A METHOD TO ESTIMATE ANNUAL AVERAGE DAILY TRAFFIC FOR MINOR FACILITIES FOR MAP-21 REPORTING AND STATEWIDE SAFETY ANALYSIS

SPR-804

Oregon Department of Transportation

A METHOD TO ESTIMATE ANNUAL AVERAGE DAILY TRAFFIC FOR MINOR FACILITIES FOR MAP-21 REPORTING AND STATEWIDE SAFETY ANALYSIS

Final Report

SPR-804

by

Avinash Unnikrishnan, Associate Professor
Miguel Figliozzi, Professor
Mostafa Khademi Moughari, Graduate Research Assistant
Sara Urbina, Graduate Research Assistant

Portland State University
1930 SW 4th Avenue, Portland OR 97201

for

Oregon Department of Transportation
Research Section
555 13th Street NE
Salem OR 97301

and

Federal Highway Administration
400 Seventh Street S.W.
Washington, DC 20590

June 2018

Technical Report Documentation Page

1. Report No. FHWA-OR-RD-18-17	2. Government Accession No.	3. Recipient's Catalog No.
4. Title and Subtitle A Method to Estimate Annual Average Daily Traffic for Minor Facilities for MAP-21 Reporting and Statewide Safety Analysis		5. Report Date June 2018 6. Performing Organization Code
7. Author(s) A. Unnikrishnan, M. Figliozzi, M. K. Moughari and S. Urbina		8. Performing Organization Report No.
9. Performing Organization Name and Address Portland State University 1930 SW 4th Avenue, Portland OR 97201		10. Work Unit No. (TRAIS) 11. Contract or Grant No. SPR-804
12. Sponsoring Agency Name and Address Oregon Dept. of Transportation Research Section and Federal Highway Admin. 555 13th Street NE, Suite 1 — 400 Seventh Street, SW Salem, OR 97301 — Washington, DC 20590-0003		13. Type of Report and Period Covered Final Report 14. Sponsoring Agency Code

15. Supplementary Notes

16. Abstract: This research project develops a simple and reliable method to estimate AADT on non-state roadway segments. Two separate analysis was conducted - non-state upper functional classification roadway segments with AADT less than 10000 and local roads. As the AADT varied based on location, we categorized region 2 non-state upper functional classification roads into four sub-regions and developed default values for each functional classification. For local roads, we determined default values based on sub-regions and the presence of Google Street View. Local roads without Google Street View had lower ADT than local roads with Google Street View. These default values can be used to quickly predict AADT if no other information is available. The analysis of the models developed in the literature review revealed that roadway and geometric information is more important than land use and socio-demographic information in predicting missing AADT. In this research, a simple point based model was developed to predict AADT based on the roadway, geometric, and signage information. We developed a stratified random sampling procedure to select roadway segments while ensuring all sub-regions and functional classification was adequately represented. The relevant variables were collected using Google Street View. An overall region 2 as well as separate sub-regional models were developed. The prediction accuracy of the models was tested on separate validation data. For the non-state upper functional classification roadway system model, the model errors were found to be reasonable on roadway segments with an AADT less than 5000. The sub-regional models provided lower median errors for the coast and valley-rural sub-regions. For local roads, the overall model had a median error of -32 which indicates that the model slightly under-predicts the ADT. The overall model has the lowest median error of 4 for the valley-rural sub-region. The gains in accuracy by using the sub-region model are not high.

17. Key Words AADT, Performance Measures, Models, Functional Classification, Planning		18. Distribution Statement Copies available from NTIS, and online at: https://www.oregon.gov/ODOT/Programs/Pages/Research-Publications.aspx	
19. Security Classification (of this report)	20. Security Classification (of this page)	21. No. of Pages 158	22. Price

Technical Report Form DOT F 1700.7 (8-72) Reproduction of completed page authorized Printed on recycled paper

SI* (MODERN METRIC) CONVERSION FACTORS

APPROXIMATE CONVERSIONS TO SI UNITS

Symbol	When You Know	Multiply By	To Find	Symbol
		LENGTH		
in	inches	25.4	millimeters	mm
ft	feet	0.305	meters	m
yd	yards	0.914	meters	m
mi	miles	1.61	kilometers	km
		AREA		
in^2	square inches	645.2	millimeters squared	mm^2
ft^2	square feet	0.093	meters squared	m^2
yd^2	square yards	0.836	meters squared	m^2
ac	acres	0.405	hectares	ha
mi^2	square miles	2.59	kilometers squared	km^2
		VOLUME		
fl oz	fluid ounces	29.57	milliliters	ml
gal	gallons	3.785	liters	L
ft^3	cubic feet	0.028	meters cubed	m^3
yd^3	cubic yards	0.765	meters cubed	m^3
NOTE: Volumes greater than 1000 L shall be shown in m^3.				
		MASS		
oz	ounces	28.35	grams	g
lb	pounds	0.454	kilograms	kg
T	short tons (2000 lb)	0.907	megagrams	Mg
		TEMPERATURE (exact)		
°F	Fahrenheit	(F-32)/1.8	Celsius	°C

APPROXIMATE CONVERSIONS FROM SI UNITS

Symbol	When You Know	Multiply By	To Find	Symbol
		LENGTH		
mm	millimeters	0.039	inches	in
m	meters	3.28	feet	ft
m	meters	1.09	yards	yd
km	kilometers	0.621	miles	mi
		AREA		
mm^2	millimeters squared	0.0016	square inches	in^2
m^2	meters squared	10.764	square feet	ft^2
m^2	meters squared	1.196	square yards	yd^2
ha	hectares	2.47	acres	ac
km^2	kilometers squared	0.386	square miles	mi^2
		VOLUME		
ml	milliliters	0.034	fluid ounces	fl oz
L	liters	0.264	gallons	gal
m^3	meters cubed	35.315	cubic feet	ft^3
m^3	meters cubed	1.308	cubic yards	yd^3
		MASS		
g	grams	0.035	ounces	oz
kg	kilograms	2.205	pounds	lb
Mg	megagrams	1.102	short tons (2000 lb)	T
		TEMPERATURE (exact)		
°C	Celsius	1.8C+32	Fahrenheit	°F

*SI is the symbol for the International System of Measurement

ACKNOWLEDGEMENTS

The authors would like to thank the members of the ODOT Research Section for their sage advice and assistance in the preparation of this report.

DISCLAIMER

This document is disseminated under the sponsorship of the Oregon Department of Transportation and the United States Department of Transportation in the interest of information exchange. The State of Oregon and the United States Government assume no liability of its contents or use thereof.

The contents of this report reflect the views of the author(s) who are solely responsible for the facts and accuracy of the data presented herein. The contents do not necessarily reflect the official policies of the Oregon Department of Transportation or the United States Department of Transportation.

The State of Oregon and the United States Government do not endorse products of manufacturers. Trademarks or manufacturers' names appear herein only because they are considered essential to the object of this document.

This report does not constitute a standard, specification, or regulation.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

1.0 INTRODUCTION

This report documents and puts together all the findings of this research including literature review, determining existing data sources in Oregon, developing a method to predict missing AADT information in Oregon, conducting a pilot demonstration and implement the methodology in ODOT Region 2, and recommendations on future data need to improve AADT estimation accuracy.

This research will develop and recommend practical methods to estimate AADT for non-state roads to meet MAP-21 and FAST reporting requirements and a statewide evaluation for safety. The objectives of this project are to: (i) identify data needs for estimating missing AADT information, (ii) develop a cost-effective method to estimate AADT for non-state roadway segments with AADT less than 10000 across the state of Oregon, (iii) conduct a pilot demonstration of the method to estimate missing AADT information in ODOT Region 2, and (iv) analyze the performance of the method and provide recommendations on future research directions and potential data collection efforts.

There are two key concepts that will utilized throughout the report. To avoid confusions these concepts are defined below:

- Missing *in-network* AADT: in this case some AADT values are missing but there have been counts performed in the past, for example there is no AADT estimation for the years 2014 and 2015 but there is count data from year 2013 or there is a count in a nearby section of the same highway AADT

- Missing *out-of-network* AADT: there are no records with past counts at a specific location or at nearby links; secondary data sources are necessary. This is the most challenging case and the main focus of this research project.

The words *missing AADT,* without clarification regarding in- or out-of-network, apply to both cases.

The final report is organized as follows. Chapter 2 presents a detailed literature review on the various methods used for predicting AADT. In chapter 3, we present the results of the survey of the best practices followed by other state department of transportation agencies. Chapter 4 describes the existing data sources available in Oregon. Chapter 5 and 6 present a point based model for predicting AADT on non-state upper functional classification and local streets respectively. Chapter 7 describes the results of the validation. Chapter 8 evaluates the data needs for the model. Chapter 9 summarizes the conclusions of this research.

2.0 LITERATURE REVIEW

2.1 REGRESSION ANALYSIS

Regression analysis is a statistical procedure to study the relationship between a dependent variable and multiple independent variables. The commonly used regression techniques for estimating AADT are multiple linear regression, nonlinear regression, and geographically weighted regression techniques.

2.1.1 Multiple Linear Regression

Multiple linear regression techniques assume that the AADT or the logarithm of AADT can be predicted as a linear function of several independent land use, socio-economic and demographic variables. The models found in the literature review had either a statewide scope of application or were more focused on a smaller county or urban level.

2.1.1.1 County level/Urban Area Models:

Xia et al. (1998) developed a multiple linear regression model to predict AADT on non-state roads in urban (population over 1 million) Broward County in Florida. Data was collected on roadway, socio-economic, and accessibility characteristics. The following independent variables were used with respect to roadway characteristics: number of lanes (L), land use type (AREA1 = 1 for rural, 2 for central business district and fringe area, 3 for residential area, and 4 for outlying business district), and functional classification (FCLASS1 = 0 for local and unclassified, 1 for city and county collector, 2 state and county minor arterial). Data was collected for the following socio-economic variables within a buffer zone of 0.25 mile from the count location – population, dwelling units, automobile ownership (AUTO), industrial employment, commercial employment, service employment (SEREMP), total employment, school enrollment, and hotel occupancy. Two indicator variables were used to characterize easy access to state roads and easy access to non-state roads (ACCESS2). A buffer zone of 1.6 km from the count location was used and the presence of state and non-state roads were examined. Data from 399 count locations was used to calibrate the model. The final model had the following six independent variables with $R^2$of 0.6302:

$$\boldsymbol{AADT = -10759 + 4737.44L + 5071.13FLCASS1 + 1274.17AREA1 + 0.15AUTO - 816.21ACCESS2 + 0.15SEREMP}$$

(2-1)

The model was validated using 40 additional count location data points. The percentage difference between observed and predicted AADT ranged from 1.31% to 57% with an average difference of 22.7%. Roadway characteristics such as number of lanes and functional classification helped better predict the AADT variation than socioeconomic variables.

Zhao and Chung (2001) further enhanced *Xia et al. (1998)* model to predict AADT for both state and non-state roads in Broward County. Four type of predictors were used. The following independent variables were used for roadway characteristics: number of lanes (L), land use type (central business district, central business district fringe, residential, outlying business district, rural area, undefined), and functional classification (FCLASS = 0.6 for unclassified, 1.0 for urban collector, 2.2 for urban minor arterial, and 3.4 for urban principal arterial). Data on total aggregated employment (BUFFEMP), population, and dwelling units were determined for variable sized buffer around counting stations. The buffer sizes varied depending on the average spacing of roadways as well as their functional classification in the subareas containing counting stations. Data on employment (EMPBUF) and population (POPBUF) were also aggregated based on variable sized buffers which were determined based on functional classification only. An indicator variable was used to capture accessibility to expressways (DIRECTAC) which takes value of 1 if the count station is located on a road which connects to expressway and 0 otherwise. Other variables characterizing accessibility such as minimum distance and travel time to expressway access point and number of expressway access points within a 4 mile radius were found to be insignificant. An index for accessibility of count station k to employment centers ($REACCESS_k$) was defined as:

$$\boldsymbol{REACCESS_k = \sum_{j=1}^{N_E} E_j e^{-0.054 t_{kj}}} \tag{2-2}$$

Where N_E, is the number of employment centers, E_j is the employment of the j^{th} employment center, and t_{kj} is the travel time from count station k to the j^{th} employment center. The employment centers were determined by visualizing employment distribution and density maps. An index for distance to mean centers of population (DPCNTR) was also estimated. The mean center of population was defined as the population weighted spatial mean center of all traffic analysis zones (TAZ). Other regional accessibility measures indicating distance to regional mean centers of employment and joint regional accessibility to population and employment centers were found to be insignificant. Regression models were calibrated based on 816 data points. AADT values for model calibration and validation were estimated from quarterly traffic counts in 1998. Four linear regression models were calibrated of which the following two – Model 1 ($R^2 = 0.81$) and Model 3 ($R^2 = 0.7624$) were selected as having the best prediction ability:

$$\begin{aligned}\boldsymbol{AADT} = {} & \boldsymbol{-9.520386 + 8.480001FCLASS + 3.428939LANE + 0.596752REACCESS} \\ & \boldsymbol{+ 2.991573DIRECTAC + 0.069086BUFFEMP}\end{aligned} \tag{2-3}$$

$$AADT = -4.66034 + 4.95341LANE + 0.5119REAccess + 4.52713DIRECTAC - 0.10689DPCNTR + 0.00112POPBUF \quad \textbf{(2-4)}$$

The models were validated using 82 additional data points. Model 1 had the lowest mean square error (MSE) of 50.00. In comparison, Model 3 had a MSE of 56.20. The percentage difference between predicted and observed AADT varied from 0.3% to 155.6% for Model 1 and from 0% to 288.1% for Model 3. For model 1, close to 90% of the errors were lower than 50%. In comparison for model 3 close to 83% of the errors were lower than 50% reflecting a minor improvement in accuracy. Similar to *Xia et al. (1999)*, roadway characteristic variables such as functional class and number of lanes were found to be more important than socio-economic variables.

