On the Estimation of Volumes of Roadways: An Investigation of Stop-Controlled Minor Legs

Joel Stephen Barnett
Portland State University

Winter 2-19-2015

Let us know how access to this document benefits you.
Follow this and additional works at: https://pdxscholar.library.pdx.edu/open_access_etds
Part of the Civil and Environmental Engineering Commons

Recommended Citation

10.15760/etd.2193

This Thesis is brought to you for free and open access. It has been accepted for inclusion in Dissertations and Theses by an authorized administrator of PDXScholar. For more information, please contact pdxscholar@pdx.edu.
On the Estimation of Volumes of Roadways:

An Investigation of Stop-Controlled Minor Legs

by

Joel Stephen Barnett

A thesis submitted in partial fulfillment of the requirements for the degree of

Master of Science
in
Civil and Environmental Engineering

Thesis Committee:
Christopher Monsere, Chair
Kelly Clifton
Robert Fountain

Portland State University
2015
ABSTRACT

This effort seeks to answer the question; can a transferable model be developed from easily obtainable, publicly available land-use, census, roadway, and network data for the use in safety performance functions? 474 stop-controlled minor legs were used as the training data set using ordinary least squares regression. A best-fit model of maximum independent variables, n=12 was chosen using an exhaustive approach using Mallow's Cp to select the model with least bias in the predictors. The results of the analysis revealed that the combination of variables from Washington, Ohio, and North Carolina did not have a strong relationship. The best-fit model incorporated functional class information of the major-leg, minor leg functional class information, longitudinal markings, access to a parking lot, and population density of census tract. Validation of the model demonstrated an average 59 percent error between the model estimated and actual AADT values for validation data set (n=54). Furthermore, separate models for each state revealed a lack of uniformity in the dependent variables, and more variance description of the state specific AADT.
ACKNOWLEDGEMENTS

Nothing instilled motivation in me more than the people who believe in who and what I can become. Thanks Mom!

I would like to thank my thesis committee, Dr. Chris Monsere, Dr. Kelly Clifton, and Dr. Robert Fountain for their time and guidance regarding this effort.

More specifically, I would like to thank Dr. Monsere for the opportunity to serve under his charge as a research assistant; I appreciate you affording me the opportunity to work on some truly interesting projects, and providing excellent guidance throughout.

To the City of Portland, thank you for the bicycle facilities. Quietly attacking your roads and facilities provided exceptional release for my graduate woes regarding this thesis.
TABLE OF CONTENTS

ABSTRACT ........................................................................................................................................ i
ACKNOWLEDGEMENTS....................................................................................................................... ii
LIST OF TABLES.................................................................................................................................... vii
LIST OF FIGURES.................................................................................................................................. ix
1.0 INTRODUCTION .......................................................................................................................... 1
  1.1 Problem Statement ...................................................................................................................... 1
  1.2 Organization .............................................................................................................................. 1
2.0 LITERATURE REVIEW ................................................................................................................. 3
  2.1 Travel Demand Modeling ........................................................................................................... 3
  2.2 Geospatial Methods .................................................................................................................. 4
  2.3 Regression Methods .................................................................................................................. 6
    2.3.1 Geographically Weighted Regression .................................................................................. 6
    2.3.2 Ordinary Least Squares ...................................................................................................... 7
  2.4 Summary .................................................................................................................................. 12
3.0 METHODOLOGY .......................................................................................................................... 15
  3.1 Study Location .......................................................................................................................... 15
  3.2 Data ......................................................................................................................................... 18
    3.2.1 Average Annual Daily Traffic ............................................................................................ 19
3.2.2 Roadway ................................................................. 20
3.2.3 Functional Class ......................................................... 20
3.2.4 Land Use ................................................................. 21
3.2.5 Socioeconomic ......................................................... 22
3.2.6 Network ................................................................. 22
3.3 Statistical Methods .......................................................... 23
4.0 DESCRIPTIVE STATISTICS ................................................. 26
4.1 AADT ................................................................. 26
   4.1.2 Site Roadway Characteristics .................................... 30
   4.1.3 Nearest State Highway Characteristics .................... 39
4.2 Land Use ................................................................. 49
   4.2.1 Developed or Not .................................................. 50
   4.2.2 Residence plus Open Space ....................................... 51
   4.2.3 Parking Lot Access ............................................... 54
   4.2.4 Within City Limits ................................................ 56
4.3 Network Characteristics .................................................. 57
   4.3.1 Freeways .......................................................... 58
LIST OF TABLES

Table 1: Summary of factors accounted for in previous efforts. ..............................14
Table 2: Mann-Whitney Post Hoc test for Longitudinal Markings. .........................34
Table 3: Mann-Whitney U-test post hoc results for lane configuration.................36
Table 4: Significance testing results for differences in AADT values across states based on presence of HPMS freeways within buffers.................................58
Table 5: Significant testing results from Kruskal-Wallis testing of difference in AADT values between states.................................................................60
Table 6: Significance testing results for differences in AADT values across states based on presence of HPMS minor arterials within buffers..............................62
Table 7: Significance testing results for differences in AADT values across states based on presence of HPMS major collectors within buffers...............................63
Table 8: Descriptive statistics of percent of civilian labor force in census tract by industry. ........................................................................................................69
Table 9: Summary of model performance and significant variables for three state sample........................................................................................................73
Table 10: Model result for minor legs with 3,000 or more annual average daily traffic.

Table 11: Multiple Regression model results for count sites in Ohio.

Table 12: Multiple Regression model results for count sites in Washington.

Table 13: Multiple regression modeling results for sites located in North Carolina.
LIST OF FIGURES

Figure 1: Map of Ohio sample site locations ............................................................. 16
Figure 2: Map of Washington sample site location ............................................... 17
Figure 3: Map of North Carolina sample sites ....................................................... 18
Figure 4: AADT boxplots for measured and log transformed measured AADT 27
Figure 5: Comparison of empirical and fitted CDF (normal and lognormal) ..... 28
Figure 6: Box plots of AADT by state and longitudinal markings ........................ 31
Figure 7: Mean AADT by longitudinal markings and sample group ................... 32
Figure 8: Average AADT grouped by state and turn lane configuration ........... 35
Figure 9: Boxplots grouped by minor functional class ........................................ 37
Figure 10: Average AADT grouped by state and turn lane presence ............... 39
Figure 11: Comparison of site mean AADT by type of divided roadway and state. ................................................................................................................................. 41
Figure 12: Boxplot of minor AADT with the mean AADT for the major functional class group. ................................................................................................................................. 43
Figure 13: Minor mean AADT grouped by whether the minor intersects a principal arterial or major arterial functional class in an urban or rural context then grouped by state. 45

Figure 14: Mean AADT by major leg rural or urban minor arterials and state. 48

Figure 15: Mean AADT by observed developed land use and state. 50

Figure 16: Mean AADT by observed single residence large lot and state. 53

Figure 17: Mean AADT by whether or not parking lot accesses minor road and state. 55

Figure 18: Mean AADT for sites within city or town boundaries by state. 56

Figure 19: Mean AADT for minor legs accessing city or town by state. 66

Figure 20: AADT specific identification of assumed outliers. 75

Figure 21: Diagnostic plot checking linear assumptions for data. 77

Figure 22: Histogram of sample data model residuals demonstrating normality assumption. 79

Figure 23: Comparison of Three-legged Stop Controlled SPF generated crash values using Predicted and Actual AADT Values. 87
Figure 24: Comparison of Four-legged Stop Controlled SPF generated crash values using Predicted and Actual AADT Values.
1.0 INTRODUCTION

1.1 Problem Statement

The influence of traffic volumes on safety is well known, and subsequently accounted for in many crash prediction methodologies. Most notably the Highway Safety Manual safety performance functions which use major and minor leg volume to estimate crashes at stop-controlled intersections (AASHTO, 2010).

However, traffic counting programs are resource intensive, and often exclude stop-controlled minor legs to focus on the higher functional class roadways. Thus, there exists a need to account for these volumes without putting undue constraint on public agencies. This effort attempts to address this problem by developing a transferable model for estimating stop-controlled minor leg volumes for use in safety studies.

1.2 Organization

The subsequent chapter of this report discusses the previous research regarding traffic volume estimation using statistical approaches for safety and
operational usages. Chapter three introduces the data used in the analysis, the processing methods, and the statistical methods used to conduct the analysis. Chapter four presents comparison analysis of AADT values within variables. Chapter five presents the developed model and validation. Chapter six presents the discussion, and chapter seven concludes the body of this effort.
2.0 LITERATURE REVIEW

The following sections provide a comprehensive discussion of the methods used to estimate traffic volumes on roadways not traditionally part of counting programs. Previous research presented in this text discusses neural networks, research using network wide analysis, generalized linear models using transformed dependent variables, negative binomial link functions, and ordinary least squares regression attempts.

2.1 Travel Demand Modeling

Previous research efforts estimated volumes on roadways not included in count programs using four stage modeling processes. Zhong and Hanson (2009) estimated volumes using trip generation, trip attraction, and route choice step of the popular four step travel demand model for the road network in York County and Beresford Census Consolidated Subdivision (CCS), New Brunswick, Canada. The researchers used TransCAD’s built-in four step model finding that traffic assignment to only 65 percent of the network forced overestimated traffic on the roadway network.
Wang (2012) estimated average annual daily traffic for local roads using parcel-level travel demand analysis. Trips derived from the ITE trip generation rates were assigned to tax parcels within Broward County, Florida. A gravity model was used to distribution trips to nearby count locations. Assignment was conducted using an equilibrium assignment.

2.2 Geospatial Methods

Generally, spatial methods approach estimation of average annual daily traffic values by attaching weights that decrease with an increase in Euclidean or network distance between the measured traffic count and the estimation location. Wang and Kockelman (2009) used 28,304 short term program counts collected over eight years to test the ability of Kriging to estimate AADT in Texas. A median error percentage of 33 percent was found when compared to the actual counts. Furthermore, roads of less than 1000 vehicles per day were found to have larger errors while high volume roadways were under estimated. The researchers suggested that a component of the error was due to the lack of controlling for variables like number of lanes, speed limit, and functional class.
In further research using Kriging techniques, Selby and Kockelman (2011) explored the same data set accounting for speed, number of lanes, employment per acre, population per acre, and functional class as covariates. Accounting for the variance attributed to these variables decreased the absolute error within mini-regions, but had limited effect on the high error percentages associated with low-volume counts.

Other research investigated the effects of polygon-based network buffers compared to Euclidean distances (Pulugurtha and Kasam, 2012). AADT values along with intersection and roadway characteristics were gathered from the North Carolina Department of Transportation, and the city of Charlotte, North Carolina for 30 study links located in Charlotte, NC. Bandwidths of one, one and a half, two, three, four, and five mile distances were compared to similar circular buffers defined by Euclidean radii. Weights were assigned based on the different bandwidth using a distance decay function, and generalized linear models were developed using Poisson and negative binomial link functions. A distance decay function was used to assign the weights to the data from different bandwidths.

Three final models were developed based on functional class of the roadway. The number of lanes, downstream link speed limit, downstream cross
street number of lanes, population, and presence of manufactured housing was accounted for in the model for freeways and expressways. Similarly, major thoroughfare volumes were estimated using an urban indicator, speed limit, upstream link speed limit, downstream number of lanes, population, and manufacturing housing. Minor thoroughfare volumes were estimated by urban indicator, upstream link speed limit, and rural district indicator.

