Heavy industrial workers	0	0.14	0.83	46.20	2.62
Retail workers	0	0.20	0.48	15.74	0.95
High retail employees	0	0.14	0.52	34.34	1.41
Office employees	0	0.10	0.43	16.96	0.97
Service employees	0	0.19	0.66	53.70	1.88
Government employees	0	0.47	1.19	66.36	2.57
Educational employees	0	0.05	0.24	13.21	0.80
Urban local road	0	0	0.41	1	-
Rural local road	0	1	0.59	1	-
Population density	3.77	239.88	346.52	345.31	2,883.49
Employment density	0.57	54.02	128.12	231.50	4,849.40
Road density (1-mile)	3.50	14.42	15.58	6.83	50.58
Distance to the nearest nonlocal road (miles) (Dis-nonlocal)	0.04	0.17	0.41	0.56	4.62
AADT at the nearest nonlocal road (AADT-nonlocal)	110	4,800	7,400	8,979	151,000

Table A6 Descriptive statistics of the explanatory variables – speed limit = 50 mph or 55 mph

Variable	Minimum	Median	Mean	Maximum	Std. Dev.
# of lanes	1	2	2	4	0
Surface type indicator (unpaved)	0	0	0.01	1.00	-
Surface type indicator (Bitumen)	0	0	0.83	1.00	-
Surface type indicator (Concrete)	0	0	0.16	1.00	-
Population	0.11	3.33	5.64	86.52	6.95
# of households	0.05	1.34	2.23	44.63	2.76
Workers	0.05	1.56	2.64	65.06	3.41
Industrial workers	0	0.07	0.36	33.71	1.31
Heavy industrial workers	0	0.07	0.22	20.81	0.61
Retail workers	0	0.04	0.19	16.12	0.61
High retail employees	0	0.03	0.18	17.18	0.58
Office employees	0	0.05	0.25	33.96	0.89
Service employees	0	0.15	0.51	68.00	1.51
Government employees	0	0.02	0.14	64.38	1.00
Educational employees	0	0.05	0.19	10.58	0.48
Urban local road	0	0	0.10	1	-
Rural local road	0	1	0.90	1	-
Population density	2.84	88.04	148.98	2,284.23	183.39
Employment density	0.00	17.84	54.82	6,186.68	145.42
Road density (1-mile)	2.00	8.84	10.37	43.88	5.56
Distance to the nearest nonlocal road (miles) (Dis-nonlocal)	0.00	0.25	0.66	9.49	0.86
AADT at the nearest nonlocal road (AADT-nonlocal)	50	2,600	4243	103,000	5,468

Table A7 Descriptive statistics of the explanatory variables – Population density < 200 people/ square mile

Variable	Minimum	Median	Mean	Maximum	Std. Dev.
# of lanes	1	2	2	4	0
Speed limit	20	55	51	55	8
Surface type indicator (unpaved)	0	0	0	1	0
Surface type indicator (Bitumen)	0	1	1	1	0
Surface type indicator (Concrete)	0	0	0	1	0
Population	0.03	3.00	3.22	7.57	1.83
# of households	0.01	1.19	1.29	3.78	0.73
Workers	0	1.33	1.48	4.99	0.89
Industrial workers	0	0.05	0.25	32.68	1.09
Heavy industrial workers	0	0.06	0.12	23.48	0.47
Retail workers	0	0.03	0.11	17.00	0.50
High retail employees	0	0.02	0.09	7.84	0.21
Office employees	0	0.04	0.13	53.70	0.95
Service employees	0	0.12	0.27	66.36	1.18
Government employees	0	0.02	0.08	12.19	0.27
Educational employees	0	0.04	0.09	5.76	0.17
Urban local road	0	0	0.04	1	-
Rural local road	0	1	0.96	1	-
Employment density	0	15.39	30.52	4,849.41	99.77
Road density (1-mile)	2.00	8.81	10.54	43.11	5.95
Distance to the nearest nonlocal road (miles) (Dis-nonlocal)	0.02	0.24	0.64	9.49	0.88
AADT at the nearest nonlocal road (AADT-nonlocal)	70	2,500	4,076	151,000	5,118

Table A8 Descriptive statistics of the explanatory variables – Population density = 200 – 400 people/ square mile

Variable	Minimum	Median	Mean	Maximum	Std. Dev.
# of lanes	1	2	2	4	0
Speed limit	20	50	47	55	9
Surface type indicator (unpaved)	0	0	0.01	1	-
Surface type indicator (Bitumen)	0	1	0.90	1	-
Surface type indicator (Concrete)	0	0	0.09	1	-
Population	7.58	10.70	10.75	15.21	2.11
# of households	1.10	4.12	4.26	7.11	0.91
Workers	1.33	4.96	5.12	8.76	1.15
Industrial workers	0	0.24	0.80	31.81	1.95
Heavy industrial workers	0	0.29	0.53	14.56	1.05
Retail workers	0	0.18	0.42	11.61	0.81
High retail employees	0	0.17	0.36	12.02	0.67
Office employees	0	0.25	0.61	70.73	2.36
Service employees	0	0.60	1.15	49.47	2.13
Government employees	0	0.07	0.29	16.53	1.04
Educational employees	0	0.25	0.39	6.32	0.55
Urban local road	0	0	0.43	1	-
Rural local road	0	1	0.57	1	-
Employment density	1.62	73.76	122.41	4,970.13	211.80
Road density (1-mile)	2.95	15.13	16.58	49.15	7.07
Distance to the nearest nonlocal road (miles) (Dis-nonlocal)	0.01	0.16	0.38	3.63	0.47
AADT at the nearest nonlocal road (AADT-nonlocal)	50	4,500	6,713	83,000	6,859

Table A9 Descriptive statistics of the explanatory variables – Population density = 400 - 600 people/ square mile

Variable	Minimum	Median	Mean	Maximum	Std. Dev.
# of lanes	1	2	2	4	0
Speed limit	20	50	47	55	9
Surface type indicator (unpaved)	0	0	0.01	1	-
Surface type indicator (Bitumen)	0	1	0.90	1	-
Surface type indicator (Concrete)	0	0	0.09	0	-
Population	15.15	18.69	18.73	22.71	2.32
# of households	0.98	7.39	7.42	10.83	1.17
Workers	1.14	8.57	8.85	13.71	1.64
Industrial workers	0.00	0.41	1.34	33.89	2.58
Heavy industrial workers	0.00	0.48	0.78	17.69	1.14
Retail workers	0.00	0.43	0.99	20.35	1.80
High retail employees	0.00	0.56	0.94	12.21	1.22
Office employees	0.00	0.61	1.18	29.17	1.96
Service employees	0.00	1.25	2.11	37.26	2.98
Government employees	0.00	0.11	0.62	23.66	1.91
Educational employees	0.00	0.48	0.69	24.69	1.24
Urban local road	15.15	18.69	18.73	22.71	2.32
Rural local road	0.98	7.39	7.42	10.83	1.17
Employment density	18.05	3,296.42	233.53	143.85	287.15
Road density (1-mile)	5.58	52.11	20.69	19.33	7.86
Distance to the nearest nonlocal road (miles) (Dis-nonlocal)	0.07	3.31	0.37	0.14	0.47
AADT at the nearest nonlocal road (AADT-nonlocal)	330	83,000	9,183	7,200	8,425

Table A10 Descriptive statistics of the explanatory variables – Population density = 600 - 800 people/ square mile

Variable	Minimum	Median	Mean	Maximum	Std. Dev.
# of lanes	1	2	2	3	0
Speed limit	20	45	43	55	9
Surface type indicator (unpaved)	0	0	0.01	1	-
Surface type indicator (Bitumen)	0	1	0.91	1	-
Surface type indicator (Concrete)	0	0	0.08	1	-
Population	22.73	27.00	26.91	30.25	2.26
# of households	6.65	10.87	10.74	15.58	1.39
Workers	8.09	12.96	13.13	19.92	2.10
Industrial workers	0	0.59	1.60	46.20	3.49
Heavy industrial workers	0	0.82	1.20	23.40	2.03
Retail workers	0	0.60	1.26	9.43	1.60
High retail employees	0	0.70	1.15	12.48	1.32
Office employees	0	0.98	1.90	21.02	2.56
Service employees	0	2.65	3.28	35.30	3.34
Government employees	0	0.25	0.74	13.21	1.71
Educational employees	0	0.79	1.07	7.89	1.11
Urban local road	0	1	0.83	1	-
Rural local road	0	0	0.17	1	-
Employment density	15.60	258.32	330.93	2,112.45	311.31
Road density (1-mile)	6.19	21.70	22.76	55.62	8.12
Distance to the nearest nonlocal road (miles) (Dis-nonlocal)	0.00	0.10	0.22	1.86	0.29
AADT at the nearest nonlocal road (AADT-nonlocal)	470	6,300	8,985	119,000	10,685

Table A11 Descriptive statistics of the explanatory variables – Population density > 800 people/ square mile

Variable	Minimum	Median	Mean	Maximum	Std. Dev.
# of lanes	1	2	2	4	0
Speed limit	20	35	40	55	9
Surface type indicator (unpaved)	0	0	0.02	1	-
Surface type indicator (Bitumen)	0	1	0.93	1	1
Surface type indicator (Concrete)	0	0	0.05	1	1
Population	30.37	43.58	47.89	219.65	18.59
# of households	6.86	16.79	18.84	68.97	7.44
Workers	6.70	20.58	22.72	79.52	9.49
Industrial workers	0	0.93	2.73	36.96	4.83
Heavy industrial workers	0	1.14	1.96	22.06	2.61
Retail workers	0	1.45	2.89	54.72	4.87
High retail employees	0	1.52	2.52	60.86	3.71
Office employees	0	2.21	4.36	112.26	7.88
Service employees	0.29	4.30	6.83	72.63	8.26
Government employees	30.37	43.58	47.89	219.65	18.59
Educational employees	6.86	16.79	18.84	68.97	7.44
Urban local road	0	1	0.95	1	-
Rural local road	0	0	0.05	1	-
Employment density	32.71	467.60	712.57	1,4347.69	956.64
Road density (1-mile)	6.88	28.44	28.81	74.00	10.13
Distance to the nearest nonlocal road (miles) (Dis-nonlocal)	0.01	0.09	0.21	2.20	0.28
AADT at the nearest nonlocal road (AADT-nonlocal)	135	8,300	12,030	103,000	11,503

Table A12 Correlation matrix for functional classification type - rural

Variable	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)	(17)	(18)	(19)	(20)	(21)
AADT (1)																					
Speed limit (2)	LN																				
# of lanes (3)	LP	LN																			
Unpaved (4)	LN																				
Bitumen (5)	LP	LN		LN																	
Concrete (6)	LN	LP		LN	HN																
Road density (7)	MP	HN	LP		LP	LN															
Dis-nonlocal (8)	LN	LP			LN	LP	LN														
AADT-nonlocal (9)	LP	LN					LP	LN													
Population (2015) (10)	MP	LN		LN	LP	LN	MP	LN	LP												
# of Households (11)	LP	LN		LN	LP	LN	MP	LN	LP	HP											
Workers (12)	MP	LN		LN	LP	LN	MP	LN	LP	HP	HP										
Industrial (13)	LP	LN			LP		LP	LN	LP	MP	LP	LP									
High industrial (14)	LP	LN			LP	LN	LP	LN	LP	HP	HP	HP	LP								
Retail (15)	LP	LN			LP	LN	LP	LN	LP	MP	MP	MP	LP	MP							
High retail (16)	LP	LN			LP		LP	LN	LP	HP	HP	HP	MP	HP	HP						
Office (17)	LP	LN	LP		LP	LN	LP	LN	LP	HP	HP	HP	MP	HP	HP	HP					
Service (18)	LP	LN			LP	LN	LP	LN	LP	HP	HP	HP	MP	HP	HP	HP	HP				
Government (19)	LP	LN			LP	LN	LP	LN	LP	MP	MP	LP	MP	LP	LP	MP	MP	MP			
Education (20)	LP	LN					LP	LN	LP	HP	HP	HP	LP	MP	MP	MP	MP	MP	MP		
Population density (21)	MP	LN		LN	LP	LN	MP	LN	LP	HP	HP	HP	MP	HP	MP	HP	HP	HP	MP	HP	
Employment density (22)	LP	LN			LP	LN	MP	LN	LP	HP	HP	HP	HP	HP	HP	HP	HP	HP	HP	HP	HP

Note: HP, MP, LP, HN, MN, and LN are high positive, moderate positive, low positive, high negative, moderate negative, and low negative, respectively.