Anderson et al. (2006) developed a multiple linear regression model to predict AADT in a small urban community of Anniston, Alabama. The independent variables selected in the model are: (i) functional classification (FCLASS), (ii) number of lanes (NUMLANES), (iii) population within a 0.5 mile buffer around a traffic count station (POPBUFF), (iv) employment within a 0.5 mile buffer around a traffic count station (EMPBUFF), and (v) variable indicating if a roadway section is a through or destination street (TVIFLOW). The model was calibrated using 2004 data from 58 roadway segments and is shown below ($R^2 = 0.801$):

$$ADT = -11189 + 4328FCLASS + 4360NUMLANES - .021POPBUFF + 0.17EMPBUFF + 2666TVIFLOW \quad \textbf{(2-5)}$$

The model prediction was validated using 38 additional data points. No statistically significant differences were found between the model predictions and the observed counts.

Yang and Wang (2014) develop a multiple linear regression model using the smooth clipped absolute deviation (SCAD) procedure to estimate the coefficients as well as select significant variables in one step. Data was assembled for four categories of input variables – driving behavior, roadway characteristics, satellite data, and socio-economic variables. With respect to driving behavior, the initial data considered include loading factor or the contribution of each household to roadway sections. The following variables was considered for roadway characteristics – number of lanes (LANES), length, connectivity to local roads, connectivity to high-level roads, AADT of nearest collector, and location of road. From Google map images, data on number of cars on the road (CARS) and the car intensity - number of cars per unit length (CARINTENSITY) was extracted. The following socio-economic variables were assembled at the zip code level – population, population density, housing units (HOUSINGUNITS), land area, water area, median income (INCOME), percentage of unemployed, and percentage of people below poverty line (BELOWPOVLINE). AADT and relevant explanatory variable data was assembled for 243 count locations in Mecklenburg County, NC in 2007 for local functional class roads. All dependent and independent variables were standardized by subtracting the mean and dividing by the standard deviation. Two hundred out of the 243 data points was used for model calibration. The final model ($R^2 = 0.6594$) obtained was:

$$\boldsymbol{AADT = 0.2959CARS + 0.3217LANES + 0.1200HOUSINGUNITS + 0.2765INCOME + 0.2648BELOWPOVLINE + 0.2729CARINTENSITY} \tag{2-6}$$

The remaining 43 data points was used to validate the model. The above model obtained from SCAD variable selection procedure was found to outperform the multiple linear regression model obtained from forward stepwise regression procedure. The final model was also compared to the spatially weighted regression model of *Zhao and Park (2004)* without including employment and access related variables. In the SCAD based model, nearly half of the errors defined as absolute percentage difference between true and predicted value was found to be less than 37%. The corresponding performance metric for *Zhao and Park (2004)* model was 47%. However one should note that *Yang and Wang (2014)* do not include three of the five predictors in *Zhao and Park (2014)* model during performance comparison.

2.1.1.2 Statewide Models:

Dadang et al. (1998) used multiple linear regression to predict AADT on county roads in Indiana. The following independent variables were selected initially - population (CPOP), households, vehicle registration, employment, per capita income, state highway mileage, arterial mileage (ART), and collector mileage aggregated at the county level, and indicator variables for location (LOCALE =1 for urban and 0 for rural), presence of interstate highways, and accessibility to state highway (ACCESS = 1 for easy access or close to state highway and 0 otherwise). Coverage count stations were located on county roads in 40 counties from February to August of 1996 to collect 48 hour traffic volume data. Four out of the eleven variables were found to significant in the final model. The independent variables in the final model were centered to reduce multi-collinearity. The final model selected had a $R^2$of 0.7726 and is shown below:

$$\boldsymbol{Log_{10}(AADT) = 4.82 + 0.82LOCALE + 0.84ACCESS + 0.24CPOP - 0.46Log_{10}(ART)} \tag{2-7}$$

Traffic volume data was collected from eight additional counties to validate the model. Two performance metrics were used for model validation – mean squared prediction error (MSPR) and percentage difference of observed and predicted AADT. The MSPR which is based on average squared difference of the observed and predicted logarithm of AADT for the final model was found to be 0.0510. The average percentage difference of observed and predicted AADT was found to be 16.78. The minimum and maximum percentage difference ranged from 1.56% to 34.18% across 18 validation observations.

Pan (2008) developed several regression based and manual methods for estimating AADT on roadway segments in Florida depending on roadway types and county characteristics. The AADT of Type 1 street segments comprising of all freeway and major state highways which have at least one traffic count per county were estimated from the closest traffic count location. The counties in Florida were classified into three groups based on 2005 population into large metropolitan (population > 400000), small-medium urban area (population between 100000 and 400000), and rural area (population < 100000). For each county category, separate multiple regression models were developed for state county highways and local streets. The following roadway characteristics were considered in model development - number of lanes in both directions (NUMBERLANE) and indicator variables for median type (DIVIDED), location (LOCATION - urban vs rural), land use (public-semipublic (SEMIPUBLIC), agriculture (AGRICULTURE), commercial (COMMERCIAL), institutional (INSTITUTIONAL), residential (RESIDENTIAL), recreation (RECREATION), industrial (INDUSTRIAL), and others), and accessibility to freeways defined as presence of freeways within buffers of 0.5(0.5MILE), 1.0, and 1.5(1.5MILE) miles. Seven different county level socio-economic variables were considered initially including population (POPULATION), total lane mileage of highways (MILEAGE), vehicle registration (VEHICLE), personal income (INCOME), retail sales (SALES), population within incorporated areas (MUNICIPALITIES), and labor force (LABORFORCE). The models were calibrated using 26721 traffic counts provided by Florida DOT. The final models obtained were:

Large Metropolitan Area, State/County Highway Model ($R^2 = 0.186$)

$$\begin{aligned} \boldsymbol{AADT} = & -\mathbf{848.8} + \mathbf{13.541}(\boldsymbol{VEHICLE}) + \mathbf{1273.347}(\boldsymbol{DIVIDED}) \\ & + \mathbf{2983.4}(\boldsymbol{COMMERCIAL}) + \mathbf{6259.6}(\boldsymbol{LOCATION}) \\ & - \mathbf{8.845}(\boldsymbol{LABORFORCE}) - \mathbf{2839.18}(\boldsymbol{AGRICULTURE}) \\ & + \mathbf{421.252}(\boldsymbol{NUMBERLANE}) + \mathbf{1311.231}(\boldsymbol{INSTITUTIONAL}) \\ & + \mathbf{129.069}(\boldsymbol{INCOME}) + \mathbf{769.6}(\mathbf{0.5}\boldsymbol{MILE}) - \mathbf{782.6}(\boldsymbol{RESIDENTIAL}) \\ & - \mathbf{585.4}(\boldsymbol{SEMIPUBLIC}) \end{aligned} \qquad \textbf{(2-8)}$$

Large Metropolitan Area, Local Street Model ($R^2 = 0.244$)

$$\begin{aligned} \mathbf{AADT} = & -\mathbf{2738.448} + \mathbf{3.806}(\mathbf{MUNICIPALITIES}) + \mathbf{1349.659}(\mathbf{DIVIDED}) \\ & - \mathbf{452.459}(\mathbf{RESIDENTIAL}) - \mathbf{562.182}(\mathbf{1.5MILE}) \\ & + \mathbf{2745.195}(\mathbf{LOCATION}) + \mathbf{259.492}(\mathbf{NUMBERLANE}) \\ & + \mathbf{1040.226}(\mathbf{SEMIPUBLIC}) + \mathbf{769.194}(\mathbf{COMMERCIAL}) \\ & - \mathbf{19.545}(\mathbf{LABORFORCE}) + \mathbf{17.369}(\mathbf{POPULATION}) - \mathbf{4.345}(\mathbf{VEHICLE}) \end{aligned} \qquad \textbf{(2-9)}$$

Small-Medium Urban Area, State/County Highway Model ($\mathbf{R^2 = 0.261}$)

$$\boldsymbol{AADT = 770.374 + 5566.145(LOCATION) + 122.079(LABORFORCE) + 2760.767(COMMERCIAL) + 960.82(NUMBERLANE) + 27.673(VEHICLE) - 70.869(POPULATION) + 0.994(SALES) - 13.311(MUNICIPALITIES) + 952.963(1.5MILE) - 431.282(RESIDENTIAL) + 765.103(SEMIPUBLIC) - 0.43(MILEAGE) + 1072.666INDUSTRIAL} \quad \textbf{(2-10)}$$

Small-Medium Urban Area, Local Street Model ($\mathbf{R^2 = 0.172}$)

$$\mathbf{AADT = 1533.94 + 2482.69(DIVIDED) - 679.405(RESIDENTIAL) + 2107.874(1.5MILE) + 2707.119(LOCATION) + 18.468(VEHICLE) - 14.468(POPULATION) + 0.9437(MUNICIPALITIES) + 3320.91(INDUSTRIAL) + 1491.556(COMMERCIAL) + 1464.231(INSTITUTIONAL) + 2011.814(RECREATION)} \quad \textbf{(2-11)}$$

Rural Area, State/County Highway Model ($\mathbf{R^2 = 0.382}$)

$$\boldsymbol{AADT = 3015.747 + 3878.551(LOCATION) + 17.722(VEHICLE) + 57.071(MUNICIPALITIES) - 1656.733(AGRICULTURE) + 22.293(LABORFORCE) - 1.931(SALES) - 3312.919\ (RECREATION)2324.493(INDUSTRIAL) + 33.239(POPULATION) - 747.708(RESIDENTIAL)} \quad \textbf{(2-12)}$$

Rural Area, Local Street Model ($\mathbf{R^2 = 0.432}$)

$$\boldsymbol{AADT = 1225.505 + 62.168(POPULATION) + 1458.501(LOCATION) - 1445.085(AGRICULTURE) - 1017.873(RESIDENTIAL)} \quad \textbf{(2-13)}$$

While the R^2 obtained were lower than previous research, the models scope of application was significantly higher with focus on predicting AADT from a statewide perspective. The models were validated using 1149 traffic counts from three randomly selected counties. The average mean absolute percentage error was found to vary from 31.99% to 159.49%. The lowest prediction errors were found for the rural county model.

Seaver et al. (2000) developed a multiple regression cluster models to estimate Average Daily Traffic (ADT) in a county on rural roads based on pavement condition and Metropolitan Statistical Area (MSA) status. The models were calibrated using data from 80 out of 159 counties in Georgia. The model used 45 independent variables which was reduced using principal component analysis. Multiple regression based clustering was performed to arrive at the final models. The independent variables used in the three models for rural paved roads located outside the MSA include percent population change from 1990 to 1996, median travel time, number of agricultural farms, percentage of farms with 500+ acres, median household income, median time to leave for work, and distance to MSA. The corresponding independent variables used for two cluster models developed for rural paved roads within MSA include population density per square mile, unemployment rate, median travel time, percentage of farms with 500+ acres, median time to leave for work, and number of persons working outside of county. The independent variables used for two cluster models for ADT estimation in rural unpaved roads outside MSA include population density per square mile, per capita income, median travel time, unemployment rate, and median time to leave for work, number of persons working outside of the county, and the distance to the MSA. The models show a high R^2 ranging from 0.80 to 0.94 for all cases considered except one. Note that the goal of this research is to predict one ADT for each roadway type for each county. The model does not predict ADT for all roadway segments.

Dixon et al. (2011) developed two multiple linear regression models to predict AADT on minor legs of rural intersections. The dependent variable used was the logarithm to the base 10 of AADT. The two models developed along with the significant variables are provided below:

Model 1 ($\mathrm{R}^2 = 0.6231$)

$$\begin{aligned}\mathbf{log_{10}}(\boldsymbol{AADT}) = \mathbf{2.0281} + \mathbf{0.6810}\boldsymbol{MIA} + \mathbf{0.4148}\boldsymbol{MAC} + \mathbf{0.1761}\boldsymbol{RIGHT} \\ + \mathbf{0.2060}\boldsymbol{RIGHTCROSS} + \mathbf{0.2125}\boldsymbol{LANDUSE} + \mathbf{0.3028}\boldsymbol{CENTERLINE} \\ + \mathbf{0.1268}\boldsymbol{EDGELINE}\end{aligned} \tag{2-14}$$

Model 2 ($\mathrm{R}^2 = 0.6395$)

$$\begin{aligned}\mathbf{log_{10}}(\boldsymbol{AADT}) = \mathbf{2.0246} + \mathbf{0.6634}\boldsymbol{MIA} + \mathbf{0.4132}\boldsymbol{MAC} + \mathbf{0.2073}\boldsymbol{RIGHTCROSS} \\ + \mathbf{0.2229}\boldsymbol{LANDUSE} + \mathbf{0.2988}\boldsymbol{CENTERLINE} + \mathbf{0.1381}\boldsymbol{EDGELINE}\end{aligned} \tag{2-15}$$

The independent variables used in the final regression models are: indicator variable for if the cross street is a minor arterial (MIA), if the cross street is a major collector (MAC), presence of right-turn lane on minor road (RIGHT), presence of right turn lane on major road (RIGHTCROSS), development in adjacent land (LANDUSE), presence of centerline on minor road (CENTERLINE), and the presence of striped edge line on minor roads (EDGELINE). The above models were used to predict AADT to be used in safety performance functions and the validation information was not available.