Authors concluded that negative binomial regression performed better than Poisson, and modeling by functional class performed better than an overall model.

2.3 Regression Methods

2.3.1 Geographically Weighted Regression

Zhao and Park (2004) developed geographically weighted regression models to estimate annual average daily traffic in Broward County, Florida. The models were derived from 1998 AADT data collected from 775 count stations. Of the 775 count stations used in model development, 164 were on principal arterials, 262 minor arterials, 300 collectors, and 49 local roads. Five independent variables were used; number of lanes, population concentration, employment
concentration, regional accessibility, whether or not the road that had the count station was directly connected to an expressway. Population and employment concentration were varied depending on the functional classification and longitudinal (east, central, west) location of the roadway; buffer size increased with higher functional classification and western direction. Two separate weighting functions were used, bi-square, and Gaussian. The authors concluded that GWR models predict better than OLS regression due to the ability of the models to better explain the variation in data by accounting for spatial variation in the predictor variables.

2.3.2 Ordinary Least Squares

Mohamad et al. (1998) used multiple regression to estimate annual average daily traffic on urban and rural county roads using county population, the number of occupied housing units (number of households), county vehicle registration, county employment, and per capita income at the county level. The transportation system was taken into consideration by accounting for county state highway mileage, arterial mileage, and collector mileage. A measure of accessibility was used as a predictor of AADT. The authors defined accessibility
as a binary determination of whether the system was closest to the state highway system or an urban area. The resultant model was calibrated using backward elimination, centered predictor variables, and log_{10} transformed average annual daily traffic. The final model contained four variables; location, accessibility, county arterial road mileage, and county population accounting for 75 percent of the variation in traffic volumes.

Xia et al. (1999) used ordinary least squares regression to estimate the volumes of 450 non-state urbanized roads in Florida. Initial independent variables were comprised of network and roadway characteristics, accessibility, and socioeconomic variables.

The road network was accounted for by including number of lanes, and a categorical representation of functional classification. City and county collectors, city and county minor arterials, and local roads were categorized by 1, 2, and 0, respectively. Additionally, accessibility was accounted for by binary representation of proximity of count location to state roads, or county roads.

Furthermore, Xia et al. accounted for socioeconomic characteristics using population, number of dwelling units, auto ownership, industrial employment,
commercial employment, service employment, total employment, school enrollment, and hotel occupancy to account for variation in volumes.

Land-use was also accounted for by developing categorical representation of rural, central business district, residential, and outlying business district. Each category was assigned a numerical value one through four, respectively.

The resultant model included number of lanes, functional class, land-use, auto-ownership, proximity to county roads, and service employment. Researchers concluded that varying Euclidean distance had little effect regarding variance prediction when accounting for population and employment.

Seaver et al (2000) developed regression models using traffic volumes collected on 1213 local roads in Georgia. Four road types were used in the model covering non-Atlanta urbanized areas, small urban areas, paved rural areas, and unpaved rural areas. 45 independent variables covering population demographics, education, transportation, income, employment, farming, and urbanization were narrowed using principal component analysis and regression. Stratification and clustering for the rural-paved and rural –unpaved roadways was required due to the poor predicting capability. Thus, road-types within and outside of a metropolitan statistical area had separate models using similar
principal independent variables. By grouping the data in strata new rural models were developed accounting for population change (1990-1996) or density, unemployment, median travel time, percent of farms greater than 500 acres, median leave time for work, and number of persons working outside of the county. For models outside of the MSA an eighth variable was used to indicate the distance from the nearest MSA. The authors concluded that the availability of data for the independent variables is a limiting factor due to the census data only being available once every ten years. The use of clusters based on road and traffic patterns instead of the demographic methodology may yield better results. Another stated issue is transferability due to the construction of the models using only data from 80 counties in Georgia. Of the three models for rural paved roads outside of metropolitan statistical areas the prediction $R^2$ values were 0.74, 0.81, and 0.96 with small sample sizes used to calibrated the models ($n=17$, $n=19$, and $n=22$, respectively).

Zhao and Chung (2001) estimated volumes using ordinary least squares for roads in Broward County, Florida using roadway, socioeconomic, expressway accessibility, and employment accessibility as predictors. The independent variables used were number of lanes, functional class, direct access
to expressways, employment within varying Euclidean distances, and accessibility defined by;

\[ REAccess_k = \sum_{j=1}^{N_E} E_j e^{-0.0954t_{kj}} \]  \hspace{1cm} (1)

- \( E_j \): Employment
- \( t_{kj} \): travel time at the \( k \)th count station
- \( N_E \): number of employment centers

Four models were generated with the best model performance of \( R^2 = 0.818 \).

Dixon et al. (2012) estimated minor annual average daily traffic for rural roadways using county and nearest city populations, regional average per capita income, distance to the freeway, functional classification, within a city boundary, presence of a minor and/or major right turn lane, land-use, centerline, and striped edge lines.

The researchers used a \( \log_{10} \) transformation to the minor average annual daily traffic volumes as well as the other continuous variables included in the model. The method of analysis tested all possible subsets of models to obtain the model with that best explained the variance in the AADT on the minor
roadways. $C_p$ statistic was calculated using the best model derived from each possible number of independent variables (i.e IV=1, IV=2).

The resulting analysis generated two models, one model for rural roadways, sans MR4SG ($R^2=0.62$), and another for use when estimating minor volumes at MR4SG intersections ($R^2=0.64$). The reason for two models was the larger effect of the right-turn lane on accounting for variance in volumes along the minor approach for MR4SG intersections than all other types of intersections. Also, functional classification of the cross street as a minor arterial or major collector had a smaller positive effect on AADT for MR4SG when compared to the combination of all other models.

2.4 Summary

The research reviewed for this effort did not reflect an attempt of generating linear models using data from each major region in the United States. Also, complex methodologies have the capacity to generate more accurate results than regression methods but are limited in application for practitioners due to the complex, time-consuming modeling processes. The methodology of
preference is ordinary least squares regression, and was chosen for this research effort.

Based on the previous work done on this subject several categories of variables were identified for use in this effort (Table 1). The most prominent demographic variable included in previous models was population followed by employment and income. The aggregation levels varied due to available data including levels at the census tract and county levels. The most readily available socio-demographic information is at aggregated at the census tract level by the United States Census Bureau.

Functional class was the most prominent roadway characteristic used in previous research. Speed limit, number of lanes, and longitudinal markings were other characteristics accounted for including the accessibility, or the distance to the nearest connection of the site to higher functional classes.

Researchers accounted for land-use using urban and rural location. In some cases, models included accessibility to central business districts and employment centers. The research effort presented here uses a combination of immediate land-use near the intersections including whether the intersection is within an urban or rural environment.
### Table 1: Summary of factors accounted for in previous efforts.

<table>
<thead>
<tr>
<th>Roadway</th>
<th>Network</th>
<th>Socio-demographic</th>
<th>Land Use</th>
<th>Other</th>
</tr>
</thead>
<tbody>
<tr>
<td>Distance to nearest freeway, interstate miles, state highway miles, arterial mileage</td>
<td>County-Population, household, vehicle registrations, employment, per capita income</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of lanes, functional class</td>
<td>Access to state roads, county roads</td>
<td>Population, dwelling units, auto ownership, employment</td>
<td>Proximity to central business district, rural area</td>
<td>Farms, travel time to work, leave time to work</td>
</tr>
<tr>
<td>Number of lanes, functional class</td>
<td>Access to nearest expressway</td>
<td>Employment, housing, population</td>
<td>Proximity to central business district, rural area</td>
<td>Proximity to employment centers</td>
</tr>
<tr>
<td>Speed limit, number of lanes, functional class</td>
<td></td>
<td>Population, employment</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Functional class, centerline, edgeline, turn lane presence</td>
<td>Distance to nearest freeway</td>
<td></td>
<td>Adjacent land use is developed</td>
<td></td>
</tr>
</tbody>
</table>
3.0 METHODOLOGY

The following text discusses the data, data processing, and statistical methods used herein. Chapter four includes the descriptive statistics and non-parametric analysis of the AADT values by variable.

In summary, five ordinary least squares regression models were developed using land-use, socioeconomic, network characteristics, and roadway characteristics. Model selection included an exhaustive methodology with models of varying number of variables chosen by comparison of Mallow’s Cp.

3.1 Study Location

The study locations are composed of 532 sites near stop-controlled intersections along state, county, and local roads in 59 counties throughout North Carolina, Ohio, and Washington.
Figure 1: Map of Ohio sample site locations
Site selection within each state was based on equal distribution between urban and rural locations, lane configuration of the major, and median type of the major roadway with emphasis in collecting data in corridors to facilitate faster data collection. This is demonstrated in the maps of Washington, Ohio, and North Carolina in Figures 1, 2, and 3.
3.2 Data

Random selection in the statistical program R selected 478 of 532 sites, or 89.85 percent of the available sites reserving the remaining ten percent for model validation. Of the 478 sites, four records did not have minor AADT associated with the record leaving 474 records for analysis in the training set.
The following sections reflect development of roadway characteristics including AADT for both the minor and major roads, socio-demographic characteristics at the census tract level, functional classification of major and minor roadways, land use, and network characteristics. A key and description of the 39 variable headers are located in Appendix A.

3.2.1 Average Annual Daily Traffic

Collection of average annual daily traffic involved public record requests, internet searches, and data collection efforts by the research team tasked with National Cooperative Highway Research Project 17-59. Local public agencies, and state transportation agencies assisted the researchers providing counts collected in previous years from local counting programs. However, not all selected roadways were part of counting programs. The data collection teams used manual counting methods, and pneumatic counters to supplement the counts by collecting counts ranging in duration from peak hour counts to twenty-four hour traffic volumes. The traffic counts were then factored to AADT using the factoring methods outlined in FHWA Traffic Monitoring Guide.
3.2.2 Roadway

To assist in identifying characteristics that differed due to traffic demand at each intersection, several roadway characteristics were included in the analysis to attempt to define differences in lower volume roadways, and higher functional class roadways; presence of minor leg right or left turn lanes, longitudinal markings, major leg median barrier, and major leg two-way left turn lane.

Visual inspection determined the presence of longitudinal markings using the street view available in Google Earth. Field data collection efforts captured the presence of a median and presence of a minor or major turn lane.

3.2.3 Functional Class

Many of the minor roadways did not have a functional classification defined by the respective roadway managing departments, and were aggregated as local roadways. Collectors and arterials defined the other two categories of ‘collectors’ and ‘arterials’, thus grouping all collectors and all arterials into separate groups. Since the focus of the data collection effort was at stop-controlled intersections, lower volume roadways were the majority case. Minor
leg arterials accounted for 14 legs in the analysis, but ranged in functional classification suggesting the higher level aggregation. More is discussed in Chapter 4 regarding the counts, and AADT associated with these categorizations.

Dummy variables defined the seven different functional classes of the nearest state highways (major leg). The categories defined were urban principal arterials, rural principal arterials, rural major arterials, urban minor arterials, rural minor arterials, rural major collectors, and urban local roads.