Table A13 Correlation matrix for functional classification type - urban

Variable	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)	(17)	(18)	(19)	(20)	(21)
AADT (1)																					
Speed limit (2)	LN																				
# of lanes (3)	LP																				
Unpaved (4)		LP																			
Bitumen (5)		LN	LP	LN																	
Concrete (6)		LP			HN																
Road density (7)	LP	MN	LP		LP	LN															
Dis-nonlocal (8)	LN	LP					LN														
AADT-nonlocal (9)	LP	LN		LP			LP	LP													
Population (2015) (10)	LP	LN	LP	LP	LP	LN	MP	LN	LP												
# of Households (11)	LP	LN	LP	LP	LP	LN	MP	LN	LP	HP											
Workers (12)	LP	LN	LP	LP	LP	LN	MP	LN	LP	HP	HP										
Industrial (13)	LP	LN	LP				LP	LN	LP	LP	LP	LP									
High industrial (14)	LP	LN	LP	LP		LN	LP	LN	LP	LP	LP	LP	MP								
Retail (15)	LP	LN	LP				LP	LN	LP	MP	MP	MP	LP	MP							
High retail (16)	LP	LN	LP				MP	LN	LP	HP	HP	MP	LP	MP	HP						
Office (17)	LP	LN	LP				MP	LN	LP	MP	MP	MP	MP	MP	HP	HP					
Service (18)	LP	LN	LP				MP	LN	LP	MP	HP	MP	MP	HP	HP	HP	HP				
Government (19)	LP	LN	LP			LP	LP	LN	LP	LP	LP	LP	LP	LP	LP	MP	MP	HP			
Education (20)	LP	LN					LP			MP	LP	LP	LP	LP	LP	HP	LP	LP	LP		
Population density (21)	LP	LN	LP	LP	LP		MP	LN	LP	HP	HP	HP	LP	LP	MP	HP	MP	MP	LP	MP	
Employment density (22)	LP	LN	LP				MP	LN	LP	HP	HP	MP	HP	HP	HP	HP	HP	HP	HP	HP	HP

Note: HP, MP, LP, HN, MN, and LN are high positive, moderate positive, low positive, high negative, moderate negative, and low negative, respectively.

Table A14 Correlation matrix for speed limit ≤ 25 mph

Variable	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)	(17)	(18)	(19)	(20)	(21)
AADT (1)																					
# of lanes (2)																					
Functional class type (3)																					
Unpaved (4)																					
Bitumen (5)				MN																	
Concrete (6)					HN																
Road density (7)	LP		MP																		
Dis-nonlocal (8)	LN		LN																		
AADT-nonlocal (9)			MP				LP	LN													
Population (2015) (10)			MP				MP		LP												
# of Households (11)			MP				HP		LP	HP											
Workers (12)			MP				MP		LP	HP	HP										
Industrial (13)			LP						LP	LP	LP	LP									
High industrial (14)			LP	MP	LN		LP		LP	MP	MP	MP	LP								
Retail (15)			LP				MP		LP	HP	HP	HP	LP	LP							
High retail (16)			LP				LP		LP	HP	HP	HP	LP	LP	HP						
Office (17)			LP		LN	LP	LP		LP	HP	HP	HP	LP	MP	HP	HP					
Service (18)	LP		LP	LP			LP		LP	HP	HP	HP	LP	MP	HP	HP	HP				
Government (19)			LP		LN	LP	LP		LP	HP	MP	MP	MP	MP	HP	HP	HP	HP			
Education (20)							LP			HP	MP	HP		LP	HP	HP	HP	HP	HP		
Population density (21)			MP				MP		LP	HP	HP	HP	LP	MP	HP	HP	HP	HP	HP	HP	
Employment density (22)			LP				LP		LP	HP	HP	HP	LP	MP	HP	HP	HP	HP	HP	HP	HP

Note: HP, MP, LP, HN, MN, and LN are high positive, moderate positive, low positive, high negative, moderate negative, and low negative, respectively.

A15 Correlation matrix for speed limit = 30 mph or 35 mph

Variable	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)	(17)	(18)	(19)	(20)	(21)
AADT (1)																					
# of lanes (2)	LP																				
Functional class type (3)	MP																				
Unpaved (4)																					
Bitumen (5)	LP	LP	LP	LN																	
Concrete (6)	LN		LN		HN																
Road density (7)	MP	LP	MP		LP	LN															
Dis-nonlocal (8)	LN		LN				LN														
AADT-nonlocal (9)	LP	LP	MP				LP														
Population (2015) (10)	MP	LP	HP	LP	LP	LN	HP	LN	LP												
# of Households (11)	MP	LP	HP		LP	LN	HP	LN	LP	HP											
Workers (12)	MP	LP	HP	LP	LP	LN	HP	LN	LP	HP	HP										
Industrial (13)	LP	LP	LP		LP	LN	LP	LN	LP	MP	MP	MP									
High industrial (14)	LP	LP	LP		LP	LN	MP	LN	LP	MP	MP	MP	MP								
Retail (15)	LP	LP	LP		LP	LN	MP	LN	LP	MP	MP	MP	MP	MP							
High retail (16)	MP	LP	MP			LN	MP	LN	LP	HP	HP	HP	MP	MP	HP						
Office (17)	LP	LP	LP		LP	LN	MP	LN	LP	MP	MP	MP	MP	HP	HP	HP					
Service (18)	MP	LP	LP		LP	LN	MP	LN	LP	HP	HP	HP	HP	HP	HP	HP	HP				
Government (19)	LP	LP	LP				LP	LN	LP	MP	MP	LP	LP	LP	LP	MP	MP	HP			
Education (20)	LP		LP				MP	LN	LP	HP	HP	HP	MP	MP	MP	HP	MP	HP	MP		
Population density (21)	MP	LP	HP	LP	LP	LN	HP	LN	LP	HP	HP	HP	MP	MP	MP	HP	MP	HP	MP	HP	
Employment density (22)	MP	LP	MP		LP	LN	MP	LN	LP	HP	HP	HP	HP	HP	HP	HP	HP	HP	HP	HP	HP

Note: HP, MP, LP, HN, MN, and LN are high positive, moderate positive, low positive, high negative, moderate negative, and low negative, respectively.

Table A16 Correlation matrix for speed limit = 40 mph or 45 mph

Variable	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)	(17)	(18)	(19)	(20)	(21)
AADT (1)	1																				
# of lanes (2)	LN	1																			
Functional class type (3)	MP		1																		
Unpaved (4)				1																	
Bitumen (5)				LN	1																
Concrete (6)					HN	1															
Road density (7)	MP		HP				1														
Dis-nonlocal (8)	LN		LN				LN	1													
AADT-nonlocal (9)	LP	LN	LP				MP		1												
Population (2015) (10)	MP		HP		LP	LN	HP	LN	LP	1											
# of Households (11)	MP		HP		LP	LN	HP	LN	LP	HP	1										
Workers (12)	MP		HP		LP	LN	HP	LN	LP	HP	HP	1									
Industrial (13)	LP		LP				LP	LN	LP	LP	LP	LP	1								
High industrial (14)	LP		MP				MP	LN	LP	MP	MP	MP	MP	1							
Retail (15)	LP		LP				LP	LN	LP	MP	MP	MP	LP	MP	1						
High retail (16)	LP		LP				MP	LN	LP	MP	HP	MP	LP	MP	HP	1					
Office (17)	LP		LP		LP	LN	LP	LN	MP	MP	MP	MP	MP	HP	HP	HP	1				
Service (18)	LP		LP			LN	LP	LN	MP	MP	MP	MP	MP	HP	HP	HP	HP	1			
Government (19)	LP		LP				LP	LN		LP	LP	LP	LP	LP	LP	LP	LP	LP	1		
Education (20)	LP		LP		LP	LN	LP	LN	LP	MP	MP	MP	LP	LP	LP	MP	MP	LP	LP	1	
Population density (21)	MP		HP		LP	LN	HP	LN	LP	HP	HP	HP	LP	MP	MP	MP	MP	MP	LP	MP	1
Employment density (22)	LP		MP		LP	LN	MP	LN	MP	MP	MP	MP	HP	HP	HP	HP	HP	HP	MP	MP	MP

Note: HP, MP, LP, HN, MN, and LN are high positive, moderate positive, low positive, high negative, moderate negative, and low negative, respectively.

Table A17 Correlation matrix for speed limit = 50 mph or 55 mph

Variable	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)	(17)	(18)	(19)	(20)	(21)
AADT (1)																					
# of lanes (2)	LP																				
Functional class type (3)	LP																				
Unpaved (4)	LN		LP																		
Bitumen (5)		LP	LP	LN																	
Concrete (6)			LN	LN	HN																
Road density (7)	MP		HP	LP	LP	LN															
Dis-nonlocal (8)	LP		LN			LP	LN														
AADT-nonlocal (9)	LP		LP	LP			MP	LN													
Population (2015) (10)	MP	LP	HP	LP	LP	LN	HP	LN	LP												
# of Households (11)	MP		HP	LP	LP	LN	HP	LN	LP	HP											
Workers (12)	MP	LP	HP	LP	LP	LN	HP	LN	LP	HP	HP										
Industrial (13)	LP		LP				LP	LN	LP	LP	LP	LP									
High industrial (14)	LP		MP		LP	LN	MP	LN	LP	MP	MP	MP	MP								
Retail (15)	LP		MP	LP	LP	LN	MP	LN	LP	HP	HP	HP	MP	MP							
High retail (16)	LP		MP	LP			MP	LN	LP	HP	HP	HP	MP	MP	HP						
Office (17)	LP		MP	LP		LN	MP	LN	LP	HP	HP	HP	MP	MP	HP	HP					
Service (18)	LP		MP	LP		LN	MP	LN	LP	HP	HP	HP	MP	MP	HP	HP	HP				
Government (19)	LP		LP				LP	LN	LP	LP	LP	LP	LP	LP	LP	HP	MP	HP			
Education (20)	LP	LP	MP				LP	LN	LP	HP	HP	HP	LP	LP	MP	HP	MP	HP	MP		
Population density (21)	MP	LP	HP	LP	LP	LN	HP	LN	LP	HP	HP	HP	LP	MP	HP	HP	HP	HP	LP	HP	
Employment density (22)	LP		MP	LP		LN	MP	LN	LP	HP	HP	HP	HP	HP	HP	HP	HP	HP	HP	HP	HP

Note: HP, MP, LP, HN, MN, and LN are high positive, moderate positive, low positive, high negative, moderate negative, and low negative, respectively.