Barnet et al. (2015) developed a multiple linear regression model to predict AADT on minor legs of stop controlled intersections based on land use, census, roadway, and network data. Data was collected from 474 locations in Oregon, Washington, and North Carolina. The dependent variable was the natural logarithm of the AADT. The following independent variables were found to be significant in the final three state model: indicator variable for urban minor arterial (UMART), local street (LSTREET), collector

(COLLECT), population density (POPDEN), does not connect to city/town within two miles (NOT2CITY), number of principal arterials within one mile (NUMPART), access to parking lot (PLOTACC), presence of edge striping (PEDGESTRIP), and Agricultural Employment (AGEMP). The final three state model with $R^2 = 0.339$ is shown below:

$$\begin{aligned}\mathbf{ln}(\boldsymbol{AADT}) = \mathbf{7.879 + 0.232}\boldsymbol{UMART} - \mathbf{1.164}\boldsymbol{LSTREET} \\ - \mathbf{0.523}\boldsymbol{COLLECT} - \mathbf{.048}\boldsymbol{POPDEN} - \mathbf{0.478}\boldsymbol{NOT2CITY} \\ + \mathbf{0.154}\boldsymbol{NUMPART} + \mathbf{0.251}\boldsymbol{PLOTACC} + \mathbf{0.604}\boldsymbol{PEDGESTRIP} \\ - \mathbf{4.310}\boldsymbol{AGEMP}\end{aligned} \tag{2-16}$$

Developing a separate model for minor legs with AADT of greater than equal to 3000 was found to increase the R^2 to 0.87. The R^2 was also found to increase when separate models were developed for Ohio, Washington, and North Carolina.

Ohio (R^2= 0.476)

$$\begin{aligned}\mathbf{ln}(\boldsymbol{AADT}) = \mathbf{3.725 + 0.885}\boldsymbol{UPART} + \mathbf{0.952}\boldsymbol{UMART} + \mathbf{0.00008}\boldsymbol{PCINC} \\ - \mathbf{0.459}\boldsymbol{NUMFW} + \mathbf{0.107}\boldsymbol{NUMMC} + \mathbf{0.607}\boldsymbol{CSTRIP} + \mathbf{7.312}\boldsymbol{AGEMP} \\ + \mathbf{8.447}\boldsymbol{TREMP}\end{aligned} \tag{2-17}$$

Washington (R^2= 0.496)

$$\begin{aligned}\mathbf{ln}(\boldsymbol{AADT}) = \mathbf{7.908 - 1.428}\boldsymbol{MLSTREET} - \mathbf{0.723}\boldsymbol{MCOLLECT} - \mathbf{0.425}\boldsymbol{NOT2CITY} \\ + \mathbf{0.635}\boldsymbol{PEDGESTRIP} + \mathbf{0.871}\boldsymbol{LTLANE} - \mathbf{5.202}\boldsymbol{AGEMP}\end{aligned} \tag{2-18}$$

North Carolina (R^2= 0.362)

$$\begin{aligned}\mathbf{ln}(\boldsymbol{AADT}) = \mathbf{6.056 + 0.383}\boldsymbol{UMART} - \mathbf{0.109}\boldsymbol{POPDEN} + \mathbf{0.275}\boldsymbol{NUMFW} \\ + \mathbf{0.561}\boldsymbol{PLOTACC} + \mathbf{0.660}\boldsymbol{CSTRIP}\end{aligned} \tag{2-19}$$

The following independent variables were used for the state specific regression models: urban principal arterials (UPART), urban minor arterials (UMART), per capita income (PCINC), number of freeway within 2 miles (NUMFW), number of major collectors within 2 miles (NUMMC), presence of center striping (CSTRIP), transportation employment (TREMP), minor local street (MLSTREET), minor collector (MCOLLECT), and presence of left turn lane (LTLANE). The models were validated using data from 54 sites. The percentage error ranged from 1% to 214% with an average error of 59%.

2.1.2 Nonlinear Regression

Nonlinear regression techniques assume that the AADT or the logarithm of AADT can be predicted as a nonlinear function of several independent land use, socio-economic and demographic variables. *Staats (2016)* and *Souleyrette et al. (2016)* compared several regression approaches and developed a poisson regression approach to predict AADT on non-state roads in Kentucky. *Zhao and Chung (2001)* model was found to have high errors with an average error of 402% when applied to several counties in Kentucky. The following linear regression model tested only in Meade County was found to provide the best performance:

$$\boldsymbol{ADT = 565.93 + 6.99RESIDENTIAL + 6.73COMMERCIAL}$$

(2-20)

Where RESIDENTIAL and COMMERCIAL correspond to total number of residential and commercial properties assigned to each roadway section determined from 911 databases. Since the 911 database was not available at the state level, this approach was discontinued. Nonlinear poisson regression models were found to provide the lowest errors. The model was calibrated and validated using AADT count data from 2011 to 2013. The following independent variables were assembled: residential properties – non-commercial vehicle registration addresses was obtained from vehicle registration database and assigned to nearest roadway segments (RESIDENTIAL), commercial properties - commercial vehicle addresses was obtained from vehicle registration database and assigned to nearest roadway segments, probe counts - 2012 daily average probe count for each roadway section was obtained from HERE corporation (PROBE), and roadway curvature defined as the actual length of the roadway segment divided by straight line distance between the end points (CURVE). Two separate models were developed for urban and rural areas. Rural area model considered roadway segments with traffic counts between 20 and 1000 only whereas no restrictions were enforced on the urban model. For both urban and rural models, three separate models were calibrated for the West, North-Central, and East regions.

Rural West

$$\boldsymbol{AADT = e^{5.7696+0.0059PROBE-0.5210CURVE+0.0041RESIDENTIAL}}$$

(2-21)

Rural North Central

$$\boldsymbol{AADT = e^{5.2644+0.0058PROBE-0.0776CURVE+0.0055RESIDENTIAL}}$$

(2-22)

Rural East

$$\boldsymbol{AADT = e^{5.50546+0.0057PROBE-0.0151CURVE+0.0023RESIDENTIAL}}$$

(2-23)

Urban West

$$AADT = e^{6.4706+0.0065PROBE-0.1258CURVE+0.0029RESIDENTIAL}$$

(2-24)

Urban North Central

$$AADT = e^{5.8139+0.0112PROBE+0.21916CURVE+0.0115RESIDENTIAL}$$

(2-25)

Urban West

$$AADT = e^{7.0093+0.0073PROBE-0.0792CURVE+0.0002RESIDENTIAL}$$

(2-26)

The model AADT estimates were scaled using adjustment factors derived from ratios of county level Daily Vehicle Miles Traveled (DVMT) obtained from a different power regression model and the county level DVMT estimated from the above AADT models. The mean absolute percentage difference for rural areas were found to vary between 61% and 97% for the three geographic areas. The corresponding values for the urban models were significantly higher and ranged between 354% and 1956.

Multiple linear regression appears to be one of the most popular methods to predict missing AADT information. Potential reasons for the large number of studies applying multiple linear regression could be its ease of use, high degree of comfort of practitioners with the method, and proven to work in transportation engineering and planning in other contexts. One interesting insight common in several studies is that roadway characteristics are more useful than socio economic and land use characteristics in predicting AADT information. In addition, several models are misspecified and in many models there seems to be a tendency towards overfitting the data and no adequate validation approach. In the last decade researchers have started to develop more sophisticated models and also tried to transfer earlier models. It seems that some of the earlier models were not correctly specified and transfer poorly to other regions.

2.1.3 Spatial Regression

Spatial regression models account for dependencies and correlations between variables based on geographic locations. Spatial regression models are increasingly becoming popular in transportation applications due to their ability to better capture spatial variations. The two methods for spatial regression used for AADT estimation applications are geographically weighted regression and universal kriging.

Zhao and Park (2004) extended *Zhao and Chung (2001)* multiple regression model by adopting a spatially weighted multiple regression approach where the regression coefficients varied based on locations. During coefficient estimation weights were used to capture the spatial relationship between count locations. The same dataset as *Zhao and Chung (2001)* was used to calibrate and validate the model. The multiple linear regression model used as a baseline to compare the new geographic spatially weighted regression model was:

$$\boldsymbol{AADT = -5.59096 + 4.87520LANE + 4.16904DIRECTAC + 0.51856REACCESS + 0.00105POPBUF + 0.00023EMPBUF}$$

(2-27)

The variable definitions are provided earlier in the description for *Zhao and Chung (2001)* model. The multiple linear regression model was found to have a R^2 of 0.764. Two spatially weighted regression models were developed with the same predictors and using two weighting functions – bi-square and gaussian. The R^2 value for the bi-square and gaussian weighted regression models were found to be significantly higher at 0.8756 and 0.87 respectively. Similar to *Zhao and Chung (2001)* the model was validated using data from 82 count locations. Nearly 85% of the data points had errors less than 50% for the above ordinary linear regression model. The corresponding number for spatial regression models were 96.34% using the bi-square weighting function and 95.12% for the Gaussian weighting function. Overall the geographically weighted regression models outperformed the multiple linear regression model. The bi-square weighting function was recommended as a better choice over the gaussian weighting function.

Kriging is a popular geostatistics method originally used in the mining industry for predicting ore reserves. In Kriging application for predicting AADT, the AADT at a location s¸$s(Z(s))$ can be written as a function of a deterministic trend $\mu(s)$ and an error $\epsilon(s)$ as follows:

$$\boldsymbol{Z(s) = \mu(s) + \epsilon(s)}$$

(2-28)

The error terms are assumed to be spatially correlated. There are three different types of Kriging depending on the nature of the assumption in describing $\mu(s)$. In simple kriging, the trend is assumed to be a known constant. In ordinary kriging, the trend is assumed to be an unknown constant. In universal kriging, the trend is assumed to be a function of independent variables. A semivariogram function is used to capture the spatial correlations. The three commonly used functions in AADT estimation are exponential, spherical, and gaussian.

Eom et al. (2006) developed a universal kriging model to predict AADT in Wake County, North Carolina. Data was assembled for three types of independent variables. The first category of data was latitude and longitude of the points where AADT prediction is to be made. The second category of independent variables relates to roadway characteristics - area type (urban vs rural vs suburban), number of lanes, speed limit, functional classification (12 types), signal density per mile, and presence of median left-turn. The following sociodemographic data was assembled from 2000 census data and the regional planning model aggregated at the block level - total population, number of households, household size, number of households with a child under 6 years, median income, and workers living within and outside the boundary. The independent variable was chosen to be AADT to the power of 0.15. A total of 200 counts was used to calibrate the model and 954 counts was used to validate the model. First multiple linear regression model was developed to predict the average transformed AADT. The final model had an $R^2 = 0.6487$ and had the following variables - latitude, longitude, speed, median income, number of lanes, area type, and functional class 1,2,3, and 4. Next three type of semivariogram models – exponential, gaussian, and spherical – was fitted to study the spatial correlation in error terms using multiple calibration procedures. The spherical semivariogram fitted by weighted least squares and exponential semivariogram fitted by restricted maximum likelihood procedure was found to provide the best performance. The spatial regression models outperformed multiple linear regression for urban arterial AADT prediction. However, their prediction did not show any improvement over simple multiple linear regression models for collector roads.

Wang and Kockelman (2009) developed a two-step procedure to predict AADT on Texas highways. In the first step, linear interpolation was used to predict AADT at 27,738 sites in the year 2006 based on 1999 to 2005 AADT data. Next ordinary kriging was used to predict AADT values at missing sites. Eighty percent of the data was used for calibration and the remaining 20% was used for validation. The exponential semivariogram specification was found to provide the best fit. Different parameters of the exponential semivariogram were estimated for different functional roadway classes. The median error was found to be a 33% overestimation. The predictions were found to be better for locations with higher AADT (> 1000 vpd).

Selby and Kockelman (2013) compared the performance of universal kriging, and geographically weighted regression with non-spatial regression techniques for AADT prediction. The Box-Cox transformation was applied to the dependent variable AADT to stabilize the variation. The independent variables considered include – speed, number of lanes, functional class of roadway segments, county level population, and employment densities. Three types of accessibility indexes were tested for each census tract – distance to a given population, population within a given distance, and sum of inverse distance weighted populations. The sum of inverse distance weighted population within 50 miles was found to be the most appropriate accessibility index. The models were tested on 2005 statewide AADT information (>= 200 vpd only). There were 25183 samples which were broken down into several regional samples centered on urban areas and statewide datasets based on interstates only, urban only, and minor roads only. For the universal kriging, three different types of semivariogram functions – exponential, spherical, and gaussian were tested. The exponential function was found to provide the best predictions which is consistent with Wang and Kockelman. In spatial regression, distances can be calculated using simpler Euclidean distances or the more complex network distances. No additional benefits were observed in prediction performance by using the more complex network distances. The most significant independent variables were road type, speed limit, number of lanes, and population accessibility index. The spatial regression models had absolute error reductions between 16% and 63% compared to non-spatial regression. The universal kriging yielded lower absolute errors ranging from 3% to 8% compared to geographically weighted regressions.

Shamo et al. (2015) compared the three different kriging techniques in combination with five variogram models on AADT data sets from 2008 to 2010 in the state of Washington. The dependent variable chosen was the logarithm of AADT. The researchers found that the ideal combination of kriging techniques and semivariogram varied from dataset to dataset and using the same kriging technique variogram combination would result in a decrease in accuracy.