3.2.4 Land Use

To account for land-use, visual inspection using Google Earth determined the presence of parking lot entrance or exit, presence of developed land, and the presence of a single family residence on a large lot. The latter category, single family residence on a large lot, was determined by a building with an access road to the minor leg located near the intersection on a large plot of land that was not noticeably developed or within an urban area. The actual residential status of each parcel defined as ‘single family farm’ was unknown regarding this research effort. Thus, the classification of the open space was used only to describe the variable collected to account for potential variations in traffic.
Land use can vary significantly between urban and rural areas. Sites included in the full sample were selected so that urban and rural sites were equally represented. Suburban environments were not taken into consideration due to the likely introduction of error between the data collection teams charged with classifying such sites. Thus, classification of the roadways were made based on the functional classification of the roadway as urban or rural, or in some cases the nearest functionally classified roadway.

3.2.5 Socioeconomic

Socioeconomic variables were compiled by relating each site to the surrounding census tract. Population, acres of each tract, vehicles available, population of at least 16 years of age, urban population, rural population, and industry employment were gathered using the American Community Survey 5-year Estimates for 2008-2012.

3.2.6 Network

The localized network was accounted for using shapefiles collected from the Highway Performance Monitoring System (HPMS) to potentially account for density changes in the network. Each state’s shapefile is composed of line
features with attributes describing functional class, segment length, volume, and other specific features.

Of importance to this effort were road identifier, functional class, and segment length. Unique road identifiers were used to capture the number and total miles for each functional class. The records obtained from the HPMS have info for five functional classes; freeways, principal arterials, other principal arterials, minor arterials, and major collectors. ArcGIS 10.1 was used to spatially relate the number of unique functional class roadways to each minor leg as well as the total miles of each functional class within an Euclidean distance of one, two and a half, five, and ten miles.

3.3 Statistical Methods

An exhaustive approach was conducted on the variables hypothesized to increase the explanation of the variance in the dependent variables. The regsubsets() function within the R package, leaps employing a branch and bound methodology calculated the best model for each possible subset of the variables up to the user defined maximum of independent variables (Lumley, 2014).
Conceptually, the algorithm calculates the residual sum of squares (RSS) and maintains the model with the least RSS until another better RSS value is generated within the described parameters for all possible models. In the case of this research a model of each possible subsets up to n=12 was generated. Or, if \( k \in (1,2,3,\ldots,11) \) and is the number of variables to be included in a model then a model was generated for \( k+1 \) resulting in eleven models. For each \( k+1 \) the branch and bound algorithm captures the best model and provides an output of the independent variables.

Mallow’s \( C_p \) values were used to select the model with the most appropriate amount of variables in an effort to limit the influence of bias and overfitting of the model. The principle behind Mallow’s \( C_p \) is that if the MSE for the full model is an accurate estimate of the population variance then the \( C_p \) value can be considered a measure of bias in the predictors with the least bias model being the model with the number of predictors nearest \( C_p \) (Gilmour, 1996). Particular to this effort and Dixon et al. (2010), Mallow’s method of model selection allowed for the comparison of models of differing number of predictors to account for the inflated difference in \( r^2 \) based on the inclusion of more variables in the regression equation.
Application in this effort involved calculation of the $C_p$ using the `regsubsets()` function accounting for the sum of the squares and the number of predictors. A plot of the included variables against $C_p$ was used to capture which variables were included in the best model (Appendix ?).

Models nearest and below $C_p=p$ were selected and reran using the R linear multiple regression function `lm()` to test for coefficient and model significance. If coefficient models did not meet the predetermined significance level of $\alpha=0.05$ then the variable was deleted using a backwards deletion approach.
4.0 DESCRIPTIVE STATISTICS

The following chapter presents the descriptive statistics of the variables used in the analysis. The 474 count sites are described by providing an overview of the sample, and by characterizing the differences between the states in terms of variations of minor road volumes and other key factors discovered in the analysis.

4.1 AADT

A key aspect of the AADT used as the dependent variable was the intersection with the state highway. The sample AADT included minor legs with volumes ranging from 80 to 15,635 AADT. The average AADT was 1,316 with a standard deviation of 1,568.

Figure 1 supports a right hand skew demonstrated by the outlier values up to the maximum of 15,635 annual average vehicles per day. Furthermore, a skew value of 3.89 was associated with the sampled traffic volumes.
Previous literature suggested a lognormal distribution of traffic volumes for positive skewed data sets (Mohammed et al, 1998; Dixon et al, 2011). The volumes were transformed using a natural logarithm (log\(_e\)) and distribution characteristics were analyzed.

Box plots of the transformed data were included in Figure 1 as a side-by-side comparison of the untransformed and transformed AADT values. The skewness value of the log-transformed data was 0.07 supporting the assumption of lognormal distribution, further supported by a kurtosis value of 2.89.
Comparison of empirical cumulative distribution function and fitted normal and lognormal cumulative distribution functions (Figure 2) further confirmed the assumption of lognormal distributed data for the entire sample as well as the state specific samples demonstrated in Appendix B.

Figure 5: Comparison of empirical and fitted CDF (normal and lognormal).
4.1.1.1 State AADT

The Ohio mean AADT was 1,554 vehicles per day, the highest of the states, with a minimum AADT of 80 and a maximum of 9,000 vehicles per day. Ohio also had the highest variation within the sample of AADT values with a standard deviation of 1,828 for 140 sites, the smallest number of sites of the three states.

Similar variation was found within the 174 Washington sites included in the analysis. The mean AADT was 1,356 vehicles per day with a standard deviation of 1,787. The large variation in the Washington data was attributed to the large volumes experienced at two locations, both greater than 9000 AADT as well as other locations that contributed to a skewed right data set.

North Carolina average minor annual daily traffic was 1,051 with a standard deviation of 858 for the 160 sites included in the analysis. The distribution was less skewed to the right compared to the other three states (Figure 1) as the maximum AADT was 5,000 vehicles per day for 160 sites with a minor value of 90 vehicles per day.
A Kruskal-Wallis test between the ranked AADT of the sites between the three states suggested that no significance difference existed between the median ratings for the state average AADT values (K=0.451, p=.798, ns).

4.1.2 Site Roadway Characteristics

4.1.2.1 Longitudinal Markings

The percentage of sites with edge line striping was almost 51 percent of the sample. Sixty-three percent of the sample sites had striped centerlines near the intersections. In Washington, 45 percent of the sites had edge lines and 68 percent of the sites had centerlines. Ohio was distributed similar to the entire sample. 51 percent of the Ohio roadways had edge lines and 65 percent of the roadways had centerlines. North Carolina sample sites had edge lines on 57 percent of the sites, and 56 percent had centerlines.
Three sites in North Carolina had edge line stripes, but no center line striping. Ohio had two sites of similar configuration, while only one site in Washington was found with the similar conditions. Sixty-six sample sites had centerline longitudinal markings only with a single site located in North Carolina, and an average of 1,216 annual average vehicles per day. Additionally, 234 sites across all three states had centerline and edge line striping with an average AADT of 1,592 vehicles per day. One hundred sixty eight sites were
found to have no striping near the intersection. The annual average daily traffic for these sites was 944 vehicles per day.

Comparatively, a decrease in the provided striping also reflected a decrease in the overall volume, an intuitive result as more traffic control features are likely to be implemented due to more traffic demand. Figure 3 presents the box plots for all four groups and demonstrates the higher AADT values associated with striping configurations of both longitudinal markings compared to the no longitudinal marking methodology.

![Average AADT by Longitudinal Markings](image)

Figure 7: Mean AADT by longitudinal markings and sample group
Figure 4 presents a comparison of the means by the four groupings of longitudinal markings compared to the sample variation between the states. A lack of longitudinal markings consistently resulted in a lower AADT when compared to the other three categorizations.

Washington had a higher AADT associated with roadways having both longitudinal markings. Ohio had a higher AADT associated with edge only group, but only six sites for the entire sample fell within that group. The high AADT for North Carolina was attributed to the influence of only one site having only a centerline.

Testing revealed that the ranked values for the four groups were not all the same (K=48.2965, p<.001). Post hoc testing using Mann-Whitney U-test revealed a significant difference in median ranked values between the sites with both longitudinal markings and those sites with only centerline markings (W=9573, p=.003). Table 2 presents the post hoc results.

Also found to have a difference were sites with no markings and sites with edge line only markings (W=789, p=.019). The most significant difference in values were the sites with both markings and those sites with no markings (W=27463.5, p<.001).
Table 2: Mann-Whitney Post Hoc test for Longitudinal Markings

<table>
<thead>
<tr>
<th></th>
<th>Both</th>
<th>Center</th>
<th>Edge</th>
<th>None</th>
</tr>
</thead>
<tbody>
<tr>
<td>Both</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Center</td>
<td>p=.003</td>
<td>-</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Edge</td>
<td>ns</td>
<td>ns</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>None</td>
<td>p&lt;.001</td>
<td>ns</td>
<td>p=.019</td>
<td>-</td>
</tr>
</tbody>
</table>

4.1.2.2 Lane Configuration

Five percent of the 474 sample sites had right turn lanes, and four percent had left turn lanes. State by state analysis found that Washington had eight percent of intersections with state highways having a right turn, and three percent having a left turn lane. Six percent of the North Carolina sites had right turn lanes, and an almost seven percent with left turn lanes. Ohio had neither left or nor right turn lanes associated with any of the sites included in the sample analysis.
Figure 8: Average AADT grouped by state and turn lane configuration.

The presence of a right turn lane, left turn lane, or both would suggest an increase in average daily traffic due to the demand for conflicting movements at the intersection. Figure 5 presents the differences in mean AADT between the sites grouped by lane configuration at the intersection with the nearest state highway. A Kruskal-Wallis test confirmed that a significant difference between the values of all four groups were not all equivalent ($K=27.7894$, $df=3$, $p<.001$).
Table 3: Mann-Whitney U-test post hoc results for lane configuration

Table 3 presents the post hoc results exposing the differences in the mean AADT between the groups. The presence of left turn was significantly different when compared to right turn lanes and both turn lanes, but was not found to be significantly different than sites with no turn lanes present. AADT values for sites with no turn lanes were significantly different than the AADT values for sites with both turn lanes as well as sites with right turn lanes.

4.1.2.3 Functional Class- Minor Leg

Sites on local streets or streets unclassified were the predominant functional class within the sample accounting for 80 percent of the sample with collectors accounting for 16 percent of the sample and arterials the remaining four percent.
Classification of a roadway as a local street was expected to reflect a lower daily traffic than higher order functional classes due to methodology of how classes are assigned. Figure 6 presents the summary statistics in the form of box plots for the three groups of minor leg functional classification, and confirms that as functional class increases so does the median AADT values. However, this not the case for actual values as both local streets and collectors had volumes greater than 5,000 vehicles per day.

The three groups were tested to determine if all ranked values were equivalent. The Kruskal-Wallis test found a significant difference in at least one
of the pair comparisons (K=38.236, df=2, p<.001). Post hoc analysis using the Mann-Whitney U-test revealed that the ranked AADT values for local classified sites were significantly different from collectors (W=10184.5, p<.001), as well as significantly different from arterials (W=1123, p<.001). Also, collectors AADT values were significantly different from arterials (W=366.5, p=.001).