Table A18 Correlation matrix for population density < 200 people/square mile

Variable	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)	(17)	(18)	(19)	(20)	(21)
AADT (1)																					
Speed limit (2)	LN																				
# of lanes (3)	LP	LN																			
Functional class type (4)	LP	LN	LP																		
Unpaved (5)	LN	LP																			
Bitumen (6)	LP	LN			LN																
Concrete (7)	LN	LP			LN	HN															
Road density (8)	MP	HN	LP	MP		LP	LN														
Dis-nonlocal (9)	LN	LP		LN			LP	LN													
AADT-nonlocal (10)	LP	LN	LP	LP				MP	LN												
Population (205) (11)	LP	LN		LP	LN	LP	LN	MP	LN	LP											
# of Households (12)	LP	LN		LP	LN	LP	LN	MP	LN	LP	HP										
Workers (13)	LP	LN		LP	LN	LP	LN	MP	LN	LP	HP	HP									
Industrial (14)	LP	LN	LP	LP				LP	LN	LP	LP	LP	LP								
High industrial (15)	LP	LN	LP	LP		LP	LN	LP	LN	LP	LP	LP	LP	MP							
Retail (16)	LP	LN	LP	LP		LP	LN	LP	LN	LP	LP	LP	LP	MP	HP						
High retail (17)	LP	LN	LP	LP				LP	LN	LP	MP	MP	LP	MP	HP	HP					
Office (18)	LP	LN	MP	LP				LP	LN	LP	LP	LP	LP	MP	HP	HP	HP				
Service (19)	LP	LN	MP	LP		LP	LN	LP	LN	LP	LP	LP	LP	HP	HP	HP	HP	HP			
Government (20)	LP	LN	LP	LP				LP	LN	LP	LP	LP	LP	LP	LP	LP	LP	LP	LP		
Education (21)	LP	LN	LP	LP				LP	LN	LP	MP	MP	LP	LP	LP	LP	LP	LP	LP	LP	
Employment density (22)	LP	LN	LP	LP		LP	LN	LP	LN	LP	LP	LP	LP	HP	HP	HP	HP	HP	HP	LP	LP

Note: HP, MP, LP, HN, MN, and LN are high positive, moderate positive, low positive, high negative, moderate negative, and low negative, respectively.

Table A19 Correlation matrix for population density = 200 - 400 people/ square mile

Variable	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)	(17)	(18)	(19)	(20)	(21)
AADT (1)																					
Speed limit (2)	LN																				
# of lanes (3)																					
Functional class type (4)	LP	MN																			
Unpaved (5)																					
Bitumen (6)			LP		LN																
Concrete (7)																					
Road density (8)	LP	MN		MP																	
Dis-nonlocal (9)	LN	LP		LN				LN													
AADT-nonlocal (10)	LP	LN		LP				LP	LN												
Population (205) (11)	LP	LN		LP				LP	LN	LP											
# of Households (12)	LP	LN		LP				LP	LN	LP	HP										
Workers (13)	LP	LN		LP				LP	LN	LP	HP	HP									
Industrial (14)	LP	LN		LP				LP	LN	LP	LP	LP	LP								
High industrial (15)	LP	LN		LP	LP			LP		LP	LP	LP	LP	MP							
Retail (16)	LP	LN		LP				LP	LN	LP	LP	LP	LP	MP	MP						
High retail (17)	LP	LN		LP				LP	LN	LP	LP	LP	LP	MP	MP	HP					
Office (18)	LP	LN		LP				LP	LN	LP	LP	LP	LP	HP	MP	HP	HP				
Service (19)	LP	LN		LP				LP	LN	LP	LP	LP	LP	HP	MP	HP	HP	HP			
Government (20)		LN		LP				LP	LN	LP	LP			MP	MP	LP	MP	MP	MP		
Education (21)	LP	LN		LP				LP	LN	LP	LP	LP	LP	LP	LP	MP	MP	MP	MP	LP	
Employment density (22)	LP	LN		LP				LP	LN	LP	LP	LP	LP	HP	HP	HP	HP	HP	HP	HP	MP

Note: HP, MP, LP, HN, MN, and LN are high positive, moderate positive, low positive, high negative, moderate negative, and low negative, respectively.

Table A20 Correlation matrix for speed limit – population density = 400 - 600 people/square mile

Variable	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)	(17)	(18)	(19)	(20)	(21)
AADT (1)																					
Speed limit (2)	LN																				
# of lanes (3)																					
Functional class type (4)	LP	LN																			
Unpaved (5)																					
Bitumen (6)			LP		LN																
Concrete (7)		LP	LN			HN															
Road density (8)	LP	MN		MP		LP	LN														
Dis-nonlocal (9)		LP						LN													
AADT-nonlocal (10)	LP			LP				LN													
Population (205) (11)		LN		LP		LP	LN														
# of Households (12)		LN						LP	LN		HP										
Workers (13)				LP	LP	LP	LN	LN			HP	HP									
Industrial (14)								LP				LP									
High industrial (15)	LP	LN	LP		MP			LP		LP		LP		LP							
Retail (16)		LN	LP		MP			LP	LN			LP	LN	MP	MP						
High retail (17)	LP	LN						LP	LN		LN	LP	LN	LP	LP	HP					
Office (18)		LN						MP		LP		LP	LN	MP	MP	HP	HP				
Service (19)					LP			LP	LN			LP		MP	MP	HP	HP	HP			
Government (20)		LN			LP				LN			LP			LP	LP		LP	LP		
Education (21)					LP	LN									LP	LP	LP	LP		LP	
Employment density (22)		LN			LP			LP	LN			MP	LN	HP	HP	HP	HP	HP	HP	MP	LP

Note: HP, MP, LP, HN, MN, and LN are high positive, moderate positive, low positive, high negative, moderate negative, and low negative, respectively.

Table A21Correlation matrix for population density = 600 - 800 people/square mile

Variable	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)	(17)	(18)	(19)	(20)	(21)
AADT (1)																					
Speed limit (2)	LN																				
# of lanes (3)																					
Functional class type (4)	LP	LN																			
Unpaved (5)																					
Bitumen (6)			LP		LN																
Concrete (7)		LP	LN			HN															
Road density (8)	LP	MN		MP		LP	LN														
Dis-nonlocal (9)		LP						LN													
AADT-nonlocal (10)	LP			LP				LN													
Population (205) (11)		LN		LP		LP	LN														
# of Households (12)		LN						LP	LN		HP										
Workers (13)				LP	LP	LP	LN	LN			HP	HP									
Industrial (14)								LP				LP									
High industrial (15)	LP	LN	LP		MP			LP		LP		LP		LP							
Retail (16)		LN	LP		MP			LP	LN			LP	LN	MP	MP						
High retail (17)	LP	LN						LP	LN		LN	LP	LN	LP	LP	HP					
Office (18)		LN						MP		LP		LP	LN	MP	MP	HP	HP				
Service (19)					LP			LP	LN			LP		MP	MP	HP	HP	HP			
Government (20)		LN			LP				LN			LP			LP	LP		LP	LP		
Education (21)					LP	LN									LP	LP	LP	LP		LP	
Employment density (22)		LN			LP			LP	LN			MP	LN	HP	HP	HP	HP	HP	HP	MP	LP

Note: HP, MP, LP, HN, MN, and LN are high positive, moderate positive, low positive, high negative, moderate negative, and low negative, respectively.

Table A22 Correlation matrix for population density > 800 people/square mile

Variable	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)	(17)	(18)	(19)	(20)	(21)
AADT (1)																					
Speed limit (2)																					
# of lanes (3)																					
Functional class type (4)	LP																				
Unpaved (5)																					
Bitumen (6)																					
Concrete (7)	LP		LP		MN																
Road density (8)						HN															
Dis-nonlocal (9)	LP	MN	LP	LP																	
AADT-nonlocal (10)		LP						LN													
Population (205) (11)	LP			LP				LP	LP												
# of Households (12)	LP	LN						MP		LP											
Workers (13)	LP	LN		LP	LP			MP	LN	LP	HP										
Industrial (14)	LP			LP	LP			LP		LP	HP	HP									
High industrial (15)		LN	LP					LP	LN												
Retail (16)		LN	LP					LP	LN	LP	LP	LP	LP	LP							
High retail (17)		LN						LP			MP	MP	MP	LP	LP						
Office (18)	LP	LN				LN	LP	LP	LN	LP	MP	MP	MP	LP	LP	HP					
Service (19)		LN						MP	LN		MP	MP	LP	LP	MP	MP	HP				
Government (20)	LP	LN						LP	LN	LP	MP	MP	LP	LP	MP	MP	HP	HP			
Education (21)		LN				LN	LP	LP		LP	LP	LP	LP	LP	LP	LP	MP	MP	HP		
Employment density (22)		LN						LP			MP	MP	LP			LP	HP	MP	LP	LP	

Note: HP, MP, LP, HN, MN, and LN are high positive, moderate positive, low positive, high negative, moderate negative, and low negative, respectively.

APPENDIX B: RESULTS FROM COUNTY-LEVEL MODELS

The spatial distribution of local road AADT counts, descriptive statistics of explanatory variables, and Pearson correlation coefficient matrices for selected counties are shown in this Appendix.

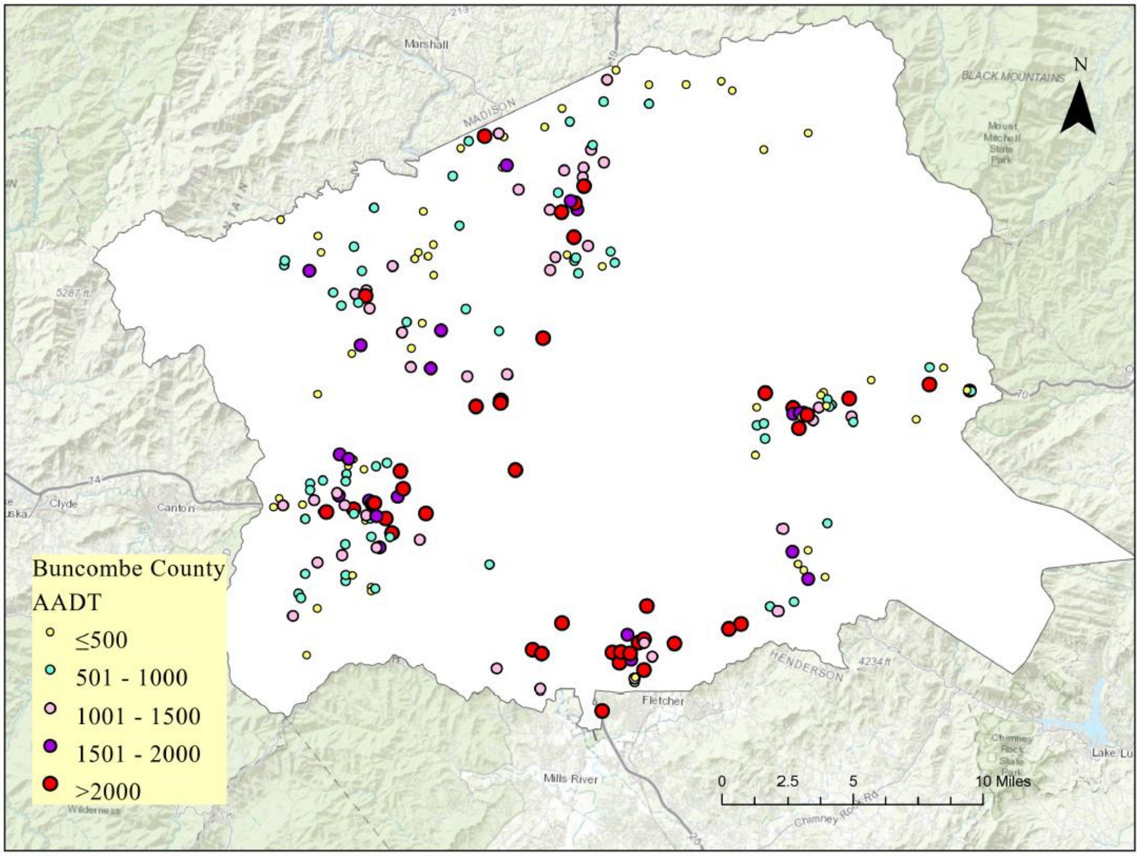


Figure B1 Spatial distribution of local road traffic count stations in Buncombe County