Zhao and Park (2004) and *Eom et al. (2006)* focus on a county level application whereas *Wang and Kockelman (2009)* and *Shamo et al. (2015)* models aim at predicting missing AADT at a statewide model. *Selby and Kockelman (2013)* provide an intermediate level model by calibrating separate models for different regions in the state and for different functional classifications.

Musunuru et al. (2017) proposed a method to estimate day and night time traffic volumes on rural, two-lane horizontal curve road segments in Utah, using Universal kriging method. The traffic count data at all 100 ATR stations in Utah from 2009 to 2013 was used to calibrate and validate the model. The dependent variable was the logarithm of the average daily traffic during day time and average daily traffic during night time. The following independent variables were used - functional classification of the road segments, number of lanes, population, and household unit counts at census block level at each ATR location. Four semivariograms models were tested: Exponential, Gaussian, Spherical and Matern M.stein's parameterization along with different covariates. A model using Matern M.stein's parameterization semivariogram and the covariates number of lanes, population or number of housing units and indicator variables for interstate and expressway/highway roads was the best fitted model for both day and night traffic prediction.

The model was validated using the K-fold cross validation procedure. Results of this process showed a positive correlation between the estimated and observed values which was near one (about 0.79). The mean square error of the selected model for day and night were 0.6674 and 0.8238 respectively. Later in the paper statistical road safety models were used with the estimated night and day traffic volumes and it was proved to work more efficiently.

The spatial regression models are increasing in popularity in recent times driven by the developments in spatial econometric approaches and computational power to calibrate these models in the last two decades. Spatial regression models hold an obvious advantage over other forms of regression in that they are more accurately able to capture the spatial relationships which are present in transportation and urban planning. However, they have more parameters to calibrate, rely on more assumptions compared to simple regression, and the transferability of the calibrated models is still an open research question.

2.2 TRAVEL DEMAND MODELING

Travel demand modeling based approaches mimic the four step travel forecasting process and estimate the missing AADTs as being equal to the volumes obtained from the traffic assignment step.

Zhong and Hanson (2009) used the four step travel demand modeling approach to estimate the missing AADT information for low-class roads in York county and Beresford regions in New Brunswick, Canada. The smallest census unit, known as dissemination area (DA) was used as the traffic analysis zone. The quick response method was used for trip generation, distribution, and demand balancing. Cross-classification method was used for generating trips in each zone based on total number of households and average household income in each zone. The total number of trips attracted to each zone was determined using a regression model bases on total number of households, retail and non-retail employment in each zone. The productions and attractions were balanced. Trip distribution was performed using a gravity model with a gamma function based on distances used for estimating impedances. The stochastic user equilibrium model was used for traffic assignment. In York County, the model predictions were validated against 26 traffic counts (9 on arterials, 6 on collectors, and 11 local highways). The errors were found to be low for arterials and collectors with significant amount of overestimation for local highways. The model predictions were further adjusted using the following equations:

$$Adjusted = \mathbf{0.4375} \times TDM\ OUTPUT + \mathbf{67.237}$$

(2-29)

The adjustment reduced the overall average error for local highways to 38.6%. The model was also applied to Beresford Census Consolidated Subdivision and the overall error was found to be 17%. Note that while the average error was lower, the number of traffic counts used for validation was also lower.

Wang et al. (2013) developed a four step travel demand modeling procedure using tax parcel level data to estimate AADT in local roads in Florida. The four step procedure involved: (i) Network Preparation: identifying the boundaries of the study area, collecting data on roadway network, traffic counts, and linking roadway sections to parcels. (ii) Trip Generation: determining number of trips generated in each parcel based on ITE trip generation rates and land use. (iii) Trip Distribution: distributing the trip generated in each parcel to nearby traffic count sites using the Gravity model with travel times as impedances. (iv) Traffic Assignment: assigning trips to routes using All or Nothing assignment. The model performance was compared with the regression model developed by *Pan (2008)* across 78 count stations in Broward County, Florida. The travel demand modeling approach significantly outperformed *Pan (2008)* regression models with an average mean absolute percentage error of 52% compared to 211% for the regression approach.

Travel demand modeling based approaches are more suitable for county level or compact urban areas compared to statewide applications. Travel demand model based AADT estimation approaches will need to be set up independently for each region where AADT is to be estimated. Ideally, this methodology is more suited for large scale urban regions where the AADT estimation procedure can leverage an existing validated travel forecasting model for planning purpose. This approach might not be suitable for rural areas where assembling the data to calibrate and validate a new travel forecasting approach will be more cumbersome.

2.3 GEOSPATIAL METHOD

Geospatial methods exploit network connectivity and topology based metrics in predicting missing AADT information. *Lowry and Dixon (2012)* develop a geographic information system (GIS) based tool to estimate missing AADT information in small sized communities. The tool embedded a linear regression model to predict missing AADT information. A unique aspect of this tool was the ability to embed network connectivity indices as an explanatory variable in the linear regression model. The recommended index was the connectivity importance index defined as the number of times a roadway segment is used in the shortest paths connecting various origins to destinations. While no detailed analysis on AADT prediction accuracy was reported, a linear regression model demonstration for Moscow Idaho using the functional classification, number of lanes, and connectivity importance index independent variables was reported to have a R^2 of 0.72. The GIS tool is flexible enough to accommodate other independent variables if available.

Lowry (2014) introduced a new explanatory variable Origin Destination (OD) Centrality as an explanatory variable in a regression framework to predict AADT in the small compact community of Moscow Idaho. The OD centrality of a link e is defined as:

$$\boldsymbol{OD\ Centrality_e} = \sum_{\boldsymbol{i \in I, j \in J}} \boldsymbol{\sigma_{ij}(e) M_i M_j}$$

(2-30)

Where $\sigma_{ij}(e) = 1$if link e lies on the shortest path connecting origin i and destination j, M_i and M_j are multipliers reflecting relative trip production and attraction potential which are estimated from ITE Trip Generation Manual. *Lowry (2014)* compared several regression models. The origins and destinations are either internal parcels or external boundary locations. The ordinary multiple linear regression with the following independent variables: internal-internal OD centrality (*x1*), internal-external OD centrality (*x2*), and external-external OD centrality (*x3*) and the Box-Cox transformed AADT (*y*) as the dependent variable was found to have the best performance with a R^2 of 0.93.

$$\boldsymbol{y = 413.30 + 183.45x1 + 105.77x2 + 1197.16x3}$$

(2-31)

Ninety percent of 341 AADT count data was used for calibration and the remaining was used for validation. The median absolute percent error in validation was found to be 22% which is similar to the performance produced by *Wang and Kockelman (2009)*'s spatial regression method. This method is particularly suited for small to medium sized compact communities.

Pulugurtha and Kusum (2012) developed a regression model to predict AADT based on roadway characteristics and land use and socio-demographic characteristics of multiple network distances based buffers of various distances around the location of interest. Circular concentric polygon based network buffers were created at distances of 1, 1.5, 2, 3, 4, and 5 miles respectively around each point of interest. Freeways, major thoroughfares, and minor thoroughfares were assumed to have a maximum accessible distance of 5, 3, and 2 miles respectively. Data on demographic, socio economic, area type, and land use related variables are aggregated for each concentric buffer with weights which decrease with distance. Several linear and nonlinear regression models were tested. Ninety percent of the 2005 AADT data for the city of Charlotte was used for calibration. Negative binomial regression models calibrated for each roadway functional class was found to yield the best predictions. The final models are:

Freeway:

$$\begin{aligned}\mathbf{log}(\boldsymbol{AADT}) = \mathbf{2.36} + \mathbf{0.133}(\boldsymbol{NUMLANES}) + \mathbf{0.018}(\boldsymbol{DLSL}) + (\mathbf{0.070}\boldsymbol{DCSNUMLANES}) \\ + \mathbf{0.029}(\boldsymbol{POP}) - \mathbf{0.036}(\boldsymbol{MANHOUSE})\end{aligned}$$

(2-32)

Major thoroughfare:

$$\begin{aligned}\mathbf{log}(\boldsymbol{AADT}) = \mathbf{1.538} + \mathbf{0.505}(\boldsymbol{URBAN}) + \mathbf{0.030}(\boldsymbol{SPEEDLIMIT}) + \mathbf{0.006}(\boldsymbol{ULSL}) \\ + \mathbf{0.053}(\boldsymbol{DCSNUMLANES}) - \mathbf{0.051}(\boldsymbol{POP}) - \mathbf{0.059}(\boldsymbol{MANHOUSE})\end{aligned}$$

(2-33)

Minor thoroughfare:

$$\mathbf{log}(\boldsymbol{AADT}) = \mathbf{1.86} + \mathbf{0.188}(\boldsymbol{URBAN}) + \mathbf{0.007}(\boldsymbol{SPEEDLIMIT}) - \mathbf{0.005}(\boldsymbol{RURAL})$$

(2-34)

where NUMLANES is the number of lanes, $URBAN$ is the indicator variable for urban land use type, $SPEEDLIMIT$ is the speed limit, $ULSL$ is the upstream link speed limit, $DLSL$ is the downstream link speed limit, $DCSNUMLANES$ is the number of lanes in downstream cross street, POP is the population, $MANHOUSE$ is the number of manufacturing houses, and $RURAL$ is the indicator for rural land use type. The absolute average percentage errors were found to be lower than 36% for all three models.

Kusam and Pulugurtha (2016) further modified *Pulugurtha and Kusum (2012)* by studying the impact of network buffers at various distances on AADT prediction. Similar to *Pulugurtha and Kusum (2012)*, negative binomial regression with separate models for three functional classes was found to yield the best predictions. For freeways, a network buffer of 2 miles was found to provide the best fit whereas 1.5 miles was found to be optimal for major and minor thoroughfares. The final models are:

Freeway:

$$\begin{aligned}\mathbf{log}(\boldsymbol{AADT}) = \mathbf{2.43} + \mathbf{0.2}(\boldsymbol{CBD}) + \mathbf{0.14}(\boldsymbol{NUMLANES}) + \mathbf{0.02}(\boldsymbol{DLSL}) \\ + (\mathbf{0.08}\boldsymbol{DCSNUMLANES}) - \mathbf{0.00006}(\boldsymbol{MOBRES}) \\ - \mathbf{0.01147}(\boldsymbol{MANHOUSE})\end{aligned}$$

(2-35)

Major thoroughfare:

$$\begin{aligned}\mathbf{log}(\boldsymbol{AADT}) = \mathbf{0.753} + \mathbf{0.509}(\boldsymbol{URBAN}) + \mathbf{0.048}(\boldsymbol{SPEEDLIMIT}) + \mathbf{0.000045}\boldsymbol{PUDEV} \\ + \mathbf{0.000384}(\boldsymbol{ROW})\end{aligned}$$

(2-36)

Minor thoroughfare:

$$\begin{aligned}\mathbf{log}(\boldsymbol{AADT}) = \mathbf{1.756} + \mathbf{0.204}(\boldsymbol{URBAN}) + \mathbf{0.006}(\boldsymbol{ULSL}) + \mathbf{0.004}(\boldsymbol{DLSL}) \\ - \mathbf{0.000027}(\boldsymbol{INSTITUTIONAL})\end{aligned}$$

(2-37)

where CBD is an indicator variable for the link being present in the central business district, $MOBRES$ denotes the area of manufactured homes and mobile parks in 1000 ft^2, $PUDEV$ refers to planned unit development area, ROW is the state owned right of way for all roadway segments, and $INSTITUTIONAL$ refers to area of major educational, medical, government, cultural, or religious institutions. The absolute average percentage errors were found to vary between 18% and 26% representing an improvement over *Pulugurtha and Kusum (2012)*.

Keehan et al. (2017) studied the use of stress and origin/destination centrality to determine AADT on all roads in a small sized city, Greenville, South Carolina. Data sources used in this study include functional classification, speed limit, number of lanes, land use, number of buildings in each parcel, area of each building, and count data. The OD centrality for each link e is calculated as:

$$\boldsymbol{OD\ CENTRALITY} = \sum_{i \in I} \sum_{j \in J} \boldsymbol{\sigma_{ij}(e) M_i M_j}$$

(2-38)

where I, J refer to the set of origins and destinations respectively, $\sigma_{ij}(e) = 1$ if link e is used in the shortest route connecting origin $i \in I$ and destination $j \in J$ and 0 otherwise, M_i and M_j are weights assigned to each zone based on weighted mean trips generated in each parcel in that zone. For the external centroid, the weight was determined based on the nearest AADT count station. Based on the origin or destinations being external points or internal centroids of the TAZs in the created network, three centrality methods were investigated (Internal-Internal, Internal-External or External-External to derive link significance variables. The link significance (LS) was derived as:

$$\boldsymbol{LS = FC \times L \times OD\ CENTRALITY}$$

(2-39)

Where FC is the functional class and L is the number of lanes. The best model for predicting AADT was a mixture of three types of link significance and speed (R^2=0.66):

$$\boldsymbol{AADT = -25359.77 + 920.11\,(Sp) + 2.69 \times 10^{-7}\,(I - I\ LS) - 1.11\,(I - E\ LS) + 62.93\,(E - E\ LS)}$$

(2-40)

In the above equation, Sp is the speed, $I - I\ LS$ is the internal-internal lane significance, $I - E\ LS$ is the internal-external lane significance, and $E - E\ LS$ is the external-external lane significance. The model was validated against AADT obtained from 109 short term count locations in Greenville. The above model was found to perform better than the city's existing travel demand model which had an R^2 of 0.61. Validation results show that the root mean square error (RMSE) of this model and the existing travel demand model were 7352.54 and 14073.19 respectively.