The range of mean AADTs associated with local functional classification varied by state. Washington had the lowest range in values, but still had a maximum value of 3,376 AADT for local streets. Ohio had a maximum of 9,000 AADT included in the data set while North Carolina had a maximum AADT of 4,900, both classified as local streets. This speaks to the possible misclassification of some roadways as local streets when a higher classification is more appropriate.
Figure 10: Average AADT grouped by state and turn lane presence

Only one site in North Carolina was classified as an arterial explaining the decrease in the average AADT within the North Carolina sample for that functional class (Figure 7). This was also the case for Ohio. The remaining 17 sites classified as arterials were located in Washington.

4.1.3 Nearest State Highway Characteristics

The traffic volumes in this analysis came from a larger data set that focused on capturing characteristics of intersections along state highways. Thus,
all count stations are assumed to be within one quarter mile of a stop controlled intersection along a state highway. To account for the potential influences of close proximity to the National Highway System available major leg characteristics of AADT, median, and functional class were accounted for in the modeling process.

4.1.3.1 Divided

A physical division of the opposing travel lanes was documented at 32 percent of the nearest state highways to the count stations. Of the sites in Washington, 37 percent of the nearest state highways were divided. Ohio had 31 percent of the nearest state highways to the count stations having a physical barrier. 28 percent of state highways nearest the respective count station were divided.
Comparing the average AADT of minor legs intersecting divided and undivided roadways suggested that no significant difference existed between the medians ($W=24484$, $p=.911$, ns). The range of AADT values for both configurations suggests that no definite delineation in AADT value exists with the exception of Ohio which had a lower mean AADT associated with the major legs being divided highways compared to undivided highways.
4.1.3.2 Two-Way Left Turn Lane

Twenty seven percent of the sample sites were located on minor legs intersecting state highways with two-way left turn lanes. Figure 8 includes the mean AADT values compared against the other median types. Similar characteristics were portrayed by the Washington and North Carolina samples when compared to the overall sample. The outlying case was the Ohio sample with the higher mean AADT for the 31 TWLTL associated sites.

4.1.3.3 Functional Class

The major roadway functional class was captured for the state highway nearest the count station. Principal arterials accounted for 46 percent of the sample with an even distribution between urban and rural environments. Urban minor arterials accounted for 30 percent of the major roadway functional class while rural major arterials, minor arterials, and major collectors were found to reflect three, nine, and seven percent of the sites, respectively.
Figure 12: Boxplot of minor AADT with the mean AADT for the major functional class group.

Box plots in Figure 9 demonstrate the distribution of the site AADT values by the functional class of the nearest state highway. The following discussion further outlines the descriptive statistics for each variable.

4.1.3.3.1 Principal Arterials

The average minor leg roadway volumes were 1,443 vehicles per day at intersections with principal arterials (n=220) compared to the 1,206 vehicles per day of legs intersecting non-principal arterials (n=254). The 25th percentiles were similar in value, 467.5 vehicles per day for principal arterials and 450 vehicles per
day for non-principal arterials. Similarly, the 75th percentiles were close in value, separated by only 19.5 vehicles per day, with legs intersecting non-principal arterials having the higher third quartile value of 1,605 vehicles per day.

Comparing legs intersecting rural principal arterials (n=111) and those intersecting urban principal arterials (n=109) discovered a higher mean AADT for those legs intersecting an urban principal arterial. The summary statistics for the two groups revealed that the legs intersecting urban principal arterials were all higher in value. The first quartile difference was 155 vehicles per day, and the third quartile difference was 266 vehicles per day while comparison of the means revealed close to an 800 vehicles per day difference.

Comparison between the states revealed that the largest variations between the functional classes occurred within the Ohio sample. Most noticeably Ohio did not have any rural major arterials included in the sample and had a mean AADT of 2,849 vehicles per day (n=33), the highest average for minor legs intersecting urban principal arterials (Figure 10).

In Figure 10 Washington is portrayed as having a high mean AADT associated with minor legs intersecting rural major arterials, but the sample size
was only two for this type of intersection. North Carolina exhibited limited variation between the high order functional classes having values similar to the sample mean AADT value.

![Average AADT by Major Leg Functional Class (PA and Major)](image)

Figure 13: Minor mean AADT grouped by whether the minor intersects a principal arterial or major arterial functional class in an urban or rural context then grouped by state.

Two tests were conducted regarding the principal arterial functional class to describe differences between rural and urban contexts. A Mann-Whitney U test found that the combined sample of urban and rural principal arterials did not tend toward higher or lower values than other functional classes (p=.90, ns).
Though, further testing of rural principal arterials in comparison to urban principal arterials suggested the AADTs associated with the intersection of urban principal arterials tended to have higher traffic volumes ($W = 7186$, $p = 0.016$).

A Kruskal-Wallis test between the states revealed that AADT values associated with minor legs intersecting urban arterials did not tend higher or lower values when compared across states ($K = 1.04$, df = 2, $p = 0.595$). Similar results across states for roadways intersecting rural principal arterials were found ($K = 0.02$, df = 2, $p = 0.99$).

4.1.3.3.2 Major Arterials

No legs intersecting classified urban major arterials were included in the sample data set. However, minor legs intersecting 14 rural major arterials were included in the sample with an average AADT of 1,372 vehicles per day. Comparison by state in Figure 10 reflects a large variation between the states. However, as mentioned earlier Washington accounted for two of the fourteen minor legs, Ohio accounted for zero minor legs, and North Carolina for the remaining 12.
When compared with legs intersecting principal arterials no significant difference was found between the AADTs of legs intersecting major arterials (n=14) and those intersecting principal arterials (n=220, p=.119, ns). Similarly, only a marginal difference was found between legs intersecting major arterials (n=14) and lower functional classes (n=240, p=0.091, ns).

4.1.3.3.3 Minor Arterials

Minor legs intersected both urban and rural minor arterials. 144 legs intersected urban minor arterials with an average AADT of 1,323 vehicles per day. Roadways intersecting 44 rural minor arterials had an average AADT of 856 (Figure 11). A test of similarity between the legs intersecting rural and urban minor arterials suggested a significant difference in AADT between legs intersecting rural and those legs intersecting urban arterials (W=4068, p=0.004).
Figure 14: Mean AADT by major leg rural or urban minor arterials and state.

Stratifying by functional class and comparing the results suggests that no significant difference exists between the AADTs associated with legs of minor roadways intersecting principal arterials and those intersecting minor arterials \((W=22760, \ p=0.54, \ ns)\). Similarly, the legs intersecting collectors and local streets did not have a significant difference in AADT \((W = 4841.5, \ p= 0.917, \ ns)\).
4.1.3.3.4 *Major Collector and Local Streets*

Comparing the lowest functional classes of major roadways revealed no significant differences between legs intersecting higher functional class roadways and those intersecting major collectors and local streets ($W=5892$, $p=0.738$, ns).

The difference between the means of AADT for minor legs intersecting arterials was 203 vehicles per day with first quarter values of 458 vehicles per day and third quarter percentiles of 1605. Both quartiles were only separated by approximately 50 vehicles per day. Comparison of the maximum values revealed that a maximum AADT of 15,635 vehicles per day ($n=422$) was contained within the higher functional classes while the legs intersecting collectors and local streets had a maximum AADT of 3310 vehicles per day ($n=52$).

4.2 *Land Use*

The following sub-sections discuss the variables generated for the purpose of accounting for the influence of land use on the roadway AADT.
4.2.1 Developed or Not

AADT for minor legs immediately connected with developed sites averaged 1,478 vehicles per day (n=328). For undeveloped sites the average minor leg AADT was 951 vehicles per day (n=146), a difference of 527 vehicles per day (Figure 19). Comparatively, a significant difference was found between the AADT values within each group (W=29122, p<.001).

Figure 15: Mean AADT by observed developed land use and state.
Comparison across the states demonstrated in Figure 12 revealed higher mean AADT values associated with Ohio, and lower mean AADT values associated with the North Carolina sample. The AADT values for roadways located in developed areas were statistically different when compared across the states ($K=9.6253$, df=$2$, $p=0.008$). Significance testing between the states revealed a significant difference in values between Washington and Ohio ($W=4257$, $p=0.035$). AADT values within the Ohio and North Carolina sample were significantly different ($W=3314$, $p=0.002$). No significant difference between Washington and North Carolina was determined ($W=7180$, $p=0.251$, ns).

### 4.2.2 Residence plus Open Space

Many intersections in the data sample were in the rural areas, but were servicing residential homes with small farms or open space, and in some cases were located in fringe areas with developed land and single residence farm land. Thus, summary values and testing accounting for both development and single family residence farm land are presented.

Only four legs within the model development sample were not determined to have residential open space or developed land use adjacent to the
 Seventy minor legs were found to have adjacent developed land in addition to open space land with residence. Minor legs with this surrounding land use averaged an AADT of 1,528 vehicles per day. When compared to non-developed land with residential farm the average traffic was 946 vehicles per day, a difference of 582 vehicles per day. A Mann-Whitney test suggested a significant difference between the values of mixed uses and farm only ($W=6692$, $p<.001$).

Of the 328 roadways attributed to developed land use, 70 of those were associated with farm land or open space. The average AADT for the remaining roadways only adjacent to developed land was 1,465 vehicles per day. Testing the groups for consistent values using the Mann-Whitney U test suggested that a significant difference exists between the two groups ($W=10434$, $p=.046$).
Ohio had the largest deviation from sample average as demonstrated by the state by state comparison presented in Figure 13. North Carolina had the least deviation from sample mean having very similar values for single residence large lots and those roadways associated with other land use. The Washington sample reflected similar variations to the sample mean data. This is likely due to similarities in distribution due to the contribution of the higher mean AADT values attributed to the Washington dataset.
4.2.3 Parking Lot Access

Another theorized influencing factor is the capture of parking lot access as a dummy variable. One hundred twenty nine minor roadways had parking lot access, either exit, entrance, or both along the leg ending at the next major roadway. The average AADT of these minor legs was 1,750 vehicles per day while minor legs without parking lot entrances and exits had an average of 1,154 vehicles per day. Testing the difference suggested a significance difference between the two groups of minor leg AADT values (W=28391, p<.001).
Figure 17: Mean AADT by whether or not parking lot accesses minor road and state.

When compared by state in Figure 14, Ohio had a higher average AADT than North Carolina and Washington. For roadways without parking lot access North Carolina had the lowest average AADT. Significance testing between the states revealed no significant difference in AADT values for roadways with parking lot access ($K=2.808$, $df=2$, $p=0.246$, ns), and no significant difference between the states for roadways without parking lot access ($K=0.459$, $df=2$, $p=0.795$, ns).
4.2.4 Within City Limits

Minor legs located with city limits averaged 1,438 vehicles per day compared to the 1,161 vehicles per day for legs located outside city boundaries. A Mann-Whitney test indicated the difference between the two groups of AADT values was non-significant ($W=29650$, $p=.18$, ns).

![Average AADT Within City/Town Boundary](image)

Figure 18: Mean AADT for sites within city or town boundaries by state.
Comparison between the states in Figure 15 reveals minor variation between the states. Supporting this analysis, significance testing suggests marginal significant difference between the states (K=5.326, df=2, p=0.07).

4.3 Network Characteristics

Descriptive statistics were calculated for the HPMS network within one, 2.5, five, and ten mile radial buffers. All legs were found to have at least one HPMS roadway within a five mile influence area. Since a five mile radial influence area is a substantial land area, inclusion in the modeling process was limited to the one and two and a half mile radial influence area. However, description of the differences in network is included to further discuss the variability in the data. Box plots demonstrating the summary statistics for the data including the mean AADT values by buffer and functional class are included in Appendix D.
4.3.1 Freeways

Table 4: Significance testing results for differences in AADT values across states based on presence of HPMS freeways within buffers.