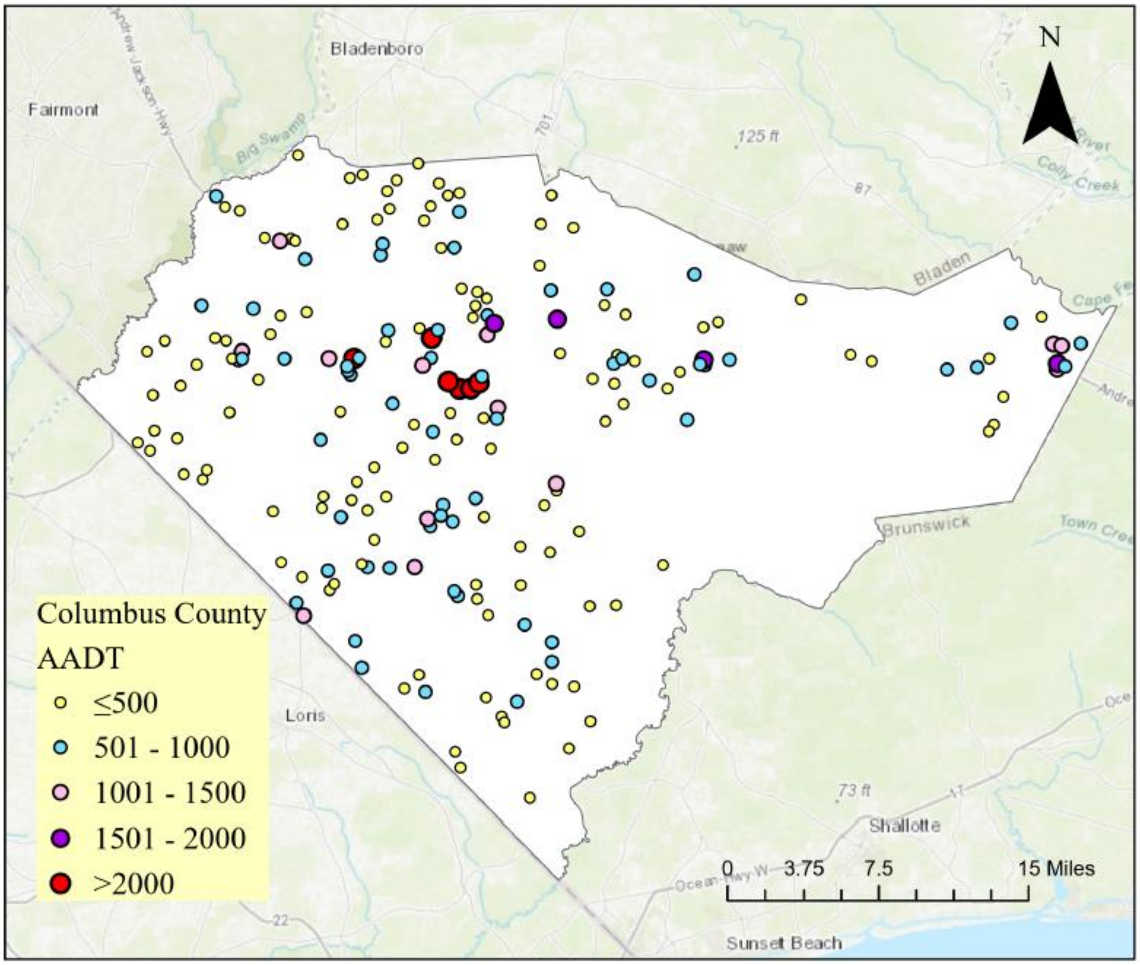


Figure B2 Spatial distribution of local road traffic count stations in Columbus County

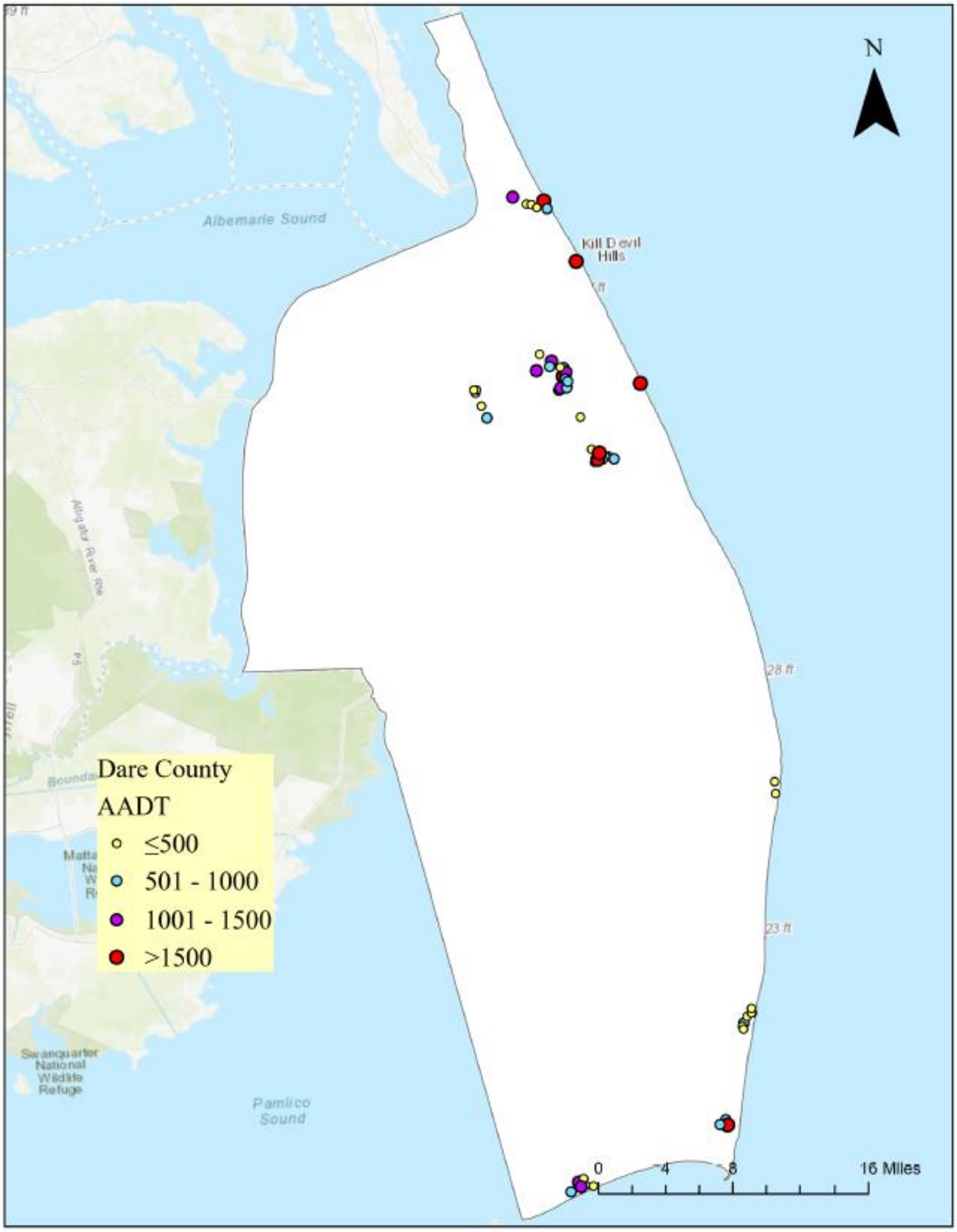


Figure B3 Spatial distribution of local road traffic count stations in Dare County

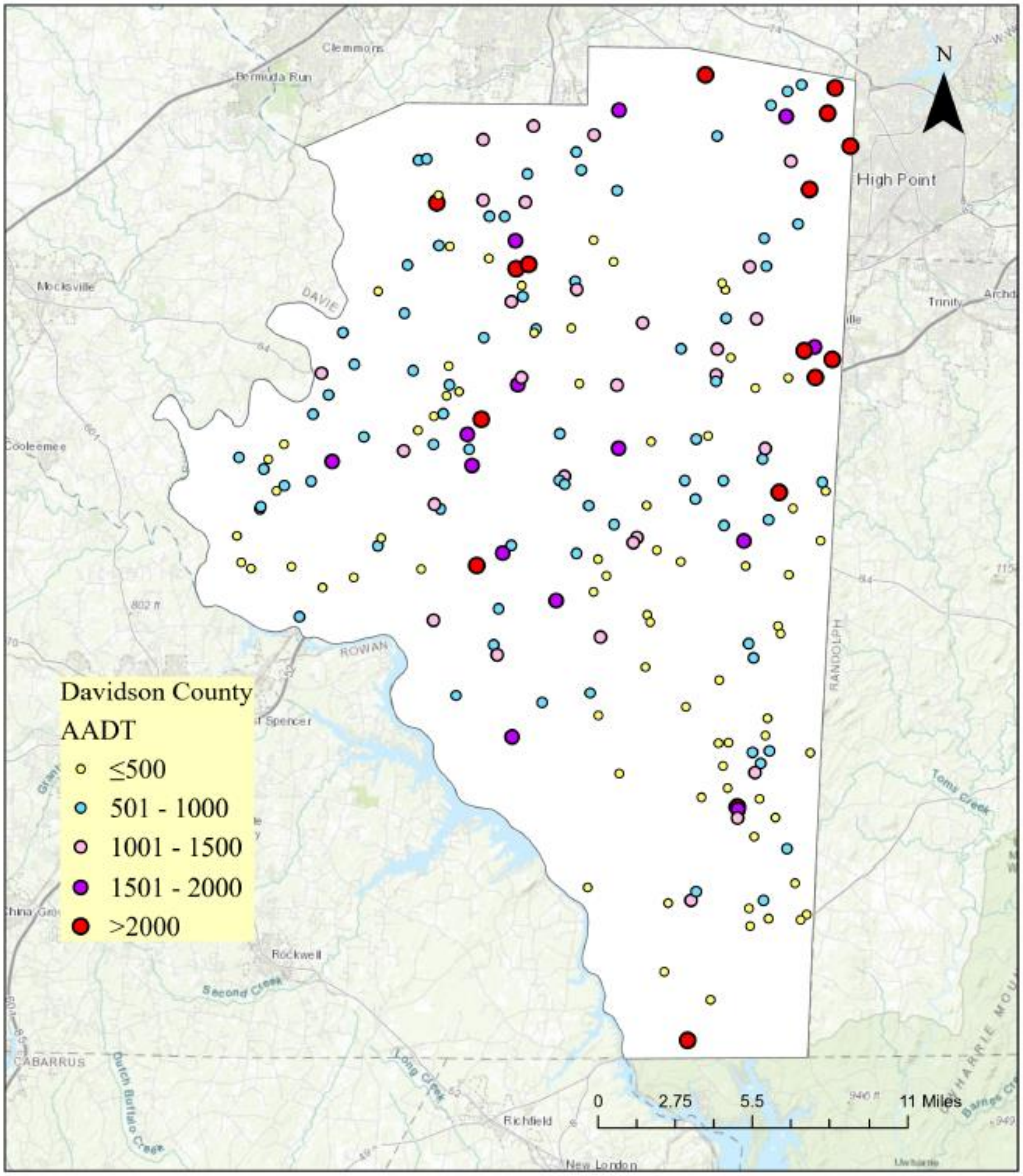


Figure B4 Spatial distribution of local road traffic count stations in Davidson County