Jayasinghe et al. (2017) proposed a more robust method using multiple weighted network centrality (MWNC) to predict AADT on road segments in Colombo metropolitan area, Sri Lanka. Using multiple centrality enabled them to capture pass-by and in-between point traffic. Weighted link cost captured the roadway characteristics in addition to topological distances. Unlike other methods this method seems to be more useful in data-constrained areas, since it does not use land use or O-D trip information. The database used consisted of road network, road type, average travel time, and road capacity in a GIS format. MWNC values for each road segment are calculated using three centrality measurements: betweenness centrality

(BCC), global closeness centrality (GCC), and local closeness centrality (LCC). Betweenness and closeness centralities account for pass-by locations and accessibility of a location, respectively. The research used 2014 AADT data from Colombo which had 1181 sample points. 90% of the data was randomly chosen for calibration of the model and the remaining 10% was used for validation. The best model had a R^2 of 0.93 and Median Absolute Percentage Error (MdAPE) of 26%:

$$\boldsymbol{log_{10}AADT = 1.214 + 0.503 * log_{10}BCC_i + 0.981 * log_{10}GCC_i + 3.209 * log_{10}LCC_i} \tag{2-41}$$

In which $BCC_{(GMD\&Ty)i}$ is the betweenness centrality using geometric distance and road type, $GCC_{(GMD\&Ty)i}$ is the global closeness centrality using geometric distance and road type, and $LCC_{(MD\&Ty)i}$ is the local closeness centrality metric using metric distance and road type.

Geospatial methods offer an intriguing middle ground between regression and travel demand based approaches by incorporating network characteristics, topology, connectivity related information, and spatial distribution of relevant land use and socio demographic variables in a regression setting. The models appear to perform reasonably well with respect to prediction errors. However, all models have been tested for compact urban areas and the performance in more rural areas is still an open avenue for research.

2.4 MACHINE LEARNING

Machine Learning is an artificial intelligence technique which relies on pattern recognition algorithms to predict missing AADT information from available land use, socio-demographic, and economic data. Three types of machine learning techniques have been applied for AADT estimation – support vector regression, neural networks, and classification based on regression trees.

2.4.1 Support Vector Regression

Sun and Das (2014) developed a support vector regression (SVR) model to predict AADT on non-state roadways for eight parishes in Louisiana, none of which had a major urban area. Four out of the eight parishes had direct access to interstates. Separate models were developed for the eight parishes. Data on the following independent variables were collected: total population and employment at the census block where traffic counts were collected and shortest distance from the count location to interstates and major US highways. The SVR model was found to outperform poisson and negative binomial regression approaches. The total percentage of AADT predictions which lie within 100 of the actual AADT counts ranged from 64% to 82% for the eight parishes tested. Similar to other regression approaches observed in the literature, the SVR model was found to underestimate AADT at higher counts (> 1500 vpd).

Castro-Neto et al. (2009) developed a SVR model to predict AADT one year into the future from past year AADT values. AADT data from 25 counties from 1985 to 2004 in Tennessee was assembled. For each county, two separate time series were created – one for rural and the other for urban roads. The 1985 to 1999 data was used for calibration and one year forecasts from 2000 to 2004 were compared with the observed values. In addition to SVR, a Holt Exponential Smoothing time series forecasting model and a simple linear regression model was used. The validation showed that SVR outperformed the Holt Exponential Smoothing technique and simple linear regression model with a mean absolute percentage error of 2.26% for urban roads and 2.14% for rural roads.

2.4.2 Neural Networks

Lam and Xu (2000) compared the performance of neural networks and regression based approaches in predicting AADT from short term counts in Hong Kong. The analysis was carried out at 13 locations in an urban area. The input variables considered were 4, 6, 8, 10, 12, 14 and 16 hour counts respectively. The study recommended use of 8 hour counts for AADT predictions. The neural network method provided more accurate AADT predictions than regression based methods. With 8 hour counts as input, the regression based approaches yielded a maximum absolute percent error of 13.26% compared to 10.88% under neural networks over 13 locations. The sum of the absolute percent error was 49.21% using linear regression approaches. The corresponding error value under neural networks was 36.49%.

Tang et al. (2003) compared the performance of time series, neural networks, nonparametric regression, and gaussian maximum likelihood approaches for predicting AADT based on historical AADT data (1994-1998) and current year short term traffic flow data (1999). Gaussian maximum likelihood models provided the best monthly weekday and all day AADT predictions with a mean absolute percent error of less than 1%. Gaussian maximum likelihood models need very little calibration. However, they require historical and AADT information and also assume that the input traffic flows are normally distributed.

Sharma et al. (1999) compared the performance of the traditional factor based AADT estimation methods with neural network based prediction on highways and major roadways. The methods were tested on data obtained from 63 Automated Traffic Recording (ATR) sites in Minnesota in 1993. The artificial neural network based estimations had an average error of 9.47%. In comparison, the traditional factor based method had a lower error of 5.66%. However in this study, in the traditional factor based method, the 48 hour count location was assigned to the appropriate ATR group with 100% accuracy which is not feasible in real world applications. Also the neural network model did not require any additional input on day of the month or month of the year information. Using two separate 48 hour counts was found to reduce the error in neural network procedures. The 95th percentile error using two separate 48 hour counts was found to vary from 14.14% to 16.68% which is competitive with the traditional factor based method which had an average 95th percentile error of 15.28%.

Sharma et al. (2000) extended the above study to compare the performance of neural networks and traditional factor based approaches on low volume roadways (<1000 vpd) in Alberta, Canada. Data was obtained from 55 ATR sites in 1996. The factor based approaches which relied on 100% accurate assignment of short period 48 hour counts to the correct factor groups provided AADT estimates with 95th percentile errors of nearly 30%. Using neural networks with two 48 hour counts provided lower 95th percentile errors of 25%. In low volume roads, the accuracy of neural network based AADT estimation was found to not vary with volume ranges.

Sharma et al. (2001) performed a detailed investigation on the application of neural network based methods to predict AADT on low volume roads in Alberta, Canada. The study recommended calibrating or training neural networks for each months. The 95th percentile error of 65% in AADT prediction was found to reduce to 35% when using month specific models. Using two 48 hour counts was found to provide the best prediction with a 95th percentile error of 25% when separate neural network models are calibrated for each month-day combination. No significant improvements in accuracy were noted in using two 72 hour counts over two 48 hour counts. Similarly using three 48 hour counts offered no notable accuracy improvements over two 48 hour counts.

2.4.3 Classification based on Regression Trees

Dixon et al. (2004) used classification and regression tree algorithms to predict AADT growth factors on rural roadway segments in Idaho. Regression, time series, and clustering based methods were found to yield poor results. Data from 42 stations were used for calibration and 10 stations for validation. The dependent variable considered was the annual AADT growth rate calculated using 1980 and 1990 AADT data which were then used to predict 2000 AADT data. The independent variables considered were county population annual growth rate, functional class of the segment, and current AADT. The mean absolute percent error was found to be lower than 14.2% in the validation data set. Note that this method relies on past AADT data to estimate growth factors which are used to forecast AADT data which is different from the scope of the project.

The number of studies adopting machine learning approaches for AADT estimation has been increasing especially over the last decade. Machine learning based prediction offers an alternative to statistical regression based approaches and have been increasing in efficiency and accuracy due to the needs of big data analytics in various industries. However, a majority of the machine learning based approaches for AADT estimation rely on short term counts or past AADT values. Hence, machine learning approaches are not useful for the estimation of AADT values in areas with no data. Furthermore, this approach for AADT prediction is relatively new and has not been tested over time or in locations without AADT base values.

2.5 IMAGE PROCESSING

Jian et al. (2006) developed a method to predict AADT by combining ground information and image based information. The AADT was obtained as a weighted linear combination of ;(i) AADT estimated from short term counts which were scaled using appropriate seasonal and growth factors ; (ii) AADT information from number of vehicles observed in an image of highway which was scaled using space mean speed on that particular hour, hourly and other seasonal factors. The weights were proportional to the inverse of the variances of the two estimates. The methodology was tested on 122 highway segments in Florida from 1994 and 2003. Including the information from image was found to improve the accuracy and increase the chances of obtaining AADT with an error of less than 10%. With increasing developments in image processing and access to street mapping tools, static roadway images provide a promising avenue to supplement count information. However, such approaches rely on access to images which capture traffic conditions which are most common to that roadway facility which might not always be available.

2.6 SUMMARY

The missing AADT estimation literature was categorized based on methodology into regression, travel demand modeling, geospatial, machine learning, and image processing based approaches. Table 2-1, Table 2-2, Table 2-3, and Table 2-4 summarizes the studies based on methodology and scope of application, input data categories (socio-demographic, economic, land use, network, and traffic), input data (objective vs subjective), and validation accuracies respectively. Note that in Table 2-3, we use objective to characterize those variables that are clearly defined and transferable and subjective to indicate those variables that are customized to one particular location or region and may not be directly applicable to Oregon. For example, functional classification is normally an objective variable. However, *Xia et al. (1998)* define functional classification related variable as being equal to 0 for local and unclassified, 1 for city and county collector, 2 state and county minor arterial. Similarly, *Zhao and Chung (2001)* define functional classification as being equal to 0.6 for unclassified, 1.0 for urban collector, 2.2 for urban minor arterial, and 3.4 for urban principal arterial. In these two specific papers, functional classification is characterized as subjective as they are customized for their specific applications. Similarly, population, employment within distance buffers is characterized as subjective as there is no science behind choosing a buffer distance threshold other than model fit. For example, the buffer distance threshold for Florida may not be applicable for Oregon. Accessibility indices are often subjective as there are several ways to define them and the choice of accessibility indices in the application is based on model fit. The following insights were obtained from the literature review:

- The performance of a methodological approach depends on scope of application (statewide vs small urban area) and the data availability. In general, the recommendation is to develop models customized to various regions (urban vs rural, north vs south) rather than rely on a single statewide model.

- Multiple Linear Regression models are the simplest and most widely used but with high variability in accuracies. Spatial regression models appear to perform better than multiple linear regression but are more complex to calibrate and may have transferability issues. Spatial approaches have not yet been tested or transferred to other areas.

- Travel demand modeling based approaches have promising performance but are more complex to set up especially in rural areas.

- Geospatial methods embedded in a regression framework provide very promising results but in smaller compact regions.

- The machine learning and image processing based approaches have been applied only to estimate missing in-network AADT values. Hence, these approaches are not recommended because they are promising but not yet practice ready.

- Some of the reviewed models may have over fitted the data or developed models with subjective variables that do not transfer well to other regions.

- The regression, travel demand modeling, and geospatial methods have been successfully applied to estimate missing out-of-network AADT values. Hence, these approaches are relevant for this research.

- Each of the out-of-network AADT methodological approaches has its advantages and disadvantages with no clear winner. There is no one particular methodological approach which dominates in terms of accurately predicting missing AADT information.

- Regardless of the methodology utilized, robust models that are parsimonious and intuitive are more likely to stand well the test of time and transferability. Models should be properly validated with independent data to avoid overfitting and misspecifications.

Table 2-1: Classification of Literature based on Methodology and Application Scope

Methodology	Literature	Scope
Multiple Linear Regression	Dadang et al. (1998)	statewide: county roads in Indiana
	Xia et al. (1999)	county: non-state roads in urban (population over 1 million) Broward county in Florida
	Zhao and Chung (2001)	county: state and non-state roads in Broward county in Florida
	Pan (2008)	statewide: county highways and local streets in rural, urban and large metropolitan areas in Florida
	Yang and Wang (2014)	county: local functional class roads in Mecklenburg county, NC
	Anderson et al. (2006)	small urban community of Anniston, Alabama
	Seaver et al. (2000)	statewide: rural roads in Georgia counties
	Dixon et al. (2011)	statewide: minor legs of rural intersections
	Barnet et al. (2015)	statewide: minor legs of stop controlled intersections in Oregon, Washington, and North Carolina
Spatial Regression	Zhao and Park (2004)	county: state and non-state roads in Broward county in Florida
	Eom et al. (2006)	county: roads in Wake county, North Carolina
	Wang and Kockelman (2009)	statewide: Texas highways
	Musunuru et al. (2017)	rural, two-lane horizontal curve road segments in Utah
	Selby and Kockelman (2013)	regional: Texas network but with specific models for each region
	Shamo et al. (2015)	county: Washington State roads
Nonlinear Regression	Staats (2016), Souleyrette et al. (2016)	regional: non-state roads in Kentucky
Travel Demand Modeling	Zhong and Hanson (2009)	county: low-class roads in York county and Beresford regions in New Brunswick, Canada
	Wang et al. (2013)	county: local roads in Broward County Florida
Geospatial Method	Jayasinghe et al. (2017)	road segments in Colombo metropolitan area, Sri Lanka

	Lowry (2014)	local, collector, minor arterial, principal arterials in Moscow, Idaho
	Pulugurtha and Kusum (2012)	Charlotte, NC: freeway, major thoroughfare, and minor thoroughfare
	Keehan et al. (2017)	small sized city, Greenville, South Carolina
	Kusam and Pulugurtha (2016)	Charlotte, NC: freeway, major thoroughfare, and minor thoroughfare
Support Vector Regression	Sun and Das (2014)	parish: non-state roadways for eight districts in Louisiana with no major urban area
	Castro-Neto et al. (2009)	statewide: roads that have a known AADT
Neural Networks	Lam and Xu (2000)	urban area in Hong Kong
	Tang et al. (2003)	urban area in Hong Kong
	Sharma et al. (1999)	statewide: highways and major roadways in Minnesota
	Sharma et al. (2000)	statewide: low volume roadways in Alberta, Canada
	Sharma et al. (2001)	statewide: low volume roadways in Alberta, Canada
Classification based on Regression Trees	Dixon et al. (2004)	statewide: rural roadway segments in Idaho
Image Processing	Jian et al. (2006)	statewide: highway segments in Florida