<table>
<thead>
<tr>
<th>Network Buffer</th>
<th>AADT Significance Test Across States - Freeways</th>
</tr>
</thead>
<tbody>
<tr>
<td>Buffer</td>
<td>One or more Princ. Arterials</td>
</tr>
<tr>
<td></td>
<td>K-W Chi-Sq</td>
</tr>
<tr>
<td>1</td>
<td>1.033</td>
</tr>
<tr>
<td>2.5</td>
<td>1.362</td>
</tr>
<tr>
<td>5</td>
<td>1.382</td>
</tr>
<tr>
<td>10</td>
<td>0.437</td>
</tr>
</tbody>
</table>

The presence of freeway was limited within a one mile radius of the sites. 423 sites within the sample have no freeways within the 3.14 square miles surrounding the sites. The average AADT for the minor legs without a freeway within one mile was 1,306 vehicles per day. 161 minor legs were located in Washington with remaining sites located in North Carolina and Ohio split equally at 131 minor legs in each site.

An extension of 1.5 miles in radius resulted in an average AADT of 1,316 from 289 minor legs. The majority of the sites were located in Washington (n=131) with 70 legs located in Ohio, and 88 legs in North Carolina.

Further extension of the radial influence area decreased the number of sites to 209 without a freeway within a radius of five miles. Washington
composed 47 percent (n=99) of the sites without a freeway included in the HPMS. The remaining legs not having a freeway within the influence area were 24 percent (n=50) of the sites located in Ohio, and the remainder in North Carolina (n=60).

The fourth buffer defined the HPMS network characteristics in 314 square mile area around the minor leg. This 10 mile radial extension decreased the number of zero freeway sites to 134 with 79 of the sites located in Washington, 25 sites in Ohio, and 30 site locations in North Carolina.

A key characteristic of the distribution of the minor legs without freeways within the defined buffers was the predominance of Washington sites that did not have a HPMS freeway within the influence area. This was the case for all four influence areas.

4.3.2 Principal Arterials

Significance testing of AADT values by comparison across the states suggested no significant difference between ranked AADT for roadways within one mile, two and a half miles, five miles, and ten miles of at least one principal
Similarly, roadways without a principal arterial within all four buffers did not have a significant difference between the AADT values across the states.

Table 5: Significant testing results from Kruskal-Wallis testing of difference in AADT values between states.

<table>
<thead>
<tr>
<th>Network Buffer</th>
<th>AADT Significance Test Across States-Principal Arterials</th>
<th>No Principal Arterials</th>
</tr>
</thead>
<tbody>
<tr>
<td>Buffer</td>
<td>One or more Princ. Arterials</td>
<td>K-W Chi-Sq</td>
</tr>
<tr>
<td>1</td>
<td>1.238</td>
<td>2</td>
</tr>
<tr>
<td>2.5</td>
<td>2.037</td>
<td>2</td>
</tr>
<tr>
<td>5</td>
<td>1.148</td>
<td>2</td>
</tr>
<tr>
<td>10</td>
<td>0.352</td>
<td>2</td>
</tr>
</tbody>
</table>

Within the smallest influence area principal arterials were more prominent than freeways. Only 181 sites did not have at least one principal arterial within one mile of the count location. The average AADT for these roadways was 1,156 vehicles per day compared to an average AADT of 1,414 vehicles per day for roadways with at least one principal arterial within one mile. State comparison of the of minor legs without a principal arterial, 32 legs were located in Washington, 62 located in Ohio, and North Carolina had the remaining 87 legs without a principal arterial.

North Carolina over represented minor legs without a principal arterial within one mile, and Washington was under represented. Comparison of the
distribution of the states within the entire sample, North Carolina represents 156 minor legs or 33 percent of the sample. However, of the 181 minor legs with no principal arterials North Carolina represented 48 percent of minor legs. Washington contributed 18 percent of the minor legs located in the state without a principal arterial within one mile.

A decrease in the number of sites without a principal arterial to 96 was experienced with a 1.5 mile increase in the radius of influence. Ohio sites decreased to 40 minor legs with the increase in radius of influence. Washington represented 20 of the sites with North Carolina having 36 minor legs with no principal arterials within the almost 20 square mile influence area.

For the influence area of almost 79 square miles a 50 percent decrease in the number of minor legs without a principal arterial was experienced. Minimal changes were found in Washington state (n=18). The primary changes occurred with legs located in Ohio and North Carolina decreasing 14 and 15 legs within the states, respectively.

All but fifteen sites had at least one principal arterial within a ten mile radius of the roadway. Washington had ten of the sites, Ohio one site, and North Carolina with the remaining four sites. Again, the state of Washington has large
population centers in and around the Seattle area, but many of the intersections in the rural areas were along rural highways away from population centers, thus the lack of high volume roadways near the locations.

4.3.3 Minor Arterials

Table 6: Significance testing results for differences in AADT values across states based on presence of HPMS minor arterials within buffers.

| Network Buffer AADT Significance Test Across States-Minor Arterials |
|-------------------------|----------------------|----------------------|
| Buffer      | One or more Princ. Arterials | No Principal Arterials |
| K-W Chi-Sq | df       | p (α=0.05)  | K-W Chi-Sq | df       | p (α=0.05)  |
| 1           | 3.71     | 2          | 0.156       | 1.947     | 2          | 0.378       |
| 2.5         | 1.746    | 2          | 0.418       | 0.139     | 2          | 0.933       |
| 5           | 0.766    | 2          | 0.682       | 0.663     | 2          | 0.718       |
| 10          | 0.377    | 2          | 0.828       | 2.25      | 2          | 0.134       |

Minor arterials were more prevalent in the HPMS data used to define the surrounding network characteristics. Thus, only 159 sites did not have a minor arterial located within one mile. Washington accounted for 60 sites, Ohio 57 sites, and North Carolina 42 sites.

Again, extending the area of influence decreased the number legs without a minor arterial. 106 legs did not have a minor arterial located within 2.5 miles. Of these sites, 43 sites were located in Washington, 41 in Ohio, and 22 in North
Carolina. The decrease was uniform across the three states with a decrease of near 20 sites experienced by each state.

Only 51 legs had zero minor arterials within five miles. Legs within Washington accounted for 28 of the sites. Ohio accounted for 11 and North Carolina accounted for 12.

467 sites had a minor arterial within ten miles. Of the seven sites not having a minor arterial, six were located in Washington and one in North Carolina.

4.3.4 Major Collectors

Table 7: Significance testing results for differences in AADT values across states based on presence of HPMS major collectors within buffers.

<table>
<thead>
<tr>
<th>Buffer</th>
<th>K-W Chi-Sq</th>
<th>df</th>
<th>p (α=0.05)</th>
<th>K-W Chi-Sq</th>
<th>df</th>
<th>p (α=0.05)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1.14</td>
<td>2</td>
<td>0.567</td>
<td>2.939</td>
<td>2</td>
<td>0.23</td>
</tr>
<tr>
<td>2.5</td>
<td>0.324</td>
<td>2</td>
<td>0.851</td>
<td>2.63</td>
<td>1</td>
<td>0.105</td>
</tr>
<tr>
<td>5</td>
<td>0.531</td>
<td>2</td>
<td>0.767</td>
<td>1</td>
<td>1</td>
<td>0.371</td>
</tr>
<tr>
<td>10</td>
<td>0.451</td>
<td>2</td>
<td>0.798</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

The lowest level functional class included in this analysis was major collectors. The HPMS system provides functional classes for minor collectors, but none were included in the largest influence areas.
111 legs were found to not have any HPMS defined roadways of major collector or higher. The state with the largest number of sites without a major collector within one mile was North Carolina (n=56). Ohio and Washington had 29 and 26 legs respectively not having a major collector within one mile.

The three larger influence areas had no more than 15 legs without a major collector within at least 2.5 miles. All sites in Ohio had at least one leg with a major collector within a 2.5 mile radius. Only one site in Washington did not have a major collector within 2.5 miles. The remaining 14 sites were located in North Carolina. 2 of the 474 minor legs did not have a major collector within 5 miles; one each located in Washington and North Carolina. Finally, a ten mile radius found no legs without at least one major collector.

4.4 Accessibility

Several variables were defined to describe the accessibility of the surrounding transportation system. Mileage values describing the accessibility of limited access facilities were analyzed as was the distance to the nearest city or town, and are presented in the following sections.
4.4.1 Accessibility to nearest city

Direct access is defined by a road connecting within two miles of the intersection to the city or town boundary as represented in Google Maps. Conversely, indirect access means the leg of the roadway does not connect with the city or town boundary within the two miles.

Comparison of the mean AADT by direct and indirect access suggested no significant difference between those roadways with direct access and those with indirect access ($W=29650$, $p=0.186$, ns). The mean AADT for direct access roadways was 1,438 vehicles per day, and for indirect roadways the mean AADT was 1,260 vehicles per day (Figure 16). The mean AADT by sample and states is presented in Figure 33.
Originally, roadways accessing the outskirts of a city or town had its own classification. However, the subjectivity of the classification could not be rationalized for the purpose of the research effort. Thus, classification of the fringe groups was incorporated into direct classification due to higher population density assumed to exist on the fringe of a city or town.

4.5 Socioeconomic

To describe the population and account for the influence of income and vehicle ownership several variables were included in the analysis. These values were captured from the ACS 5 year estimates for 2008 through 2012. Data from
190 census tracts were included in the data set. The average number of legs for each census tract was 2.5 with a first quartile of one minor roadway leg and a third quartile of three minor roadway legs per tract. The minimum was one leg and the maximum 11 minor leg roadways.

4.5.1 Population

Of the 474 minor legs, 74 roadways were located in census tracts without an urban population. 400 roadways were associated with urban populations, of those 185 were associated with urban and rural populations at the census tract level.

For roadways located within census tracts with urban populations the average urban population was 3,028 persons per tract, and the average rural population was 1,795 persons per census tract. Overall, the average total population per tract was 4,823 persons.

4.5.2 Population Density

Population density was generated using the area of the census tract and the population of the census tract. The average population density was 2.52
persons per acre with a minimum of 0.006 persons per acre and a maximum of 14.54 persons per acre.

### 4.5.3 Housing

Housing by census tract was included to attempt to suggest the variation in surrounding land use. The average number of urban households in a census tract was 1296 households per census tract with a minimum value of 0 (n=74) and maximum value of 3336 households. The average number of rural households was 789 with a minimum value of 0 (n=215) and a maximum of 3824 households per census tract.

### 4.5.4 Vehicle Ownership

The effect vehicle ownership has on roadway volume was accounted for using census tract information. The average number of private vehicles per census tract was 3634. Since not all members of the population can drive by themselves, specifically the population under 16 years of age, vehicles per persons 16 years of age and older was calculated. For all census tract this average was 0.96 vehicles per person 16 years of age or older, or almost one vehicle per assumed driver.
4.5.5 Income

To account for the variation that may exist in vehicle use along the minor legs per capita income for census tract was collected. The average per capita income was $28,136 US dollars with a minimum value of $8,386 and a maximum per capita income of $81,627.