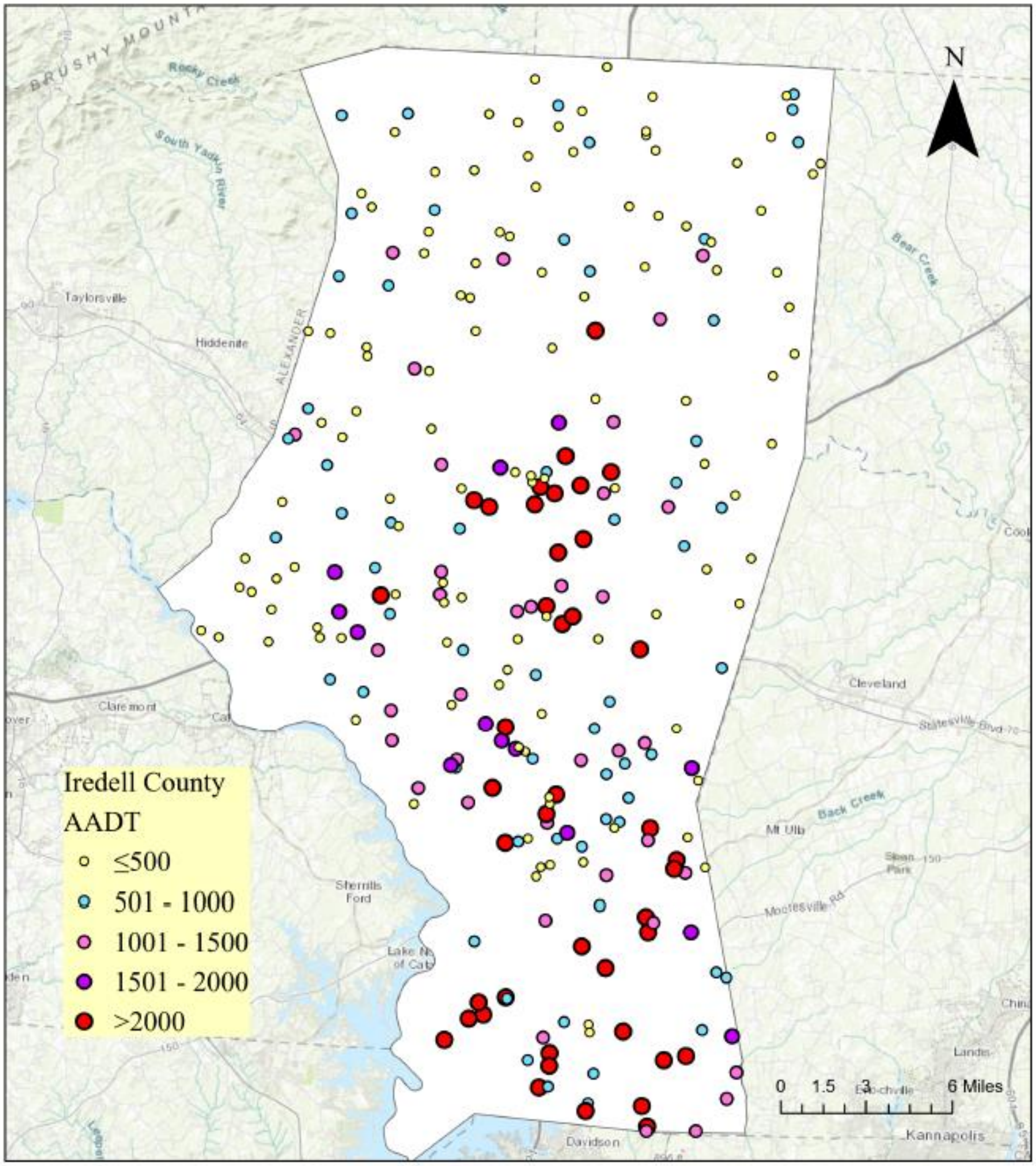


Figure B5 Spatial distribution of local road traffic count stations in Iredell County

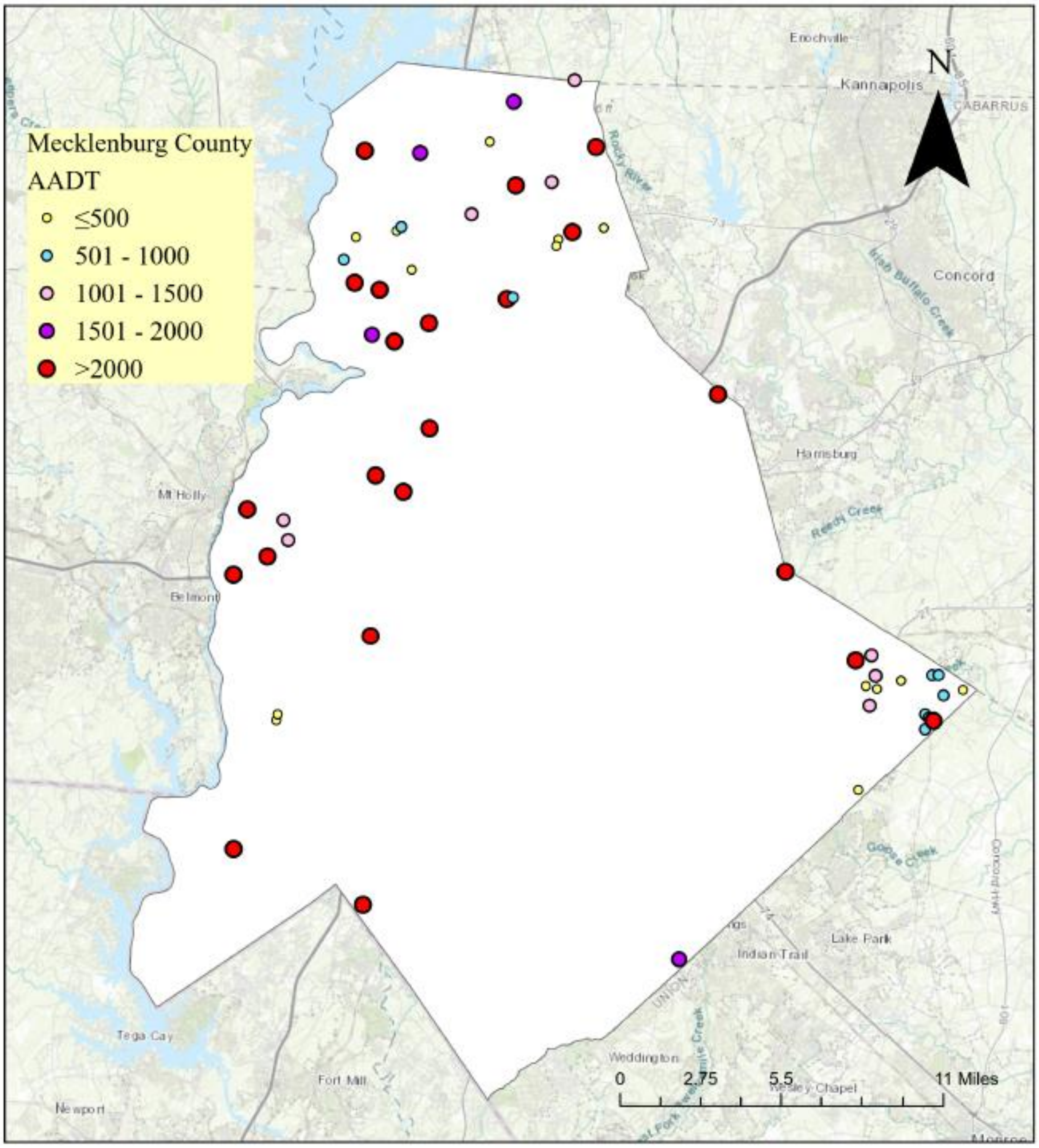


Figure B6 Spatial distribution of local road traffic count stations in Mecklenburg County

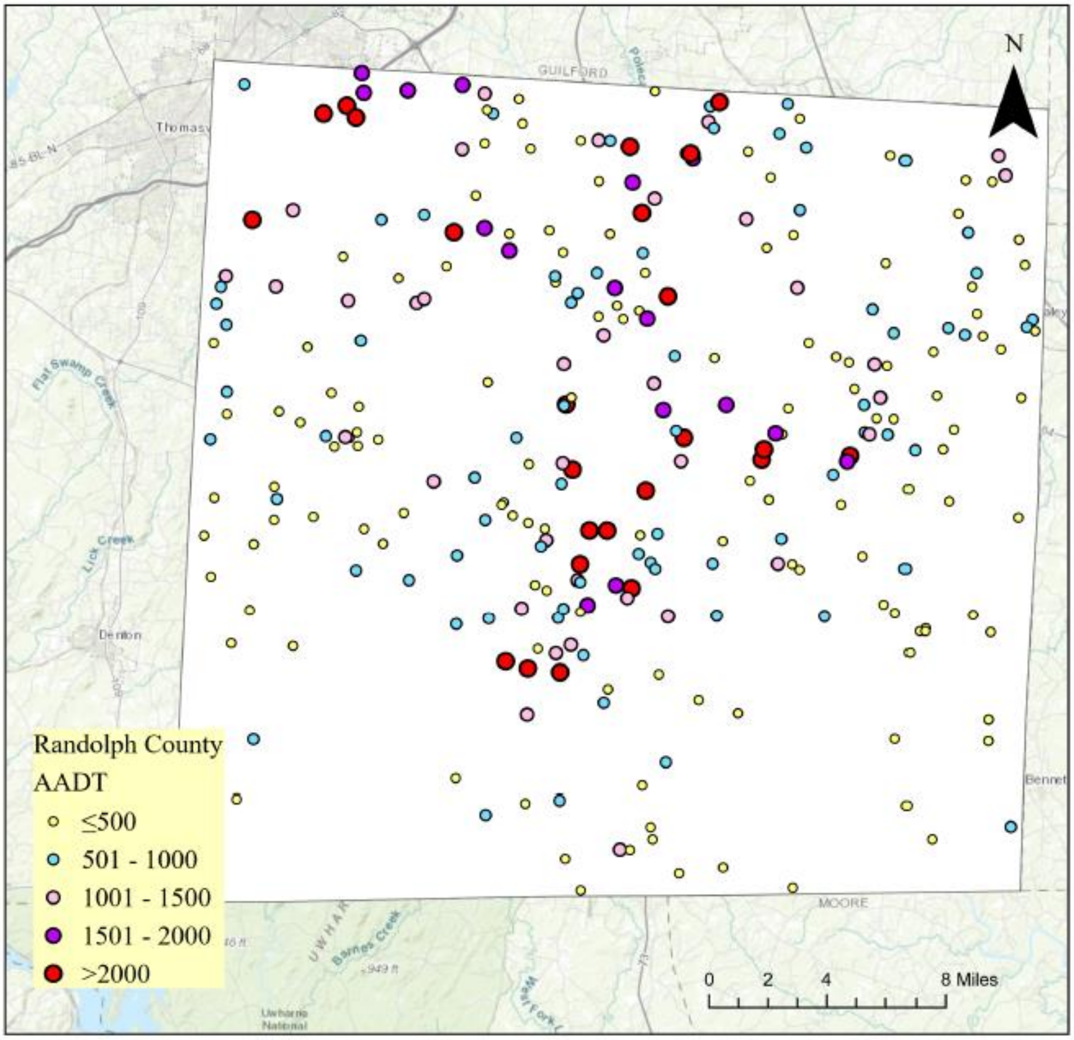


Figure B7 Spatial distribution of local road traffic count stations in Randolph County