Table 2-2: Categorizing Variables in Literature Review

Category	Variables
Socio-demographic data	• county population, distance to mean centers of population, population data aggregated based on buffer distances, population within incorporated area, population density, population accessibility index, population at Census block, percentage of population change over years, county population annual growth rate • automobile ownership • service employment, agricultural employment, transportation employment, accessibility to employment centers, employment data aggregated based on buffer distances, labor force, employment at census block, number of persons working outside of county • total number of households • number of dwelling units, number of manufacturing houses aggregated at buffers of various distances, area of manufactured homes and mobile parks, area of major educational, medical, government, cultural, or religious institutions • median travel time to work • number of buildings, area of each building
Roadway characteristics	• number of lanes, number of lanes in downstream cross street, functional classification, median type, speed limit, upstream link speed limit, downstream link speed limit, roadway curvature road capacity • whether the cross street is a minor arterial or not, whether the cross street is a major collector or not, number of downstream cross streets, indicator variable for urban minor arterial, principal arterials, local street and collector, number of principal arterials within one mile, number of freeways within 2 miles, number of major collectors within 2 miles • latitude, longitude • presence of a right turn lane on the minor road, presence of a right turn lane on the major road, presence of left turn lane, presence of a centerline on the minor road, presence of striped edgelines on the minor road, presence of edge stripping, presence of center striping
Network Characteristics	• connectivity to a city or town within two miles, accessibility to parking lot, direct access to expressway/freeway, road network connectivity importance index, whether the roadway section

	is a through or destination street • arterial mileage, total lane mileage of highways • accessibility to state highway, accessibility to freeways, shortest distance from the count location to interstates and major US highways • median travel time to work • distance to MSA
Economic characteristics	• personal income, percentage of people below poverty line, median household income, unemployment rate, per capita income • retail sales
Land use characteristics	• area type (rural or urban), land use type, residential properties, commercial properties • number of agricultural farms, percentage of farms with 500 acres or more • is the adjacent land developed or not • presence of the link in CBD
Traffic characteristics	• AADT estimates, ATR counts, current AADT, short term counts • daily average probe count • number of vehicles observed in an image of highway, number of cars on the road, car intensity

Table 2-3: Summary of Data used for AADT Estimation

Literature	Data	
	Objective	**Subjective**
Dadang et al. (1998)	area type (urban or rural), county population, and arterial mileage	accessibility to state highway
Xia et al. (1999)	number of lanes, automobile ownership, and service employment	functional classification, land use type
Zhao and Chung (2001)	number of lanes, accessibility to employment centers, accessibility to expressways, distance to mean centers of population	functional classification, population and employment data aggregated based on buffer distances
Pan (2008)	number of lanes in both directions, median type, total lane mileage of highways, vehicle registration, personal income, retail sales, population within incorporated areas, and labor force	location type (urban or rural), land use, accessibility to freeways, population
Yang and Wang (2014)	number of cars on the road from google map images, number of lanes, housing units, median income, percentage of people below poverty line, and car intensity	
Anderson et al. (2006)	functional classification, number of lanes	population and employment within a 0.5 mile buffer of a traffic count station, and whether the roadway section is a through or destination street
Seaver et al. (2000)	roads outside the Metropolitan Statistical Area (MSA): percentage of population change from 1990 to 1996, median travel time, number of agricultural farms, median household income, median time to leave for work, and distance to MSA roads inside the MSA: population density, unemployment rate, median travel time, percentage of farms with 500 acres or more, median time to leave for work, and number of persons working outside of county	percentage of farms with 500 acres or more,

Dixon et al. (2011)	whether the cross street is a minor arterial or not, whether the cross street is a major collector or not, presence of a right-turn lane on the minor road, presence of a right-turn lane on the major road, presence of a centerline on the minor road, and presence of striped edgelines on the minor road	whether the adjacent land is developed or not
Barnet et al. (2015)	indicator variable for urban minor arterial, principal arterials, local street and collector, population density, presence of edge striping, presence of center striping, agricultural employment, transportation employment, per capita income, and presence of left turn lane	connectivity to a city or town within two miles, number of principal arterials within one mile, number of freeways within 2 miles, number of major collectors within 2 miles, accessibility to a parking lot
Zhao and Park (2004)	number of lanes, accessibility to employment, and direct access to expressways	population and employment data aggregated based on buffer distances
Eom et al. (2006)	latitude, longitude, speed, median income, number of lanes	area type and functional classification
Wang and Kockelman (2009)	AADT estimates for 27,738 sites from 1999 to 2005	
Musunuru et al. (2017)	traffic count, functional classification, number of lanes, population, and household unit counts	
Selby and Kockelman (2013)	road type, speed limit, number of lanes	population accessibility index
Shamo et al. (2015)	AADT data set from 2008 to 2010	
Staats (2016) and Souleyrette et al.(2016)	2012 daily average probe count, roadway curvature	residential properties, commercial properties
Zhong and Hanson (2009)	road network data, total number of households, average income per household in each zone, retail and non-retail employment in each zone	
Wang et al. (2013)	roadway network data, traffic count, parcel level land use type data (mainly number and area of dwelling units)	

Jayasinghe et al. (2017)	road network, road type, average travel time, and road capacity	
Lowry (2014)	functional classification, number of lanes, and connectivity importance index	
Pulugurtha and Kusum (2012)	number of lanes, speed limit, upstream link speed limit, downstream link speed limit, number of lanes in downstream cross street	population, number of manufacturing houses aggregated at buffers of various distances, an indicator for rural and urban land use type
Keehan et al. (2017)	speed limit, number of lanes, number of buildings, area of each building, and count data	functional classification, land use,
Kusam and Pulugurtha (2016)	presence of the link in central business district, number of lanes, downstream link speed limit, upstream link speed limit, number of lanes in downstream cross street, number of manufacturing houses, and speed limit, state owned right of way for all roadway segments,	area of manufactured homes and mobile parks, planned unit development area, area of major educational, medical, government, cultural, or religious institutions, number of manufacturing houses aggregated at buffers of various distances,
Sun and Das (2014)	total population and employment at the census block and shortest distance from the count location to interstates and major US highways	
Castro-Neto et al. (2009)	AADT data from 25 counties from 1985 to 2004 in Tennessee	
Lam and Xu (2000)	4, 6, 8, 10, 12, 14 and 16 hour counts	
Tang et al. (2003)	historical AADT data (1994-1998) and current year short term traffic flow data (1999)	
Sharma et al. (1999)	48 hour sample traffic counts	
Sharma et al. (2000)	48 hour sample traffic counts	
Sharma et al. (2001)	48 hour sample traffic counts	
Dixon et al. (2004)	county population annual growth rate, functional classification, and the current AADT	
Jian et al. (2006)	short term counts and number of vehicles observed in an image of highway	

Table 2-4: Summary of Validation Results

Literature	Number of Data Points		Summary of Validation Results
	Calibration	Validation	
Dadang et al. (1998)	count stations on 40 counties	count stations on 8 counties	percentage difference of observed and predicted AADT ranged between 1.56% to 34.18% with an average of 16.78%
Xia et al. (1999)	399	40	percentage difference of observed and predicted AADT ranged from 1.31% to 57% with an average difference of 22.7%
Zhao and Chung (2001)	816	82	percentage difference between predicted and observed AADT varied from 0.3% to 155.6% for Model 1 and from 0% to 288.1% for Model 3
Pan (2008)	26721	1149	average mean absolute percentage error varied from 31.99% to 159.49%
Yang and Wang (2014)	200	43	nearly 50% had an absolute percentage difference of less than 37%
Anderson et al. (2006)	58	38	no statistically significant differences between the predictions and the observed counts
Seaver et al. (2000)	1213	-	-
Dixon et al. (2011)	-	-	-
Barnet et al. (2015)	420	54	percentage error ranged from 1% to 214% with an average error of 59%
Zhao and Park (2004)	775	82	about 85% of the data points had errors less than 50% for the OLR model
Eom et al. (2006)	200	954	-
Wang and Kockelman (2009)	22190	5548	median error was found to be a 33% overestimation
Musunuru et al. (2017)	data from 100 ATRS from 2009 to 2013	-	MSE=0.6674 and 0.8238 for day and night models, respectively

Selby and Kockelman (2013)	25183	-	spatial regression models had absolute error reductions between 16% and 63% compared to non-spatial regression; universal kriging had 3% to 8% lower absolute errors compared to geographically weighted regressions
Shamo et al. (2015)	for 2008, 2009 and 2010, 4992,7485 and 7734 samples respectively	-	-
Staats (2016) and Souleyrette et al. (2016)	-	-	mean absolute percentage difference varied between 61% and 97% for rural areas and between 354% and 1956% for urban areas
Zhong and Hanson (2009)	-	26	overall error of 17%
Wang et al. (2013)	-	78	mean absolute percentage error of 52%
Jayasinghe et al. (2017)	1163	118	median absolute percent error of 26%
Lowry (2014)	307	34	median absolute percent error of 22%
Pulugurtha and Kusum (2012)	90%	10%	absolute average percentage errors were lower than 36% for all models
Keehan et al. (2017)	109	-	RMSE of this model and the existing travel demand model were 7352.54 and 14073.19 respectively
Kusam and Pulugurtha (2016)	-	-	absolute average percentage errors were varied between 18% and 26%
Sun and Das (2014)	43,755	-	total percentage of AADT predictions which lied within 100 of the actual AADT counts ranged from 64% to 82%
Castro-Neto et al. (2009)	1985 to 1999 data	2000 to 2004 data	mean absolute percentage error of 2.26% (urban) and 2.14% (rural).
Lam and Xu (2000)	-	-	maximum absolute percent error - 13.26% (regression) vs 10.88% (neural networks); sum of the absolute percent error - 49.21% (linear regression) vs 36.49% (neural networks)

Tang et al. (2003)	data from Jan 1994 to December 1998	data from Jan 1999 to December 1999	gaussian maximum likelihood models: mean absolute percent error of less than 1%
Sharma et al. (1999)	63	-	average error of 9.47%; 95th percentile error using two separate 48 hour counts was found to vary from 14.14% to 16.68%
Sharma et al. (2000)	55	-	95th percentile errors of 25%
Sharma et al. (2001)	-	-	month specific models - 95th percentile error of 35%; month-day combination with two 48 hour counts - 95th percentile error of 25%
Dixon et al. (2004)	42	10	mean absolute percent error lower than 14.2%

3.0 DOT PROCEDURES

The research team surveyed state DOTs to identify any best practices followed in estimating missing AADTs. It is important to classify missing AADTs as in-network and out-of-network. For in-network missing AADTs some AADT values are missing but there have been vehicle counts performed in the past, for example there is no AADT estimation for the years 2014 and 2015 but there is count data from year 2013 or there is a count in a nearby section of the same highway. Appendix A reviews existing AADT estimation procedures, which can be applied for in-network cases. For out-of-network missing AADT values, there are no records with past counts at a specific location or at nearby links; secondary data sources are necessary. The latter is the most challenge case and the main focus of this research project.

3.1 MISSING IN-NETWORK

Most DOTs estimate in-network missing AADT utilizing or adapting the methods described in the Federal Highway Administration's Traffic Monitoring Guidelines.

3.1.1 Alaska

Alaska DOT computes AADT for segments "not counted" during the reporting year using one of the following two options: (i) by applying growth factor to a segment by road functional class and region ID (by area), and (ii) using the statewide average. The Alaska DOT generally use the first method. The annual growth factors are calculated as follows:

- Select all stations which have an AADT value calculated from actual traffic (i.e. these AADT values were calculated using actual counts, not estimated from secondary sources) in the reporting year.

- For each station found, select its next most recent actual AADT value (i.e. don't use an AADT figure from a year that was estimated).

- Calculate the growth factor *(GF)* using the formula:

$$\boldsymbol{GF = e^{ln(currentADT/previousAADT)/(currentYear - previousAADTYear)} - 1} \tag{3-1}$$

- The average growth factor is calculated by summing each of the station growth factors and dividing by the number of stations. This means one station is not weighted more than another.

- The average growth rate is capped at a maximum value. Then, these rates applied to the "not counted" by road functional class (RFC) and during the reporting year.

Alaska's travel monitoring software, is Traffic Server, provides the ability for users to set and apply their own growth rates to segments by RFC and region if those growth rates are computed outside of Traffic Server.

3.1.2 Arizona

Arizona DOT categorizes a permanent count station as urban or rural. A geographic area or polygon is associated with each permanent count station. If the permanent count station is located in a rural area, then all roadway segments in the polygon are grouped into two groups – functional class 1 roadways and functional class 2-7 roadways together. If the permanent count station is located in an urban area, then all roadway segments in the polygon are grouped into three groups – functional class 1 roadways, functional class 2 roadways, and functional class 3-7 roadways together. An average growth rate is calculated for the permanent station. The current year AADT for all roadway sections in the groups are estimated based on the previous AADT information and growth factors. Arizona DOT subscribes to the Transportation Data Management System (TDMS) web application built by MS2. This application provides a tool to review the underlying traffic data and calculate the seasonal, axle, growth factors, and finally AADT. The program calculates the AADT by applying seasonal factors and axle factors to non-permanent/short counts. If the counts are missing the program will calculate the missing AADT based on previous years AADT and growth factors.

3.2 MISSING IN AND OUT-OF-NETWORK

Some DOTs have procedures that cover both in and out-network missing AADT cases.