4.5.6 Employment

Thirteen industries were included in the analysis of disaggregate employment. The numbers were gathered from the 2008-2012 ACS Estimates Economic table DP-03 aggregated at the census tract level.

Table 8: Descriptive statistics of percent of civilian labor force in census tract by industry.

<table>
<thead>
<tr>
<th>Industry Category</th>
<th>Percent of Civ. Labor Force</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
</tr>
<tr>
<td>Ag., Forestry, Mining</td>
<td>0.03</td>
</tr>
<tr>
<td>Construction</td>
<td>0.07</td>
</tr>
<tr>
<td>Manufacturing</td>
<td>0.14</td>
</tr>
<tr>
<td>Wholesale</td>
<td>0.03</td>
</tr>
<tr>
<td>Retail</td>
<td>0.12</td>
</tr>
<tr>
<td>Transportation</td>
<td>0.05</td>
</tr>
<tr>
<td>Information</td>
<td>0.02</td>
</tr>
<tr>
<td>Finance</td>
<td>0.05</td>
</tr>
<tr>
<td>Professional</td>
<td>0.10</td>
</tr>
<tr>
<td>Education</td>
<td>0.22</td>
</tr>
<tr>
<td>Arts</td>
<td>0.08</td>
</tr>
<tr>
<td>Public Administration</td>
<td>0.05</td>
</tr>
<tr>
<td>Other</td>
<td>0.05</td>
</tr>
</tbody>
</table>
Table 8 reflects the average percent of the civilian labor force for each of the 13 industry categories included in the analysis. Education has the highest average with 22 percent of the labor force, and maximum value of 50 percent. The lowest averages are transportation, agriculture, public administration, information, finance, wholesale, and other. Of the lowest averages, the maximum percentages are relatively high with 22 percent maximum in agriculture, 18 percent in transportation and finance.
5.0 ANALYSIS

The following chapter presents the analysis and results obtained from the model generation process and verification using the reserved sites. This chapter is organized by presenting the results from modeling the entire sample and by state.

5.1 Sample Analysis

Using the methods outlined in section 3.3, Table 9 depicts the best model fit with the least bias in the predictor variables. At least one variable from the five groups was a significant predictor of the ln(AADT). The best fit model for the entire sample contained eleven significant variables accounting for approximately 32 percent of the variation in the ln(AADT).

Classification of an urban minor arterial resulted in a significant increase of 0.232 in the dependent variable. Similarly, ln(AADT) for minor legs significantly increased by 0.492 when intersecting rural major arterials. Minor leg classification as a local or unclassified roadway had a decreasing effect of -1.164. If the roadway did not access a city or town ln(AADT) experienced a significant
negative effect of -0.48. The effect of parking lot entrances or exits accessing the minor legs was positive with influence of 0.251 if present.

An increase in the number of principal arterials within a one-mile radius of the intersection had a positive increase of 0.154. While an increase in the number of major collectors within a 2.5 mile radius had a significant increasing effect of 0.022.

Two variables reflected roadway characteristics; presence of right turn lane on minor approach, and edge striping on minor leg near the intersection. The presence of a right turn lane on the minor leg approach resulted in a significant increase in ln(AADT) of 0.402. Edge line striping had the largest effect on ln(AADT) with a positive effect of 0.604 if the roadway was striped.
Table 9: Summary of model performance and significant variables for three state sample.

<table>
<thead>
<tr>
<th>Sample</th>
<th>All</th>
<th>Outliers Removed</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Coef. p-value</td>
<td>Coef. p-value</td>
</tr>
<tr>
<td>(Intercept)</td>
<td>7.879 p&lt;.001</td>
<td>6.874 p&lt;.001</td>
</tr>
<tr>
<td>Urban Min. Art.</td>
<td>0.232 0.001</td>
<td>-</td>
</tr>
<tr>
<td>Rural Maj. Art.</td>
<td>0.492 0.024</td>
<td>0.504 0.016</td>
</tr>
<tr>
<td>Local Street</td>
<td>-1.164 p&lt;.001</td>
<td>-0.546 p&lt;.001</td>
</tr>
<tr>
<td>Collector</td>
<td>-0.523 0.009</td>
<td>-</td>
</tr>
<tr>
<td>Pop. Density</td>
<td>-0.048 0.004</td>
<td>-0.044 0.004</td>
</tr>
<tr>
<td>Not to City/Town</td>
<td>-0.478 p&lt;.001</td>
<td>-0.606 p&lt;.001</td>
</tr>
<tr>
<td># of Princ. Art (1mi)</td>
<td>0.154 p&lt;.001</td>
<td>-</td>
</tr>
<tr>
<td># of Maj. Col. (2mi)</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Parking Lot Acc.</td>
<td>0.251 0.004</td>
<td>0.316 p&lt;.001</td>
</tr>
<tr>
<td>Edge Striping</td>
<td>0.604 p&lt;.001</td>
<td>0.619 p&lt;.001</td>
</tr>
<tr>
<td>Right Turn Lane</td>
<td>0.402 0.025</td>
<td>-</td>
</tr>
<tr>
<td>Ag. Employment</td>
<td>-4.310 p&lt;.001</td>
<td>-</td>
</tr>
<tr>
<td>Oth. Employment</td>
<td>-3.953 0.031</td>
<td>-</td>
</tr>
<tr>
<td>Public Admin. Emp.</td>
<td>-3.355 0.012</td>
<td>-</td>
</tr>
<tr>
<td>( R^2: )</td>
<td>0.339</td>
<td>0.344</td>
</tr>
<tr>
<td>RMSE:</td>
<td>0.78</td>
<td>0.72</td>
</tr>
<tr>
<td>p-value</td>
<td>p&lt;.001</td>
<td>p&lt;.001</td>
</tr>
</tbody>
</table>

As a measure of variability in the errors or how well the model predicts the fitted values, the root mean square error is the standard deviation of the residuals. For the full sample model the RMSE was calculated to be 0.78. Compared to the range of the dependent variable, 4.83 to 9.657, with a mean of 6.72 the variability in the error terms is exposed. To further demonstrate, one standard deviation above the mean (exp(7.5)=1808 AADT, exp(6.72)=828 AADT)
when compared to the mean results in a 218 percent error in AADT. RMSE for
the outlier free model is 0.72, with a range of 4.38 to 9.02 reflecting similar
variability, and decreased maximum ln(AADT).

5.1.1 Sample Outliers

Outliers were identified using DFFITS to measure influence of cases on
the overall model. Using the rule of thumb for large datasets identified in Cohen
et al. (2003) 29 sites with DFFITS values above 0.30 or below -0.30 were removed
from the data set resulting in a minimal increase in the described variance of
0.005.
The minor AADT range for the excluded sites included 130 vehicles per day at the minimum and the maximum value of 15,635 vehicles per day. Figure 17 provides visual representation of the excluded outliers graphed by AADT and numerical site identifier. Most notable all of the maximum AADT for each state were identified as outliers. Of the AADT assumed to be outliers 13 were from Ohio, nine from Washington, and the remaining seven outliers were from North Carolina.

Figure 20: AADT specific identification of assumed outliers.
Appendix E contains the index plots (Figures E1 and E2) for the individual independent outlier DFBETAS values. Although some variation existed with some of the variables no values were outside Cohen et al. (2003) suggested rule of thumb of +/- 1.

5.1.2 Diagnostics

The underlying assumption of linear regression is that a linear relationship between the minor leg AADTs and the chosen predicted variables exists in some capacity. A component of the lack of strong relationship between the predictor variables and AADT could be the failure of the analysis to transform a non-linear relationship, or model the data set as such.
Figure 21: Diagnostic plot checking linear assumptions for data.

Homoscedascity plots are presented in Figure 18. The upper plot suggests that the variability in the residuals decreases with an increase in the predicted
value. Similar behavior was captured with population density and the number of major collectors within 2.5 miles of the intersection. Both instances suggest that residuals approach zero with an increase in the variable. However, the higher values are not as prominent as lower values as demonstrated in Figure 18. It is also likely that the higher values of the independent variables can be attributed to higher AADT values, thus the possibility for a more accurate model exists by modeling the sites associated with higher AADT values.

A normal curve overlaying a histogram of the residuals for the full sample model is presented in Figure 19. The normality assumption regarding the residuals is further confirmed by a skewness value of 0.08 and a kurtosis value of 3.05.
5.1.3 Higher Volume Minor Legs

Table 9 contains the results of modeling the 47 stop-controlled minor legs above 3,000 AADT. An increase in AADT with the intersection of a rural or urban principal arterial was found for minor legs with 3,000 or more vehicles per day. Left turn lanes and location within city limits had a positive effect on the
higher volume minor legs. Counter-intuitively, the presence of centerline striping had a negative effect on the predicted value of the minor leg volume.

The most notable result of the subset model was the significant 87 percent of variance; the highest result of all the models included herein. As well as the low variability in the residuals (RMSE=0.14) suggesting good predictability of the model.

Table 10: Model result for minor legs with 3,000 or more annual average daily traffic.

<table>
<thead>
<tr>
<th>3,000 AADT or Larger</th>
<th>R²</th>
<th>0.872</th>
<th>RMSE: 0.14</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Estimate</td>
<td>Std. Error</td>
<td>t value</td>
</tr>
<tr>
<td>Intercept</td>
<td>8.562</td>
<td>0.124</td>
<td>69.2</td>
</tr>
<tr>
<td>Urban Prin. Art</td>
<td>0.406</td>
<td>0.062</td>
<td>6.6</td>
</tr>
<tr>
<td>Rural Princ. Art</td>
<td>0.486</td>
<td>0.113</td>
<td>4.5</td>
</tr>
<tr>
<td>Two-Way Left Turn Lane</td>
<td>0.225</td>
<td>0.061</td>
<td>3.7</td>
</tr>
<tr>
<td>Avail. Vehicles per Person</td>
<td>0.0001</td>
<td>0.00002</td>
<td>4.7</td>
</tr>
<tr>
<td>Not to City/Town</td>
<td>-0.452</td>
<td>0.086</td>
<td>-5.2</td>
</tr>
<tr>
<td>Res. Farm</td>
<td>-0.201</td>
<td>0.075</td>
<td>-2.7</td>
</tr>
<tr>
<td>Parking Lot</td>
<td>-0.127</td>
<td>0.057</td>
<td>-2.2</td>
</tr>
<tr>
<td>Centerline Striping</td>
<td>-0.292</td>
<td>0.066</td>
<td>-4.4</td>
</tr>
<tr>
<td>Left Turn Lane</td>
<td>0.883</td>
<td>0.131</td>
<td>6.8</td>
</tr>
<tr>
<td>Construction Emp.</td>
<td>-2.336</td>
<td>0.823</td>
<td>-2.8</td>
</tr>
<tr>
<td>Finance Emp.</td>
<td>-3.966</td>
<td>1.136</td>
<td>-3.5</td>
</tr>
</tbody>
</table>

F-statistic: 21.75 on 11 and 35 DF  p < .001
5.2 **State**

Exploratory analysis of the states provided insight into the variation between the states. As was expected, each state had different significant models. The model using only data for Ohio explained approximately 48 percent of the variance in the minor volumes. The Washington data explained almost 50 percent of the variance in the Washington AADT values. North Carolina had the least amount of variance (36%).