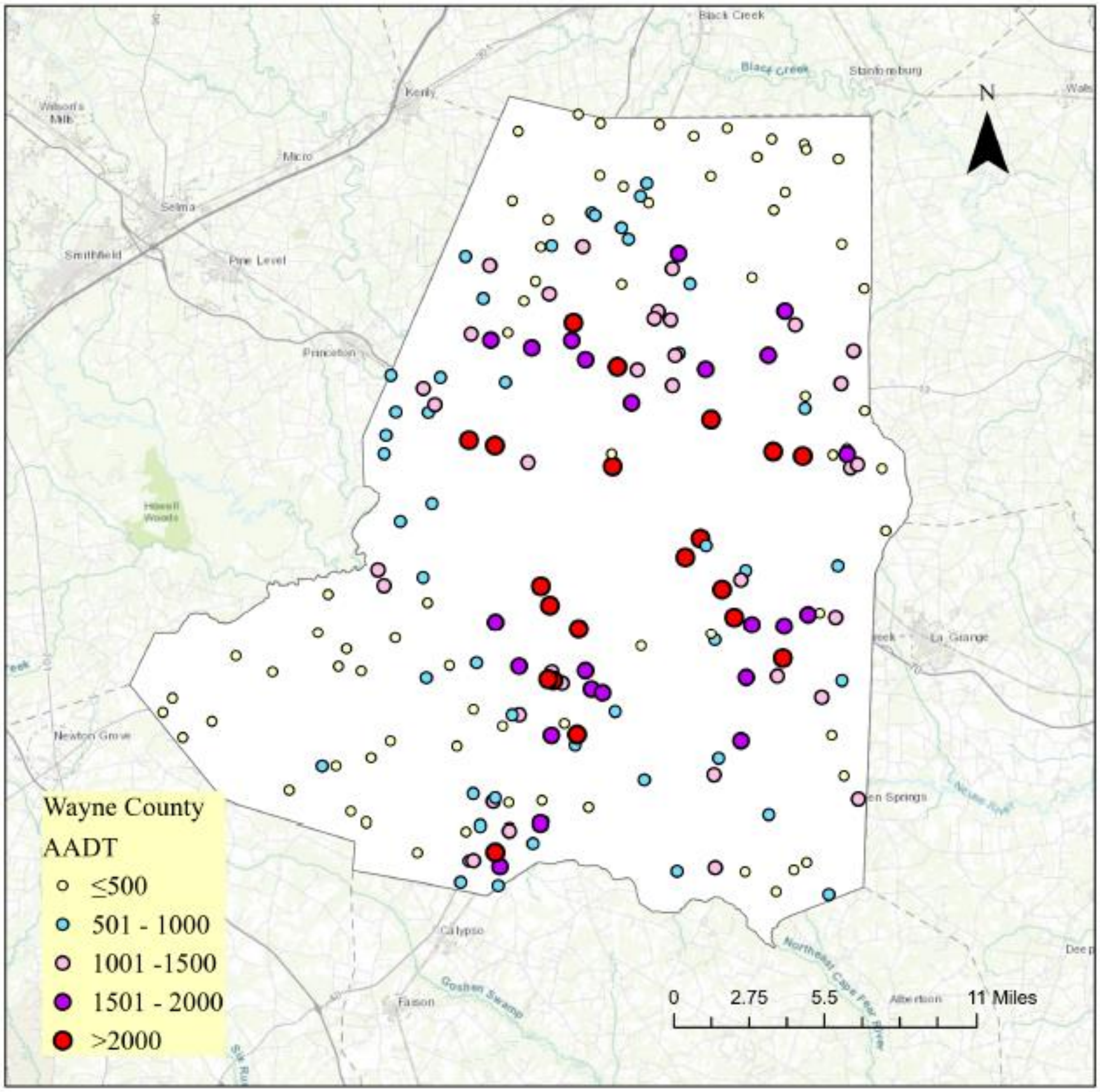


Figure B8 Spatial distribution of local road traffic count stations in Wayne County

Table B1 Descriptive statistics of the explanatory variables – Buncombe County

Variable	Minimum	Median	Mean	Maximum	Std. deviation
Speed limit (mph)	20	35	41	55	9.5
Area type	0	0	0.26	1	-
Road density	2	11.1	13.7	74.1	8.4
Dis-nonlocal (miles)	0.01	0.20	0.54	9.48	0.77
AADT-nonlocal	50	3,300	5,490	151,000	6,837
Socioeconomic variables					
Population	2.34	11.02	15.49	80.17	12.25
# of house holds	1.00	4.47	6.48	34.41	5.30
Workers	1.17	5.06	7.63	36.75	6.11
Industrial	0	0.13	0.81	16.53	2.08
Hi industrial	0	0.42	0.78	16.36	1.57
Retail	0	0.20	0.62	7.79	1.18
Hi Retail	0	0.10	0.56	9.22	1.07
Office	0	0.19	0.88	10.94	1.60
Service	0.06	0.57	1.58	15.1	2.48
Government	0	0.07	0.15	3.10	0.30
Education	0	0.19	0.45	5.00	0.68
Population density	0.81	116.2	213.5	5,798.1	312.21
Employment density	0	28.22	106.4	14,347.1	311.32
Land use					
# of Multi-family units	0	6	10	81	13
# of Single-family units	0	29	44	565	58
Agricultural area	0	0	29.15	996.62	94.50
Government area	0	0	6.27	347.41	32.54
Light commercial area	0	0	101.11	13,823.38	975.10
Heavy commercial area	0	0	0.47	60.90	5.02
Light industrial area	0	0	9.13	626.94	58.34
Heavy industrial area	0	0	8.05	419.13	43.12
Medical area	0	0	0.26	56.30	3.82
Office area	0	0	3.84	286.71	26.71
Recreational area	0	0	23.53	1,727.92	153.73
Resource area	0	0	3.64	156.17	20.84
Retail area	0	0	13.13	504.73	54.23
School area	0	0	6.42	682.47	53.68
Vacant area	0	0	27.30	942.07	100.18

Table B2 Descriptive statistics of the explanatory variables – Columbus County

Variable	Minimum	Median	Mean	Maximum	Std. deviation
Speed limit (mph)	20	55	52	55	7.58
Area type	0	0	0.04	1	-
Road density	3.34	8.24	9.72	38.04	5.43
Dis-nonlocal (miles)	0.02	0.19	0.51	3.46	0.71
AADT-nonlocal	280	2,100	3,787	22,000	4,220
Socioeconomic variables					
Population	0.90	2.13	3.04	9.60	2.09
# of house holds	0.36	0.83	1.20	3.81	0.86
Workers	0.35	0.81	1.13	3.53	0.77
Industrial	0	0.02	0.18	2.36	0.47
Hi industrial	0	0.04	0.07	0.69	0.12
Retail	0	0.03	0.13	1.22	0.27
Hi Retail	0	0.02	0.12	0.90	0.20
Office	0	0.06	0.17	1.84	0.35
Service	0	0.17	0.40	3.47	0.66
Government	0	0.02	0.14	2.00	0.36
Education	0	0.04	0.11	0.47	0.13
Population density	23.72	56.28	80.37	253.34	55.26
Employment density	2.27	11.49	35.68	302.60	58.97
Land use					
# of Multi-family units	0	0	0	45	3
# of Single-family units	0	10	13	75	11
Commercial area	0	0	13.38	317.91	47.67
Government area	0	0	1.79	85.55	10.01
Industrial area	0	0	4.11	343.90	29.01
Institutional area	0	0	5.64	198.98	21.29
Office area	0	0	4.03	228.50	21.79
Retail area	0	0	7.74	326.41	34.22
School area	0	0	5.60	439.67	43.18
Vacant area	0	449.57	428.82	1,055.69	186.23

Table B3 Descriptive statistics of the explanatory variables – Dare County

Variable	Minimum	Median	Mean	Maximum	Std. deviation
Speed limit (mph)	20	35	37	55	11.42
Area type	0	0	0.44	1	-
Road density	6.39	13.7324	15.8262	30.45	6.73
Dis-nonlocal (miles)	0.02	0.08	0.21	1.13	0.24
AADT-nonlocal	1,800	6,100	7,996	36,000	6,312
Socioeconomic variables					
Population	0.03	4.83	3.72	10.55	3.33
# of house holds	0.01	2.03	1.58	4.52	1.42
Workers	0.01	2.44	1.90	5.49	1.73
Industrial	0	0.18	0.11	0.19	0.08
Hi industrial	0	0.38	0.30	0.92	0.29
Retail	0	0.43	0.36	1.04	0.32
Hi Retail	0	0.25	0.30	1.33	0.40
Office	0	0.81	0.69	2.25	0.69
Service	0	0.47	0.41	1.37	0.42
Government	0	0.34	0.29	0.51	0.23
Education	0	0.15	0.10	0.20	0.08
Population density	0.81	127.58	98.33	278.52	88.03
Employment density	0.86	84.43	67.90	201.98	62.46
Land use					
# of Multi-family units	0	0	5	29	8
# of Single-family units	0	51	54	128	28
Commercial area	0	0	54.31	431.91	102.41
Government area	0	0	18.01	205.44	43.31
Institutional area	0	0	31.96	280.68	70.69
Office area	0	0	16.39	219.21	48.96
Resource area	0	0	2.57	133.15	17.70
Retail area	0	0	5.46	208.75	29.59
Transportation area	0	0	2.17	61.35	10.16
Vacant area	0	103.08	131.82	455.46	120.08

Table B4 Descriptive statistics of the explanatory variables – Davidson County

Variable	Minimum	Median	Mean	Maximum	Std. deviation
Speed limit (mph)	20	55	51	55	7.45
Area type	0	0	0.38	1	-
Road density	5.19	16.10	17.47	46.27	7.13
Dis-nonlocal (miles)	0.01	0.12	0.33	2.25	0.44
AADT-nonlocal	540	3,700	5,723	28,000	4,969
Socioeconomic variables					
Population	0.79	9.81	12.62	78.55	11.70
# of house holds	0.33	3.95	5.09	30.79	4.62
Workers	0.39	5.20	6.25	30.75	5.09
Industrial	0	0.22	0.78	17.11	1.97
Hi industrial	0	0.18	0.54	4.48	0.93
Retail	0	0.14	0.75	35.84	2.84
Hi Retail,	0	0.13	0.50	12.48	1.18
Office	0	0.33	1.13	43.77	4.67
Service	0	0.71	1.59	39.89	3.56
Government	0	0.03	0.15	4.67	0.44
Education	0	0.10	0.31	3.63	0.48
Population density	20.79	258.89	333.15	2,073.69	308.93
Employment density	3.29	64.21	153.70	2,552.40	322.84
Land use					
# of Multi-family units	0	0	0	14	2
# of Single-family units	0	22	26	99	20
Commercial area	0	0	17.33	595.54	69.38
Government area	0	0	1.97	279.08	20.52
Industrial area	0	0	13.40	522.64	63.89
Institutional area	0	0	28.69	739.19	90.28
Office area	0	0	1.05	81.81	7.89
Resource area	0	0	0.47	55.30	4.56
Retail area	0	0	8.22	358.12	38.41
School area	0	0	7.89	300.27	32.02
Transportation area	0	0	1.04	86.28	7.86
Vacant area	0	224.32	229.08	722.28	168.05

Table B5 Descriptive statistics of the explanatory variables – Iredell County

Variable	Minimum	Median	Mean	Maximum	Std. deviation
Speed limit (mph)	25	55	49	55	7.57
Area type	0	0	0.31	1	-
Road density	5.68	12.44	14.78	49.22	7.54
Dis-nonlocal (miles)	0.02	0.24	0.63	3.41	0.77
AADT-nonlocal	360	3,900	7,694	83,000	9,788
Socioeconomic variables					
Population	1.98	8.47	11.55	56.40	9.95
# of house holds	0.78	3.18	4.52	24.65	3.95
Workers	0.98	4.19	5.69	28.15	4.71
Industrial	0	0.13	0.91	9.47	1.74
Hi industrial	0	0.18	0.57	20.81	1.85
Retail	0	0.10	0.66	19.21	2.06
Hi Retail	0	0.08	0.53	12.42	1.25
Office	0	0.14	0.80	15.12	1.87
Service	0	0.47	1.42	21.34	2.57
Government	0	0.04	0.21	7.58	0.70
Education	0	0.13	0.32	4.61	0.68
Population density	52.21	223.50	304.91	1,489.03	262.67
Employment density	2.09	43.41	145.18	1,997.36	247.90
Land use					
# of Multi-family units	0	0	2	36	6
# of Single-family units	0	29	33	98	20
Agricultural area	0	0	2.91	270.66	24.07
Commercial area	0	0	27.60	727.10	96.13
Government area	0	0	1.54	151.18	13.21
Industrial area	0	0	19.54	652.64	92.41
Institutional area	0	0	11.60	482.27	47.35
Medical area	0	0	0.30	78.57	4.82
Office area	0	0	1.10	157.26	11.37
Recreational area	0	0	1.99	368.72	24.13
Resource area	0	0	2.11	194.37	16.40
School area	0	0	0.04	8.60	0.55