3.2.1 Arkansas

Arkansas DOT locate "like" segments that have an existing AADT and apply that AADT to the new segment. "Like" segments are identified based on the following characteristics: county, rural vs urban, functional classification, paved vs unpaved, number of lanes, and one-way vs two-way.

3.2.2 Florida

Florida DOT uses the results of the Turnpike State Model (TSM) statewide transportation model to estimate missing traffic volumes for street and roads in Florida. The existing data come from five sources to get a "best estimate" for AADT volumes on target roads: (i) a statewide parcel layer published annually from Department of Revenue to determine the number of housing units along target segments, (ii) data from InfoUSA is used to determine the employment sites and the number of employed, (iii) a shapefile that is derived from the observed AADT, (iv) estimated AADT values generally for all Florida's major roadways as well as total number of trips by

Traffic Analysis Zones (TAZ) provided by TSM, and (v) detail street level linear GIS information provided by Navteq Street Network.

Navteq Street Network is a commercial available street GIS database. Street segments in the Navteq Street Network are divided into two categories (Tier 0 and Tier 1-N). Each TAZ is analyzed separately as a unit and is categorized. If the segment does not have the estimated AADT from TSM, the Tier rank of route within each TAZ must be calculated by obtaining the number of employees and housing unit per TAZ. Final AADTs within a TAZ were developed from the calculation of trips on routes using employees, housing, and trip factor. The Allocator Process maintains a list of how each route is connected to other routes. The final volume of a route is equal to the trips for the route plus the accumulation of trips from higher tiered routes that are connected to the route.

3.2.3 Georgia

Georgia DOT generates AADTs according to 4 methods. (i) Actual – based on data collected at the location in the reporting year, either permanent or portable and factored. (ii) Estimate – based on data collected at the location from a previous year with a growth factor applied. (iii) Calculated – based on data collected at adjacent locations on the same route/facility. For example, assume State Route 1 is sectioned into traffic segments A, B, C, D, and E. If the AADTs on A, D, and E are actual or estimated, then the AADT on segment B and C can be calculated from those values on the same route. (iv) Applied – based on counts on other routes of similar functional class, urban code, pavement type (paved/unpaved), and geographic location.

3.2.4 Illinois

Illinois DOT conducts approximately 20000 traffic counts annually. The AADT for off system roadway sections are common sense estimated based on reviewing: (i) routes in the adjacent area and actual counted AADT values, (ii) aerial imagery for factors which could affect traffic patterns (dead end roads, major trip generators/attractors etc.), and (iii) roadways with similar functional class in the area with physical counts.

3.2.5 Iowa

Iowa DOT reviews new city and secondary roadways for their land use characteristics such as number of houses if residential, size and type of businesses and the roadway's proximity to other traffic generators and estimates a traffic volume. This traffic volume is assigned to the roadway. For the new roads with no traffic data available, a constant value between 6 and 15 will be set as a trip generation for each household and then the AADT will be calculated based on this value. If the roadway is functionally classed higher than local, then counts are conducted at some point and the estimated traffic volume is replaced with the count. Iowa DOT collect both manual counts and tube counts for minimum 48 hours at the selected locations on state highways. For secondary roads, the tube counts is used to collect data for minimum 24 hours. Iowa DOT also coordinates with the local city and counties to collect traffic data on roadways they want current traffic data collected on. These counts are usually collected between

May and September. Although, for some locations the counts are collected only in April. That's because there are three major university in the state and traffic counts shows lower volume between May to September due to the summer time. Iowa DOT used a vendor software (Traits) to calculate the factors to estimate the AADT from the counts volume. Recently, the DOT is in a transition to another software called High Desert Traffic (Jackalope system) to estimate AADT.

3.2.6 Kansas

Kansas DOT has a counting program that samples roads from all functional classes. Roadway segments without a count are estimated using one of five possible methods: (i) using route flows, if adjacent counts are available, (ii) using route average, if the route is sampled, (iii) city functional class average, (iv) population group functional class average, or (v) county functional class average for non-corporate areas. For local roads in single-family developments, estimates can be developed from count of the driveways (houses) by a manual process.

3.2.7 Montana

Montana DOT assigns default values to the not counted roadway segments. Actual traffic counts were used to generate county level averaged default values for a roadway section categorized by functional class, urban vs rural, and paved vs unpaved.

3.2.8 Mississippi

Mississippi DOT use blanket counts that are calculated for each functional class and county for functional classified routes. For the local routes, the DOT conducts a local sample counting and use those to cover the local system.

3.2.9 Nevada

Nevada DOT estimates AADT by using nearby roads with similar functional classifications on the roadway network. For example, a known minor collector AADT can be used to estimate a missing minor collector AADT in the same neighborhood. Nevada DOT is using local knowledge of experienced field technicians to estimate AADT. They are familiar with what a 500 AADT road looks like, versus a 300, 800 or 1000 car AADT roadway segment. A reasonable estimate can be obtained through field experience if they are also familiar with the roadway that DOT wish to estimate an AADT.

3.2.10 South Carolina

South Carolina DOT collects the counts yearly and factors up the counts based on the roadway's functional classification growth factors provided by the Planning department's traffic projection group. SC DOT enters a default count based on rural/urban area into the system if the data for a new route is not available.

Chowdhury (2015) from Clemson University is conducting a research project for estimating statewide AADT with SC DOT. There are no published methodologies yet.

3.2.11 Vermont

Vermont DOT determines actual AADT from an automatic traffic recorder count on the federal aid routes. They also use tube counts to collect traffic counts for a week between May to October. DOT conducts 12-hour turning movement counts at intersections. An expansion factor is calculated based on the ratio of the actual AADT to the 12-hour approach total, and applied to the 12-hour approach total on the missing AADT side of the intersection. Vermont DOT use a cloud based system (MS2) to estimate the missing AADT.

3.2.12 Washington

Washington DOT does not have any missing AADT data for state highways. The DOT ask all the cities and counties to collect AADT data and growth factors. The DOT maintains AADT values for the total system on arterial and collector roads. AADT are obtained from Local Agency count data except for 13% that have been estimated by WS DOT based on AADT for adjacent road segments. Vehicle Mile Traveled (VMT) for local access roads is estimated as a percentage based on the arterial and collector VMT. WS DOT consider rural arterial and collector VMT to represent 93% of total rural VMT, and the remaining 7% is the estimated rural local access VMT. WS DOT uses 89% to calculate total urban VMT, for urban arterial and collector VMT, and the remaining 11% of the total urban VMT is the estimated rural local access VMT.

3.2.13 Wisconsin

Wisconsin DOT uses three methods for estimating missing AADT: (i) if there is a previous count for the location, then it would be growth factored. (ii) If there isn't a previous count DOT would take one. iii) If the facility is not one DOT is responsible for (i.e., a local road), they would refer to the appropriate local unit of government.

3.3 SPECIAL CASES

New Jersey and British Columbia DOT indicated that they do not have any procedure to estimate missing AADTs and if needed they plan to follow the Federal Highway Administration's Traffic Monitoring Guidelines. Washington DC (mostly urban and densely populated) has counts for all roadway segments.

3.4 SUMMARY

Most state DOTs which responded to the survey tend to have ad hoc approaches to estimate missing AADT information. The most useful or promising approaches include:

- Classify links into groups of similar characteristics and apply an average figure for missing links (e.g. Arkansas).

- Estimate number of nearby households or employees and a trip factor to estimate AADT (e.g. Florida, Iowa).

- Sample categories with missing AADT values and apply an average value (e.g. Georgia, Mississippi).

- Utilize a statewide model and interpolate from available counts or estimations (e.g. Florida).

- Utilize default values for a given functional class (e.g. Montana).

- Ask local staff or technicians to estimate AADT (e.g. Nevada).

- Coordinate with local cities and counties to get counts (e.g. Washington and Iowa)

- A combination of approaches abovementioned (e.g. Florida).

It is worth mentioning that several state DOTs were interested in the results of this research project which seems to suggest that many DOTs may be interested in adopting better methodologies to estimate out-of-network AADT values.

4.0 DATA SOURCES IN OREGON

This chapter analyzes the various data sources available in Oregon and their coverage. The search for data sources was guided by the various traffic, roadway, and socio-demographic factors which were commonly considered in the previous research.

4.1 SOCIODEMOGRAPHIC AND ECONOMIC DATA

4.1.1 Census

Article 1, Section 2 of the Constitution mandates the census counts which take place every 10 years. The census data is used to determine the number of seats each state has in the U.S. House of Representatives and for the distribution of federal funds to local communities.

The latest available Census database is published for 2010 at block level resolution. For the whole state of Oregon, there are 196,621 blocks in this database. Figure 4-1 shows the number of blocks in each ODOT region. Region 2 has nearly 29% of the census blocks in the state of Oregon. 321 blocks are outside the ODOT regions.

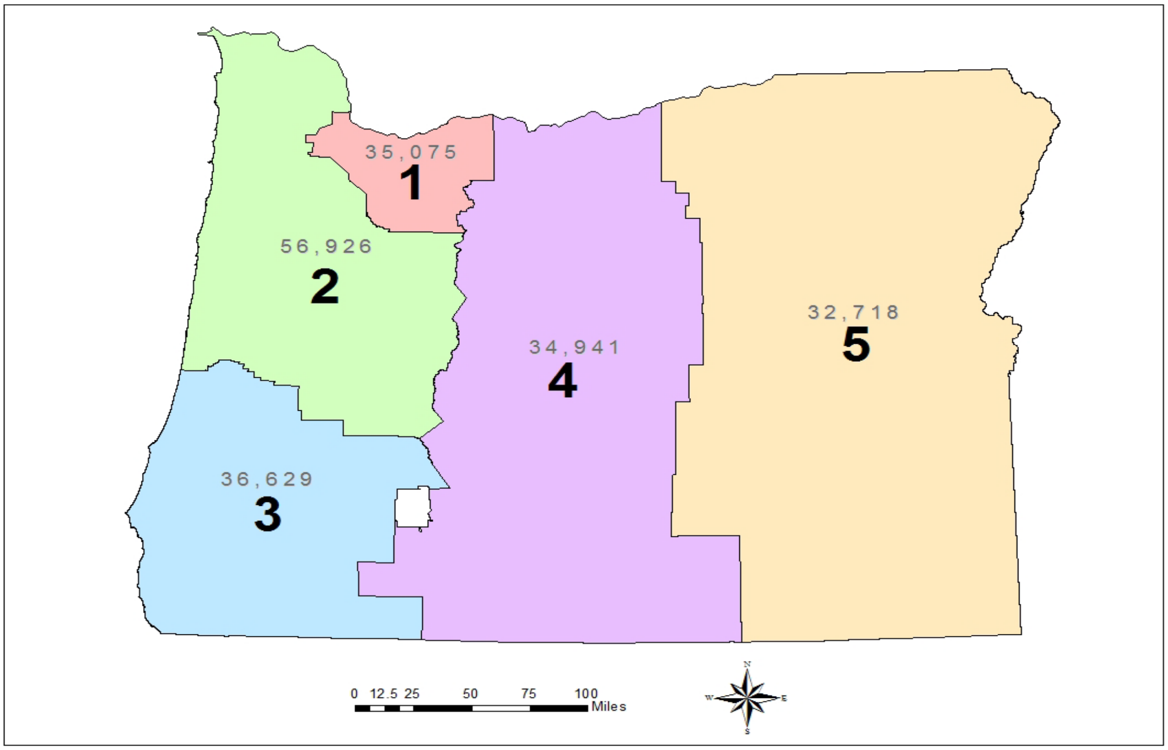

Figure 4-1: Number of census blocks in ODOT regions

Figure 4-2 and Figure 4-3 show the population and household density in region 2. As expected, the population and household densities are higher in the valleys and coastal regions compared to mountainous terrains.