5.2.1 **Ohio**

The Ohio locations were most impacted functional class and network characteristics (Table 10). Two variables within the 2.5 mile radial buffer were significant. The number of freeways in the 2.5 mile radial buffer contributed to a decrease in AADT while the number of major collectors was associated with a slight increase in AADT. Functional class of the minor leg as a local street had a decreasing effect on the traffic volume while intersecting with an urban principal arterial or urban minor arterial had an increasing effect on AADT.

Of the accessibility and demographic variables included in the exhaustive method, one from each group was significant. Per capita income accounted for a
small increase in volume with an increase in income. A minor leg not having
direct access to a town or city had a negative impact on AADT.

The percentage of civilian labor force within the respective census tract
employed in agriculture, forestry, fishing/hunting, and mining had an increasing
effect of .07 for every percent increase in employment. A similar increase (0.08)
was identified with a single percent increase in transportation related
employment.

Table 11: Multiple Regression model results for count sites in Ohio.

| Ohio               | R² | Estimate | Std. Error | t value | Pr(>|t|) |
|--------------------|----|----------|------------|---------|---------|
| Intercept          |    | 3.725    | 0.603      | 6.2     | p < .001|
| Urban Princ. Art.  |    | 0.885    | 0.260      | 3.4     | p < .001|
| Urban Minor Art.   |    | 0.952    | 0.260      | 3.7     | p < .001|
| Minor Local Street |    | -0.732   | 0.267      | -2.7    | 0.007   |
| Per Capita Income  |    | 0.00008  | 0.00001    | 5.7     | p < .001|
| Not to City/Town   |    | -0.422   | 0.175      | -2.4    | 0.017   |
| # of Freeways (2mi)|    | -0.459   | 0.113      | -4.1    | p < .001|
| # of Maj. Coll. (2mi)| | 0.107    | 0.021      | 5.2     | p < .001|
| Centerline Striping|    | 0.607    | 0.178      | 3.4     | p < .001|
| Ag. Employment     |    | 7.312    | 2.710      | 2.7     | 0.004   |
| Transport. Employment| | 8.447    | 2.500      | 3.4     | 0.002   |

F-statistic: 12.06 on 10 and 133 DF  p < .001

RMSE for the Ohio model was the highest of the three states. The
dependent variable range was 4.38 to 9.10, with a mean value of 6.767, thus a
predicted value a single standard deviation of error above the mean would result in a 1080 AADT difference.

5.2.2 Washington

No continuous variables were significant predictors of Washington AADT included in the analysis with the exception of the percent of the civilian labor force working agriculture, forestry, fishing, hunting, and mining (AG) (Table 11). Functional classifications of Washington count station roadways and nearest state highways were important in predicting the logarithm value of AADT. Classification of the nearest state highway as a minor arterial had significant negative effect on Washington AADT. Unclassified and local street count station roadways had the largest negative effect on the logarithm of the AADT with the exception of an increase in AG. Classification of the count station roadway as a collector also had a negative effect on the AADT.
Table 12: Multiple Regression model results for count sites in Washington

|                      | Estimate | Std. Error | t value | Pr(>|t|) |
|----------------------|----------|------------|---------|----------|
| Intercept            | 7.908    | 0.192      | 39.7    | p < .001 |
| Rural Minor Art.     | -0.683   | 0.254      | -2.7    | 0.006    |
| Minor Local Street   | -1.428   | 0.191      | -6.1    | p < .001 |
| Minor Collector      | -0.723   | 0.208      | -3.1    | p < .001 |
| Not to City/Town     | -0.425   | 0.128      | -4.9    | 0.002    |
| Edge Striping        | 0.635    | 0.122      | 4.9     | p < .001 |
| Left Turn Lane       | 0.871    | 0.303      | 3.3     | 0.004    |
| Ag. Employment       | -5.202   | 1.437      | -3.6    | p < .001 |

F-statistic: 23.37 on 6 and 166 DF  p < .001

The binary indicator variable of no direct access to the fringe or center of a city or town negatively impacted the AADT. Conversely, the presence of an edge line and left turn lane both had positive impacts on the logarithm of the AADT.

Though predictability of the individual states is less the focus of this effort, and more to explore the differences in predictors between the states, the RMSE for the Washington model was relatively low. The dependent variable range was 4.5 to 9.66, with a mean value of 6.74, thus a predicted value one single error standard deviation above the mean would demonstrate a difference of 800 AADT.
5.2.3 North Carolina

The North Carolina was the worst performing model between the states (Table 12). Using the Mallow’s Cp model selection criterion, the model with least bias was a six variable model including the intercept. Classification of the nearest state highway as an urban minor arterial had a positive increasing effect on AADT. Also having a positive effect on estimated AADT were the number of freeways within a 2.5 mile radial buffer, the presence of a parking lot entrance and/or exit, and the presence of a centerline on the count station road near the intersection with the nearest state highway. Interestingly, an increase in population density was found to have a negative effect on the logarithm of the AADT.

Table 13: Multiple regression modeling results for sites located in North Carolina.

<table>
<thead>
<tr>
<th>North Carolina</th>
<th>$R^2$: 0.362</th>
<th>RMSE: 0.612</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Estimate</td>
<td>Std. Error</td>
</tr>
<tr>
<td>Intercept</td>
<td>6.056</td>
<td>0.141</td>
</tr>
<tr>
<td>Urban Minor Art.</td>
<td>0.383</td>
<td>0.127</td>
</tr>
<tr>
<td>Population Density</td>
<td>-0.109</td>
<td>0.032</td>
</tr>
<tr>
<td># of Freeways (2mi)</td>
<td>0.275</td>
<td>0.062</td>
</tr>
<tr>
<td>Parking Lot Access</td>
<td>0.561</td>
<td>0.108</td>
</tr>
<tr>
<td>Centerline Striping</td>
<td>0.660</td>
<td>0.130</td>
</tr>
</tbody>
</table>

F-statistic: 17.04 on 5 and 150 DF p < .001
North Carolina had the lowest RMSE at 0.612. The range of the dependent variable was also the smallest with a minimum value of 4.61 and a maximum of 8.5. An error single standard deviation above the mean of 6.67 reflected a difference between of 665 AADT in the non-log transformed values.

5.3 Model Validation

Validation of the overall model with the exclusion of assumed outliers reflected high percent error. Fifty four sites were held in reserve for validation and error influence on the safety performance functions. The average percent error for the validation was 59 percent with a minimum percent error of one percent and maximum of 214 percent. Thirty-one sites had higher than a 50 percent error compared to the actual values.

5.3.1 Crash Prediction

As a compliment to the AADT model generation, crash values were estimated using the actual major and minor AADT in comparison to the predicted minor AADT values coupled with the actual major values. Using the Highway Safety Manual’s Safety Performance Functions (SPF) for three and four
leg stop-controlled intersections allowed for quantifying the difference between the actual and predicted crashes due to estimated minor volumes.

Figure 20 is a product of this analysis and demonstrates the difference between the actual and predicted values for three-legged intersections. Most notably the larger differences are associated with the higher minor leg AADT values.

Figure 23: Comparison of Three-legged Stop Controlled SPF generated crash values using Predicted and Actual AADT Values.
The crash prediction average absolute error associated with the use of actual values versus predicted values was 32 percent. Of the 25 three leg intersections five had greater than 50 percent error with a maximum error of 74 percent. Similar error percentages were calculated for the four leg intersection crash prediction function (Figure 21).

![Four-Legged SPF Crash Prediction Comparison](image)

Figure 24: Comparison of Four-legged Stop Controlled SPF generated crash values using Predicted and Actual AADT Values.

An absolute error percentage of 38 percent was calculated with the maximum error of 100 percent attributed to a minor leg AADT of 102 vehicles per day. Interestingly, the crashes predicted at this site were less than 0.4 using
the actual data, and almost 0.8 using the estimated AADT (320 vehicles per
day). Although the error associated with this value is large, the error is likely
acceptable based on the low crash prediction value.

Generally, larger absolute percent errors were calculated with high major
AADT values and lower minor AADT values. This makes sense as the majority
of these sites also had high percent errors associated with the volume estimation
model. The error was not equally transferred to crash prediction as major AADT
plays a mathematically dominant role in both SPF equations, thus limiting the
effect of high percent errors associated to the volume estimation process.
6.0 DISCUSSION

The purpose of this modeling effort was to develop a reasonably accurate model that can be used to estimate volumes on minor legs near stop-controlled intersections, and quantify the impact in error generation when using the safety performance functions located in the Highway Safety Manual. The five models presented herein demonstrated the complexities of generating linear regression models from three states all having different characteristics and social conventions that are difficult to capture using a quantitative approach, especially a simplistic approach like ordinary least squares regression.

The redundant theme in this analysis is the lack of strong relationship of the independent variables with the dependent variables. One of the problems with network wide traffic counts is the expenditures necessary to count all the local streets. This is reflected in the use of the Highway Performance Monitoring System GIS network. It was assumed for this analysis that a denser network could be expected to infer higher levels of traffic on the overall system. However, the HPMS lacked the definition to capture the nuance of the local street environment. This is especially the case in the urban environment where minor
collectors and local street volumes are normally excluded from counting programs, but can carry higher volumes due to many factors like residential or business density.

Ideally, land use on a microscopic level could be accounted for in the modeling process. The addition of number of business, type of businesses, number of residences, and other factors could increase the predictive accuracy of the overall model by defining site specific characteristics. The intent of this effort was to include land-use and tax information at the parcel level, but inconsistent data across counties and states forced the exclusion of this microscopic data.

Instead binary variables attempting to create the ambience of the surrounding land-use were captured. Only the presence of a parking lot entrance/exit was included in the overall model as a result of the analysis. Surprisingly, the indicator for urban or rural location was not included in the best-fit model. A possible explanation for the lack of significance is the variation in volumes regardless of the urban or rural location of the roadway.

Socioeconomics were accounted for by the inclusion of census tract data. One hundred and ninety census tracts were included in the analysis, which reflects an average of approximately 2.5 sites per census tract. The focus of the
inclusion of census tract data was to include population density, income, and housing density based on the assumption that affluence and housing would have an effect on local traffic. These assumptions were not supported by the analysis.

Most interesting is the negative effect that population density had on the overall model, and the North Carolina model. A potential explanation for this counter-intuitive result could be reflective of the influence of rural higher volume roadways surrounded by low population density. Or, as population density increases in suburban areas traffic may move to local streets at lower volumes.

Exploratory state by state analysis revealed longitudinal markings as significant predictors of the minor leg volumes across all models. Edge line striping was significant in the best-fit overall model, as well as in the Washington state specific model. Center line striping was significant for both Ohio and North Carolina. The significance of roadway characteristics is unsurprising due to the influence traffic volume has on decisions regarding transportation safety and operations.

Estimating crashes provided the most interesting results. Although larger errors resulted from large errors in the volume estimation process, the effect was
not transferred at the same level due to the emphasis of crash prediction on the major roadway in the SPF equation.
7.0 CONCLUSION

The result of this effort was the conclusion that the variables collected and used in this analysis were not good estimators of minor leg volumes at stop-controlled intersections. A myriad of factors likely contribute to the volume on these roadways and are specific to the surrounding area. Disaggregation to state specific models demonstrated a better explanation of the variance, but each of the three models varied in predictors further demonstrating the lack of a definitive overall model.

A key factor described in this analysis is the high level of variability within dependent variable. Outlier analysis suggested the removal of 26 sites which increased the adjusted explanation of variance by approximately three percent to 33.3 percent, but resulted in a model not in comparison to the results of similar research efforts with similar variables (Dixon et al., 2012).