Table B6 Descriptive statistics of the explanatory variables – Mecklenburg County

Variable	Minimum	Median	Mean	Maximum	Std. deviation
Speed limit (mph)	35	45	44	55	8.00
Area type	0	1	0.79	1	-
Road density	7.57	20.45	20.55	41.97	7.73
Dis-nonlocal (miles)	0.01	0.12	0.41	2.58	0.55
AADT-nonlocal	835	13,000	13,373	34,000	8,718
Socioeconomic variables					
Population	17.10	27.22	35.95	87.30	21.18
# of house holds	5.93	10.30	13.19	34.10	8.01
Workers	7.16	13.31	18.58	53.69	11.68
Industrial	0	0.12	0.58	7.31	1.13
Hi industrial	0.14	1.13	1.53	22.06	2.99
Retail	0.06	0.26	1.00	7.98	1.49
Hi Retail	0	0.29	1.06	11.60	2.08
Office	0	0.61	1.85	21.96	3.18
Service	0.20	1.89	3.19	46.69	6.48
Government	0	0.00	0.44	9.88	1.43
Education	0	1.15	1.21	10.80	1.75
Population density	451.45	718.64	949.02	2,304.72	559.12
Employment density	27.37	141.17	293.22	3,750.10	518.02
Socioeconomic variables					
# of residential units	0	37	37	82	20
Commercial area	0	0	94.87	605.59	152.62
Industrial area	0	0	8.90	277.36	41.00
Large industrial area	0	0	8.92	275.88	42.58
Institutional area	0	0	75.54	1,164.09	237.92
Office area	0	0	4.12	94.54	16.02
Recreational area	0	0	2.98	157.73	21.67
School area	0	0	3.44	182.47	25.06
Utility area	0	0	1.03	52.25	7.17
Vacant area	0	0	20.82	280.53	56.92
Warehouse area	0	0	82.49	2,285.92	339.07

Table B7 Descriptive statistics of the explanatory variables – Randolph County

Variable	Minimum	Median	Mean	Maximum	Std. deviation
Speed limit (mph)	25	55	50	55	7.93
Area type	0	0	0.19	1	-
Road density	3.54	10.72	12.74	45.87	6.99
Dis-nonlocal (miles)	0.03	0.19	0.45	3.45	0.63
AADT-nonlocal	315	2,950	5,254	41,000	7,079
Socioeconomic variables					
Population	0.90	6.28	9.46	93.14	10.89
# of house holds	0.38	2.46	3.76	42.76	4.69
Workers	0.49	3.20	4.76	47.43	5.45
Industrial	0	0.19	1.08	33.71	3.19
Hi industrial	0	0.16	0.40	4.54	0.62
Retail	0	0.08	0.43	13.44	1.24
Hi Retail	0	0.04	0.41	9.43	1.18
Office	0	0.07	0.51	29.20	2.41
Service	0	0.44	1.18	26.80	2.56
Government	0	0	0.46	23.66	2.94
Education	0	0.06	0.34	8.95	1.15
Population density	23.74	165.75	249.73	2,459.01	287.44
Employment density	2.27	37.89	129.60	3,066.87	329.94
Land use					
# of Multi-family units	0	0	0	17	2
# of Single-family units	0	20	24	79	17
Agricultural area	0	0	78.59	701.75	127.47
Commercial area	0	0	1.80	400.55	24.77
Government area	0	0	4.13	339.36	28.58
Industrial area	0	0	15.83	722.30	73.22
Manufacturing area	0	0	1.39	104.81	8.57
Office area	0	0	15.28	538.55	52.10
Recreational area	0	0	0.00	0.04	0.00
Resource area	0	0	0.61	113.63	7.05
Retail area	0	0	9.43	379.19	44.89
Vacant area	0	98.38	118.27	579.48	106.34

Table B8 Descriptive statistics of the explanatory variables – Wayne County

Variable	Minimum	Median	Mean	Maximum	Std. deviation
Speed limit (mph)	25	55	52	55	6.30
Area type	0	0	0.20	1	0.40
Road density	3.21	8.74	11.08	49.15	6.73
Dis-nonlocal (miles)	0.00	0.32	0.59	3.04	0.58
AADT-nonlocal	250	2,900	4,713	24,500	5,059
Socioeconomic variables					
Population	2.13	6.82	8.00	32.15	6.23
# of house holds	0.87	2.61	3.10	12.86	2.45
Workers	1.05	3.12	3.71	12.43	2.72
Industrial	0	0.07	0.49	5.85	1.10
Hi industrial	0	0.06	0.13	1.38	0.25
Retail	0	0.04	0.15	1.99	0.36
Hi Retail	0	0.10	0.25	3.04	0.56
Office	0	0.04	0.26	4.76	0.79
Service	0	0.24	0.62	6.85	1.37
Government	0	0.04	0.13	1.80	0.33
Education	0	0.08	0.35	4.23	0.78
Population density	56.13	180.09	211.21	848.75	164.43
Employment density	1.62	21.71	68.27	870.55	147.53
Land use					
# of Multi-family units	0	0	0	12	1
# of Single-family units	0	3	8	82	13
# of Rural single-family units	0	16	18	125	15
Commercial area	0	0	10.13	464.59	44.89
Government area	0	0	0.78	24.41	3.70
Industrial area	0	0	9.54	358.96	42.65
Institutional area	0	0	6.00	356.56	33.16
Office area	0	0	0.87	61.27	5.36
Resource area	0	0	0.48	52.09	4.31
Retail area	0	0	2.15	158.47	15.68
School area	0	0	2.68	240.30	20.93
Transportation area	0	0	2.46	259.69	20.56
Vacant area	0	177.35	191.02	617.77	147.22

Table B9 Correlation matrix for Buncombe County

Variable	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
AADT (1)	1																			
Speed limit (2)	LN	1																		
Func. class type (3)	LP	LN	1																	
Road density (4)	MP	MN	HP	1																
Dis-nonlocal (miles) (5)	LN	LP		MN	1															
AADT-nonlocal (6)	LP	LN	MP	MP	LN	1														
Population (7)	MP	LN	MP	HP		MP	1													
# of households (8)	MP	LN	LP	HP	LN	MP	HP	1												
Workers (9)	MP	LN	LP	HP	LN	MP	HP	HP	1											
Industrial (10)	MP			LP		MP	LP	LP	LP	1										
Hi-Industrial (11)	MP		LP	MP	LN	LP	HP	HP	MP	HP	1									
Retail (12)	MP		LP	MP		LP	MP	MP	MP	HP	HP	1								
Hi-Retail (13)	MP		LP	MP	LN	LP	MP	MP	MP	HP	HP	HP	1							
Office (14)	MP		LP	MP		MP	HP	HP	HP	HP	HP	HP	HP	1						
Service (15)	MP		LP	MP		MP	HP	HP	HP	HP	HP	HP	HP	HP	1					
Government (16)			LP	LP		LP	LP	LP	LP	LP	LP	HP	HP	LP	LP	1				
Education (17)	MP		LP	LP		LP	HP	HP	HP	HP	MP	HP	HP	HP	HP	MP	1			
Population density (18)	MP	LN	LP	MP		LP	HP	HP	HP	LP	HP	MP	MP	MP	HP	LP	MP	1		
Employment density (19)	MP		LP	MP		MP	MP	MP	MP	HP	HP	HP	HP	HP	HP	MP	HP	HP	1	
# of Multi-family units (20)																LP				1
# of Single-family units (21)		LN	LP	LP		LP	LP	LP	LP									LP		LP
Government area (22)	LP						LP	LP	LP			LP				LP	LP	LP		
Commercial area (23)	LP																			
Industrial area (24)	LP			LP		LP				LP		LP	LP	LP	LP				LP	

Note: HP, MP, LP, HN, MN, and LN are high positive, moderate positive, low positive, high negative, moderate negative, and low negative, respectively.

Table B10 Correlation matrix for Columbus County

Variable	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
AADT (1)	1																				
Speed limit (2)	MN	1																			
Func. class type (3)	MP	MN	1																		
Road density (4)	MP	HN	MP	1																	
Dis-nonlocal (miles) (5)	LN			LN	1																
AADT-nonlocal (6)	LP		LP	MP		1															
Population (7)	MP	LN	MP	MP	LN	LP	1														
# of households (8)	MP	LN	MP	MP	LN	LP	HP	1													
Workers (9)	MP	LN	MP	MP	LN	LP	HP	HP	1												
Industrial (10)	LP			LP	LN	LP	LP	LP	MP	1											
Hi-Industrial (11)	LP		MP	LP	LN		HP	HP	HP	MP	1										
Retail (12)	MP	LN	HP	MP		LP	HP	HP	HP	LP	HP	1									
Hi-Retail (13)	MP	LN	HP	MP	LN	LP	HP	HP	HP	MP	HP	HP	1								
Office (14)	LP	LN	HP	MP			HP	HP	HP	LP	HP	HP	HP	1							
Service (15)	LP	LN	HP	MP			HP	HP	HP	LP	HP	HP	HP	HP	1						
Government (16)	LP	LN	MP	LP			HP	HP	HP	LP	HP	HP	HP	HP	HP	1					
Education (17)	MP	MN	MP	MP	LN		HP	HP	HP	LP	HP	HP	HP	HP	HP	HP	1				
Population density (18)	MP	LN	MP	MP	LN	LP	HP	HP	HP	LP	HP	HP	HP	HP	HP	HP	HP	1			
Employment density (19)	MP	LN	HP	MP	LN	LP	HP	HP	HP	MP	HP	HP	HP	HP	HP	HP	HP	HP	1		
# of Multi-family units (20)		LN																		1	
# of Single-family units (21)		MN		LP	LP															LP	1
Commercial area (22)				LP						LP											
Office area (23)	LP	LN	LP				LP	LP	LP	LP							LP	LP			LP
Retail area (24)	MP	LN	LP	LP			LP	LP	LP	LP							LP	LP			

Note: HP, MP, LP, HN, MN, and LN are high positive, moderate positive, low positive, high negative, moderate negative, and low negative, respectively.

Table B11 Correlation matrix for Dare County

Variable	1	2	3	4	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
AADT (1)																				
Speed limit (2)		1																		
Func. class type (3)		MN	1																	
Road density (4)		LN	HP	1																
AADT-nonlocal (6)			MP	HP	1															
Population (7)		LN	MP	HP	MP	1														
# of households (8)		LN	MP	HP	MP	HP	1													
Workers (9)		LN	MP	HP	MP	HP	HP	1												
Industrial (10)			MP	HP	MP	HP	HP	HP	1											
Hi-Industrial (11)		LN	MP	MP	MP	HP	HP	HP	HP	1										
Retail (12)	LP	LN	MP	MP	MP	HP	HP	HP	HP	HP	1									
Hi-Retail (13)		LN	MP	MP	MP	HP	HP	HP	HP	HP	HP	1								
Office (14)		LN	MP	MP	MP	HP	HP	HP	HP	HP	HP	HP	1							
Service (15)		LN	MP	MP	MP	HP	HP	HP	HP	HP	HP	HP	HP	1						
Government (16)			MP	HP	MP	HP	HP	HP	HP	HP	HP	HP	HP	HP	1					
Education (17)		LN	MP	HP	MP	HP	HP	HP	HP	HP	HP	HP	HP	HP	HP	1				
Population density (18)		LN	MP	HP	MP	HP	HP	HP	HP	HP	HP	HP	HP	HP	HP	HP	1			
Employment density (19)		LN	MP	MP	MP	HP	HP	HP	HP	HP	HP	HP	HP	HP	HP	HP	HP	1		
# of Multi-family units (20)		LN																	1	
# of Single-family units (21)																				1
Commercial area (22)	MP																			LN
Retail area (23)					HP	LP	LP	LP		LP	LP	MP	LP	LP			LP	LP		
Transportation area (24)	MP																			

Note: HP, MP, LP, HN, MN, and LN are high positive, moderate positive, low positive, high negative, moderate negative, and low negative, respectively.

Table B12 Correlation matrix for Davidson County

Variable	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22
AADT (1)	1																					
Speed limit (2)	MN	1																				
Func. class type (3)																						
Road density (4)	MP	MN	1																			
Dis-nonlocal (miles) (5)	MP	HN	HP	1																		
AADT-nonlocal (6)	LN	LP	LN	LN	1																	
Population (7)	LP	LN	MP	MP		1																
# of households (8)	MP	MN	HP	HP	LN	LP	1															
Workers (9)	MP	HN	HP	HP	LN	LP	HP	1														
Industrial (10)	MP	HN	HP	HP	LN	LP	HP	HP	1													
Hi-Industrial (11)	MP	MN	LP	MP		LP	HP	HP	HP	1												
Retail (12)	LP	MN	MP	LP		LP	LP	LP	MP	LP	1											
Hi-Retail (13)	LP	MN	LP	MP			HP	HP	HP	MP	LP	1										
Office (14)	MP	MN	MP	HP	LN	LP	HP	HP	HP	HP	MP	MP	1									
Service (15)	MP	MN	LP	MP		LP	HP	HP	HP	HP		LP	HP	1								
Government (16)	LP	MN	MP	MP	LN	LP	HP	HP	HP	MP	LP	HP	HP	MP	1							
Education (17)		LN		MP			HP	HP	MP	LP		HP	MP	LP	HP	1						
Population density (18)	LP	MN	MP	MP	LN	LP	HP	HP	HP	HP	HP	LP	MP	MP	LP		1					
Employment density (19)	MP	MN	HP	HP	LN	LP	HP	HP	HP	HP	LP	HP	HP	HP	HP	HP	HP	1				
# of Multi-family units (20)		LN	LP	LP		LP														1		
# of Single-family units (21)							LP	LP	LP			LP			LP		LP	LP			1	
Commercial area (22)	MP	MN		MP		LP	LP	LP	LP	MP			LP	MP		LP	LP	LP	MP		LN	1
Office area (23)	LP	LN		LP			LP	LP	LP		LP	LP	LP		LP	LP		LP	LP	LP		LP
Retail area (24)	MP	LN	LP	LP		LP	LP	LP	LP			LP	LP				LP	LP				LP

Note: HP, MP, LP, HN, MN, and LN are high positive, moderate positive, low positive, high negative, moderate negative, and low negative, respectively.

Table B13 Correlation matrix for Iredell County

Variable	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
AADT (1)	1																				
Speed limit (2)	MN	1																			
Func. class type (3)	MP	MN	1																		
Road density (4)	MP	HN	HP	1																	
Dis-nonlocal (miles) (5)	LN	MP	MN	MN	1																
AADT-nonlocal (6)	MP	LN	MP	HP	LN	1															
Population (7)	MP	MN	HP	HP	MN	MP	1														
# of households (8)	MP	MN	HP	HP	MN	MP	HP	1													
Workers (9)	MP	MN	HP	HP	MN	MP	HP	HP	1												
Industrial (10)	MP	LN	LP	LP	LN		LP	LP	LP	1											
Hi-Industrial (11)		LN	LP	LP		MP	MP	MP	MP		1										
Retail (12)	MP	MN	MP	HP	LN	LP	HP	HP	HP	LP	LP	1									
Hi-Retail (13)	MP	MN	MP	HP	LN	MP	HP	HP	HP	LP	LP	HP	1								
Office (14)	MP	MN	MP	HP	LN	MP	HP	HP	HP	LP	LP	HP	HP	1							
Service (15)	MP	MN	MP	HP	LN	LP	HP	HP	HP	MP	LP	HP	HP	HP	1						
Government (16)	MP	LN	LP	MP	LN		HP	HP	HP	MP		HP	MP	HP	HP	1					
Education (17)	MP	LN	MP	MP	LN	LP	HP	HP	HP			HP	MP	HP	HP	HP	1				
Population density (18)			LP			MP	LP	LP	LP		HP			LP	LP			1			
Employment density (19)	MP	LN	LP	LP			MP	MP	MP			LP	MP	LP	LP	LP			1		
# of Multi-family units (20)	MP	MN	MP	HP	LN		LP	LP	LP	LP			LP	LP	LP					1	
# of Single-family units (21)	LP	LN	LP	LP	LN																1
Commercial area (22)	LP	MN	MP	MP	LN	MP	MP	MP	MP	LP		LP	MP	MP	MP	LP	LP		LP	MP	MP
Industrial area (23)	LP	LN	LP	LP	LN	LP	MP	MP	MP	LP		MP	LP	MP	MP	HP	MP			MP	LN

Note: HP, MP, LP, HN, MN, and LN are high positive, moderate positive, low positive, high negative, moderate negative, and low negative, respectively.

Table B14 Correlation matrix for Mecklenburg County

Variable	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
AADT (1)	1																			
Speed limit (2)		1																		
Func. class type (3)			1																	
Road density (4)	MP	MN	MP	1																
Dis-nonlocal (miles) (5)		MN	MN	MN	1															
AADT-nonlocal (6)			HP	MP		1														
Population (7)		MN	MP	HP	MN	LP	1													
# of households (8)		MN	MP	HP	MN	MP	HP	1												
Workers (9)		MN	MP	HP	LN	MP	HP	HP	1											
Industrial (10)	MP			MP						1										
Hi-Industrial (11)				MP						HP	1									
Retail (12)		MN	LP	HP	LN	LP	MP	HP	MP	HP	HP	1								
Hi-Retail (13)		MN		HP	LN	MP	HP	HP	HP	HP	HP	HP	1							
Office (14)		MN		HP	LN	LP	MP	MP	MP	HP	HP	HP	HP	1						
Service (15)		MN		HP			MP	MP	LP	HP	HP	HP	HP	HP	1					
Government (16)				HP						HP	HP	HP	HP	HP	HP	1				
Education (17)		MN		HP	MN		MP	LP		HP	HP	HP	HP	HP	HP	HP	1			
Population density (18)																		1		
Employment density (19)				MP						HP	HP	HP	HP	HP	HP	HP	HP		1	
# of residential units (20)	LP	MN																		1
Commercial area (22)	MP																			
School area (23)	MP																			

Note: HP, MP, LP, HN, MN, and LN are high positive, moderate positive, low positive, high negative, moderate negative, and low negative, respectively.

Table B15 Correlation matrix for Randolph County

Variable	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
AADT (1)	1																				
Speed limit (2)	LN	1																			
Func. class type (3)	MP	LN	1																		
Road density (4)	MP	MN	HP	1																	
Dis-nonlocal (miles) (5)	LN		LN	LN	1																
AADT-nonlocal (6)	LP	LN	LP	MP	LN	1															
Population (7)	MP	MN	HP	HP	LN	MP	1														
# of households (8)	MP	MN	HP	HP	LN	MP	HP	1													
Workers (9)	MP	MN	HP	HP	LN	MP	HP	HP	1												
Industrial (10)	LP	LN	MP	MP	LN	LP	MP	MP	MP	1											
Hi-Industrial (11)	LP		MP	MP	LN	LP	HP	HP	HP	HP	1										
Retail (12)	LP	LN	MP	MP	LN	MP	HP	HP	HP	HP	HP	1									
Hi-Retail (13)	LP	LN	LN	MP	LN	MP	HP	HP	HP	HP	HP	HP	1								
Office (14)	LP	LN	LN	MP		LP	HP	HP	HP	MP	MP	HP	HP	1							
Service (15)	MP	LN	MP	HP	LN	LP	HP	HP	HP	MP	HP	HP	HP	HP	1						
Government (16)	LP	LN	LN	LP		LP	LP	LP	LP	HP	MP	MP	HP	MP	MP	1					
Education (17)	LP		LN	MP			MP	LP	LP	HP	MP	LP	HP	MP	MP	HP	1				
Population density (18)	MP	MN	HP	HP	LN	MP	HP	HP	HP	MP	HP	HP	HP	HP	HP	LP	MP	1			
Employment density (19)	LP	LN	MP	HP	LN	LP	HP	HP	HP	HP	HP	HP	HP	HP	HP	HP	HP	HP	1		
# of Multi-family units (20)	LP			LP		LP														1	
# of Single-family units (21)	LP	MN	LP	LP		LP	LP	LP	LP									LP			1
Agriculture area (22)	LN	MP	LN	MN	LP	LN	LN	LN	LN	LN	LN	LN	LN					LN	LN		MN
Government area (23)	LP	LN	LP																		

Note: HP, MP, LP, HN, MN, and LN are high positive, moderate positive, low positive, high negative, moderate negative, and low negative, respectively.

Table B16 Correlation matrix for Wayne County

Variable	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
AADT (1)	1																				
Speed limit (2)	MN	1																			
Func. class type (3)	HN	HN	1																		
Road density (4)	MN	HN	HP	1																	
Dis-nonlocal (miles) (5)		LP	LN	LN	1																
AADT-nonlocal (6)	LP		LP	LP		1															
Population (7)	MP	MP	MP	HP		LP	1														
# of households (8)	MP	MP	MP	HP		LP	HP	1													
Workers (9)	MP	MP	MP	HP		LP	HP	HP	1												
Industrial (10)	LP	MP	MP	HP	LN		HP	HP	MP	1											
Hi-Industrial (11)	MP	MP	MP	HP		LP	HP	HP	HP	LP	1										
Retail (12)	LP	MP	MP	HP	LN		HP	HP	HP	HP	HP	1									
Hi-Retail (13)	LP	MP	MP	HP	LN		HP	HP	HP	HP	HP	HP	1								
Office (14)	LP	MP	MP	HP	LN		HP	HP	HP	HP	HP	HP	HP	1							
Service (15)	LP	MP	MP	HP	LN		HP	HP	HP	HP	HP	HP	HP	HP	1						
Government (16)		MP	LP	HP	LN		HP	HP	HP	HP	HP	HP	HP	HP	HP	1					
Education (17)	MP	MP	MP	HP			HP	HP	HP	HP	HP	HP	HP	HP	HP	HP	1				
Population density (18)	MP	MP	MP	HP		LP	HP	HP	HP	HP	HP	HP	HP	HP	HP	HP	HP	1			
Employment density (19)	MP	MP	MP	HP	LN		HP	HP	HP	HP	HP	HP	HP	HP	HP	HP	HP	HP	1		
# of Multi-family units (20)	LP		LP	LP			LP	LP	LP	LP		LP	LP	LP	LP	LP	LP	LP	LP	1	
# of Single-family units (21)	LP	HN	MP	MP			LP	LP	LP	MP		LP	MP	LP	MP	LP	MP	LP	LP	LP	1
Industrial area (22)	LP	MP	LP	MP								LP	MP	LP	MP	LP	LP			LP	
Government area (23)	LN		LP					LN													

Note: HP, MP, LP, HN, MN, and LN are high positive, moderate positive, low positive, high negative, moderate negative, and low negative, respectively.