Influences at each site are likely dependent on the immediate surroundings of the intersection including land-use factors and network characteristics. The influence of HPMS network in the models suggests that the surrounding network has some predictive ability, and could benefit from more
extensive models with disaggregated land-use data; an area for further investigation.

Lastly, application of the predicted and actual AADT values demonstrated the low sensitivity of safety performance functions to poorly estimated volumes. However, the high error associated with this volume estimation model does not render this an effective method for estimating AADT for use in crash prediction.
8.0 REFERENCES


APPENDIX A: List of Variables Used in Analysis

Table A: Variables associated with Roadway Characteristics

<table>
<thead>
<tr>
<th>Variable</th>
<th>Description</th>
<th>Variable Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>Min_AADT</td>
<td>Minor road volume (dependent variable)</td>
<td>Cont.</td>
</tr>
<tr>
<td>Maj_AADT</td>
<td>Intersecting state highway</td>
<td>Cont.</td>
</tr>
<tr>
<td>min_edge</td>
<td>Count station roadway has edgeline (yes = 1, no = 0)</td>
<td>Binary</td>
</tr>
<tr>
<td>min_center</td>
<td>Count station roadway has centerline (yes = 1, no = 0)</td>
<td>Binary</td>
</tr>
<tr>
<td>right_turn</td>
<td>Count station roadway has right turn lane at major intersection (yes = 1, no = 0)</td>
<td>Binary</td>
</tr>
<tr>
<td>left_turn</td>
<td>Count station roadway has left turn lane at major intersection (yes = 1, no = 0)</td>
<td>Binary</td>
</tr>
<tr>
<td>divided</td>
<td>Major is divided roadway = 1, Other = 0</td>
<td>Binary</td>
</tr>
<tr>
<td>twltl</td>
<td>Major has two-left turn lane at intersection = 1, Other = 0</td>
<td>Binary</td>
</tr>
</tbody>
</table>

Table A 1: Variables associated with Functional Class

<table>
<thead>
<tr>
<th>Variable</th>
<th>Description</th>
<th>Variable Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>U_PA</td>
<td>Major is urban principal arterial = 1, Other = 0</td>
<td>Binary</td>
</tr>
<tr>
<td>U_MnA</td>
<td>Major is urban minor arterial = 1, Other = 0</td>
<td>Binary</td>
</tr>
<tr>
<td>R_PA</td>
<td>Major is rural principal arterial = 1, Other = 0</td>
<td>Binary</td>
</tr>
<tr>
<td>R_MajA</td>
<td>Major is rural major arterial = 1, Other = 0</td>
<td>Binary</td>
</tr>
<tr>
<td>R_MinA</td>
<td>Major is rural minor arterial = 1, Other = 0</td>
<td>Binary</td>
</tr>
<tr>
<td>R_MajC</td>
<td>Major is rural major collector = 1, Other = 0</td>
<td>Binary</td>
</tr>
<tr>
<td>min_local</td>
<td>Minor is local street or unclassified = 1, Other = 0</td>
<td>Binary</td>
</tr>
<tr>
<td>min_coll</td>
<td>Minor is collector = 1, Other = 0</td>
<td>Binary</td>
</tr>
<tr>
<td>min_art</td>
<td>Minor is arterial = 1, Other = 0</td>
<td>Binary</td>
</tr>
</tbody>
</table>

Table A 2: Variables associated with land use characteristics

<table>
<thead>
<tr>
<th>Variable</th>
<th>Description</th>
<th>Variable Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>urbRur</td>
<td>Urban = 0, Rural = 1</td>
<td>Binary</td>
</tr>
<tr>
<td>developed</td>
<td>Adjacent land is developed (yes = 1, no = 0)</td>
<td>Binary</td>
</tr>
<tr>
<td>res_farm</td>
<td>Adjacent land has single unit residence associated with large open area (yes = 1, no = 0)</td>
<td>Binary</td>
</tr>
<tr>
<td>parking_lot</td>
<td>Count station roadway has parking lot entrance and/or exit (yes = 1, no = 0)</td>
<td>Binary</td>
</tr>
<tr>
<td>city_lim</td>
<td>Count station is located within city limits (yes = 1, no = 0)</td>
<td>Binary</td>
</tr>
</tbody>
</table>

Table A 3: Network related independent variables

<table>
<thead>
<tr>
<th>Variable</th>
<th>Description</th>
<th>Variable Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mi1_freeways</td>
<td>Number of freeways within one mile of the count station</td>
<td>Cont.</td>
</tr>
<tr>
<td>Mi1_pas</td>
<td>Number of principal and other principal arterials within one mile of the count station</td>
<td>Cont.</td>
</tr>
<tr>
<td>Mi1_maj_arcs</td>
<td>Number of major arterials within one mile of the count station</td>
<td>Cont.</td>
</tr>
<tr>
<td>Mi1_njcs</td>
<td>Number of major collectors within one mile of the count station</td>
<td>Cont.</td>
</tr>
<tr>
<td>Mi2_freeways</td>
<td>Number of freeways within 2.5 miles of the count station</td>
<td>Cont.</td>
</tr>
<tr>
<td>Mi2_maj_arcs</td>
<td>Number of principal and other principal arterials within 2.5 miles of the count station</td>
<td>Cont.</td>
</tr>
<tr>
<td>Mi2_njcs</td>
<td>Number of major arterials within 2.5 miles of the count station</td>
<td>Cont.</td>
</tr>
<tr>
<td>Mi2_njcs</td>
<td>Number of major collectors within 2.5 miles of the count station</td>
<td>Cont.</td>
</tr>
</tbody>
</table>
### Table A 4: Independent variables associated with road accessibility

<table>
<thead>
<tr>
<th>Variable</th>
<th>Description</th>
<th>Variable Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dist_Inter_m</td>
<td>Distance of count station to nearest interstate access point (miles)</td>
<td>Cont.</td>
</tr>
<tr>
<td>Dist_Arterial</td>
<td>Distance of count station to nearest limited access roadway other than interstate (miles)</td>
<td>Cont.</td>
</tr>
<tr>
<td>indirect</td>
<td>Count station road does not access town/city (yes = 1, no = 0)</td>
<td>Binary</td>
</tr>
<tr>
<td>direct</td>
<td>Count station road directly accesses town/city (yes = 1, no = 0)</td>
<td>Binary</td>
</tr>
</tbody>
</table>

### Table A 5: Independent variables describing population, housing, income, and available vehicles

<table>
<thead>
<tr>
<th>Variable</th>
<th>Description</th>
<th>Variable Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>tot_p</td>
<td>Total population within census tract</td>
<td>Cont.</td>
</tr>
<tr>
<td>pop_dens</td>
<td>Population density within census tract (persons per acre)</td>
<td>Cont.</td>
</tr>
<tr>
<td>person_per_h</td>
<td>Persons per household in census tract</td>
<td>Cont.</td>
</tr>
<tr>
<td>perc_rur_pop</td>
<td>Percent rural population within census tract</td>
<td>Cont.</td>
</tr>
<tr>
<td>veh_avail</td>
<td>Vehicles available within census tract</td>
<td>Cont.</td>
</tr>
<tr>
<td>per_cap_inc</td>
<td>Per Capita Income within census tract ($ )</td>
<td>Cont.</td>
</tr>
<tr>
<td>ag_p</td>
<td>Percent of civilian labor force within census tract employed in agriculture, forestry, mining</td>
<td>Cont.</td>
</tr>
<tr>
<td>const_p</td>
<td>Percent of civilian labor force within census tract employed in construction industry</td>
<td>Cont.</td>
</tr>
<tr>
<td>manu_p</td>
<td>Percent of civilian labor force within census tract employed in manufacturing industry</td>
<td>Cont.</td>
</tr>
<tr>
<td>wholes_p</td>
<td>Percent of civilian labor force within census tract employed in wholesale related industry</td>
<td>Cont.</td>
</tr>
<tr>
<td>retail_p</td>
<td>Percent of civilian labor force within census tract employed in retail related industry</td>
<td>Cont.</td>
</tr>
<tr>
<td>transport_p</td>
<td>Percent of civilian labor force within census tract employed in transportation related industry</td>
<td>Cont.</td>
</tr>
<tr>
<td>info_p</td>
<td>Percent of civilian labor force within census tract employed in information related industry</td>
<td>Cont.</td>
</tr>
<tr>
<td>finance_p</td>
<td>Percent of civilian labor force within census tract employed in finance related industry</td>
<td>Cont.</td>
</tr>
<tr>
<td>pro_p</td>
<td>Percent of civilian labor force within census tract employed in a professional field</td>
<td>Cont.</td>
</tr>
<tr>
<td>educ_p</td>
<td>Percent of civilian labor force within census tract employed in education</td>
<td>Cont.</td>
</tr>
<tr>
<td>arts_p</td>
<td>Percent of civilian labor force within census tract employed in arts and related industry</td>
<td>Cont.</td>
</tr>
<tr>
<td>other_p</td>
<td>Percent of civilian labor force within census tract employed in other industries</td>
<td>Cont.</td>
</tr>
<tr>
<td>pub_admin_p</td>
<td>Percent of civilian labor force within census tract employed in Public Administration</td>
<td>Cont.</td>
</tr>
</tbody>
</table>
APPENDIX B: Mallow’s Model Selection Plots

Figure B 1: Full sample plot of variables and Cp values for model selection.

Figure B 2: North Carolina sample plot of variables and Cp values for model selection.
Figure B 3: Ohio sample plot of variables and Cp values for model selection.

Figure B 4: Washington sample plot of variables and Cp values for model selection.
Figure B 5: 3,000 AADT and greater sample plot of variables and Cp values for model selection.
APPENDIX C: AADT Distribution Graphics by State

Figure C 1: ECDF of Washington Sample
Figure C 2: ECDF of Ohio Sample

Figure C 3: ECDF for North Carolina Sample
APPENDIX D: Boxplots of AADT by Functional Class and Buffer

Boxplots of AADT by Major Collectors within One Mile

Boxplots of AADT by Major Collectors within 2.5 Miles
Boxplots of AADT by Minor Arterials within Five Miles

Boxplots of AADT by Minor Arterials within Ten Miles
Boxplots of AADT by Princ. Arterials within Five Miles

- NC-No Princ. Art.: n=16
- NC-With Princ. Art.: n=141
- OH-No Princ. Art.: n=14
- OH-With Princ. Art.: n=130
- WA-No Princ. Art.: n=18
- WA-With Princ. Art.: n=156
- Sample-No Princ. Art.: n=47
- Sample-With Princ. Art.: n=427

Annual Average Daily Traffic

Boxplots of AADT by Princ. Arterials within Ten Miles

- NC-No Princ. Art.: n=4
- NC-With Princ. Art.: n=152
- OH-No Princ. Art.: n=1
- OH-With Princ. Art.: n=143
- WA-No Princ. Art.: n=10
- WA-With Princ. Art.: n=164
- Sample-No Princ. Art.: n=15
- Sample-With Princ. Art.: n=459

Annual Average Daily Traffic
Figure E 1: DFBETAS Index Plots for UMinA, RMajA, Minor-Local, and Pop. Density for Sample data.
Figure E 2: DFBETAS for the remaining six independent variables in the best-fit sample model.