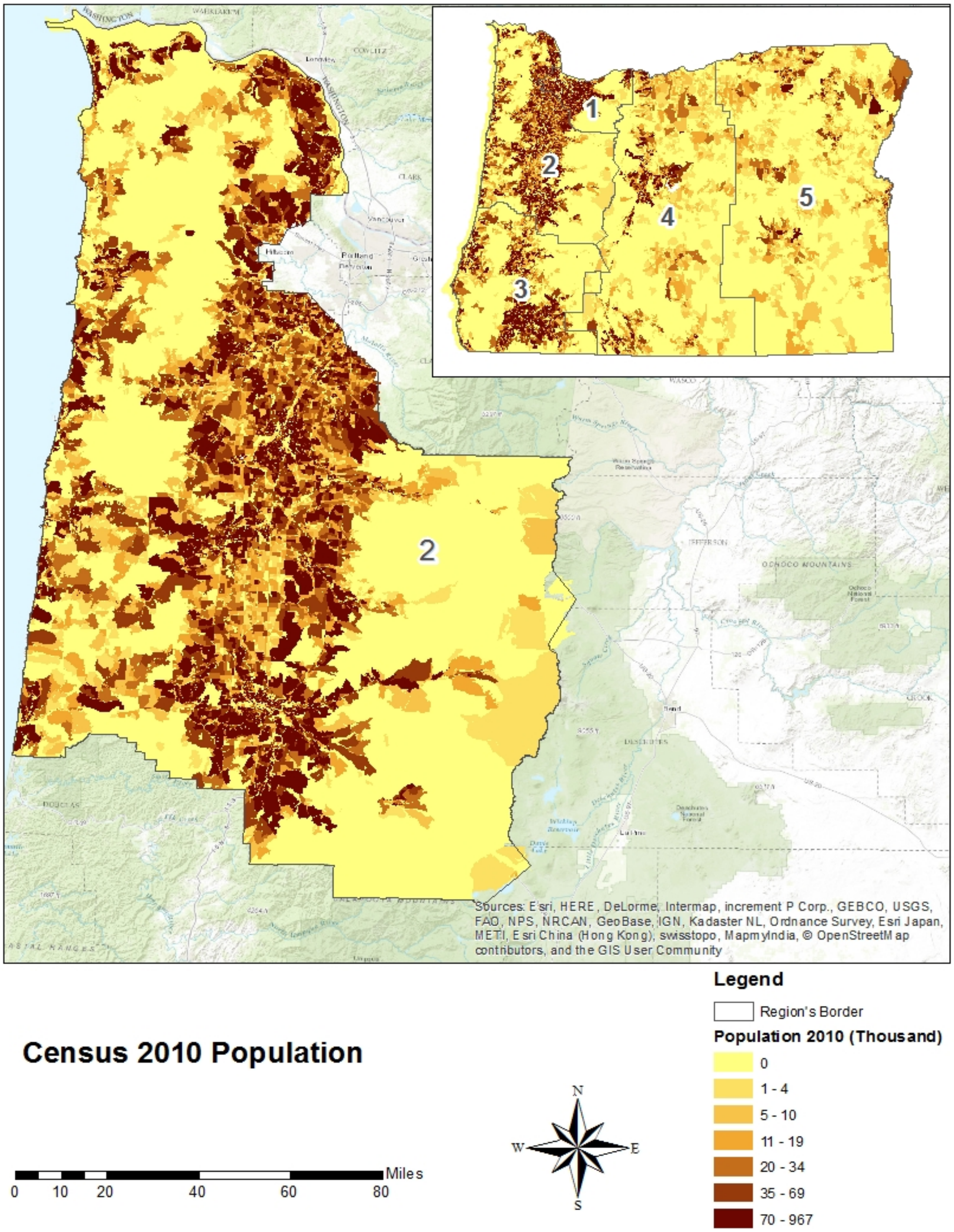

Figure 4-2: Census 2010 population density in Oregon and ODOT Region 2

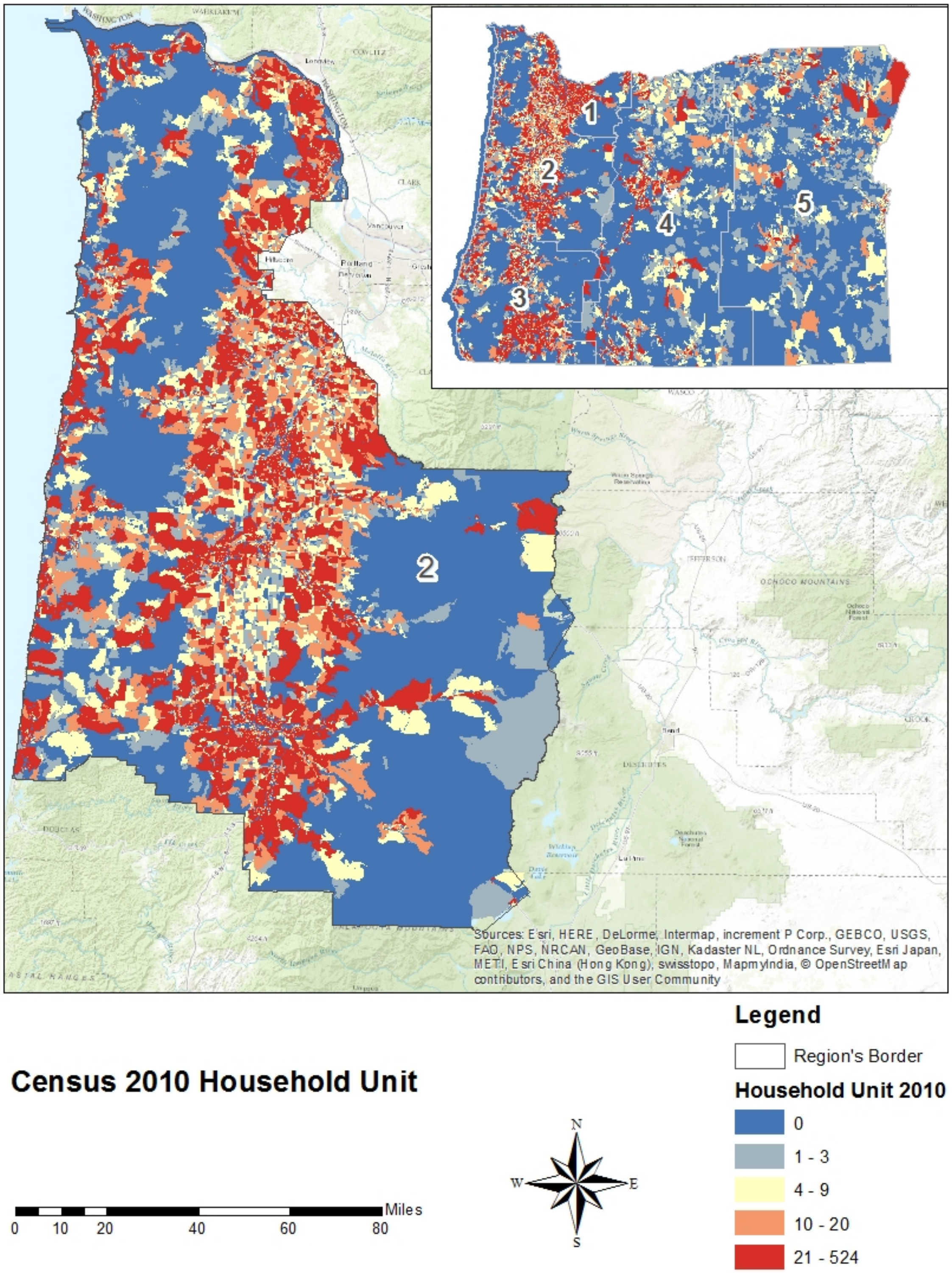

Figure 4-3: Census 2010 household unit density in Oregon and ODOT Region 2

Table 4-1 summarizes the density of blocks, population and housing units for each ODOT region. Region 2 has approximately 32% of the total population and housing units in Oregon.

Table 4-1: Density of Blocks, Population, and Housing Units for each Region

Region	Land Area (sq.mile)	Total Population	Total Housing Units	Number of Blocks per sq.mile	Population per sq.mile	Number of Housing Units per sq.mile
1	2971	1636660	693425	11.8	550.9	233.4
2	14425	1227594	530662	3.9	85.1	36.8
3	12712	478301	220760	2.9	37.6	17.4
4	27562	306261	152091	1.3	11.1	5.5
5	38050	182126	78425	0.9	4.8	2.1

From the Census 2010 GIS database, the following attributes may potentially be useful for the next stages of this research:

- Total households
- Total population
- Categorized population by age and race

Table 4-2 shows the mean, 15^{th}, 50^{th}, and 95^{th} percentiles of census 2010 block sizes (areas) for region 2. Figure 4-4 shows the cumulative percentages of frequency of block sizes (areas) for region 2 up to 5 square miles. Most of the parcels in region 2 are smaller than 1 square mile.

Table 4-2: Mean, 15^{th}, 50^{th}, and 95^{th} percentiles of Census 2010 Block Sizes (Areas) for Region 2

Mean	15^{th} Percentile	50^{th} Percentile	95^{th} Percentile
0.2563	0.0037	0.0170	1.1008

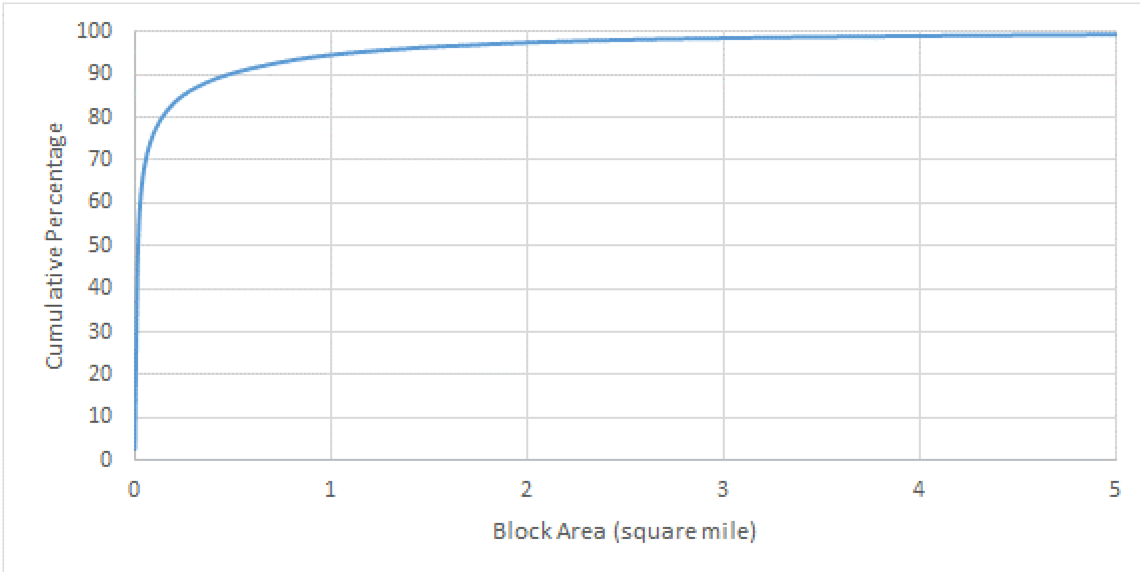

Figure 4-4: Cumulative % for census 2010 by blocks for Region 2 up to 5 square miles

To get a better understanding of the block sizes in this database, Figure 4-5 and Figure 4-6 illustrate the resolution of these data at a part of the city of Salem and Corvallis. Block sizes increase as we move away from the urban core.

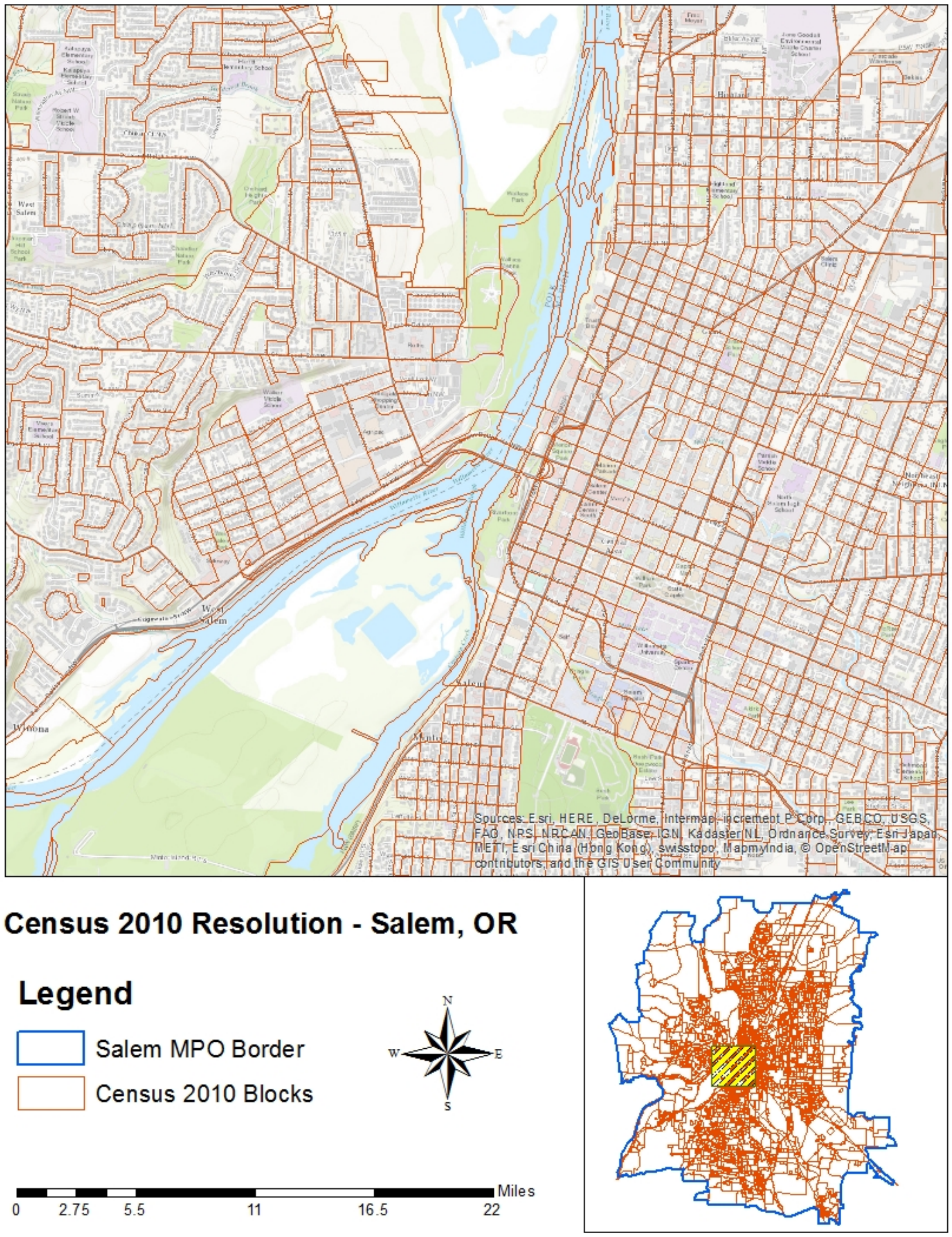

Figure 4-5: Census 2010 block sizes in Salem

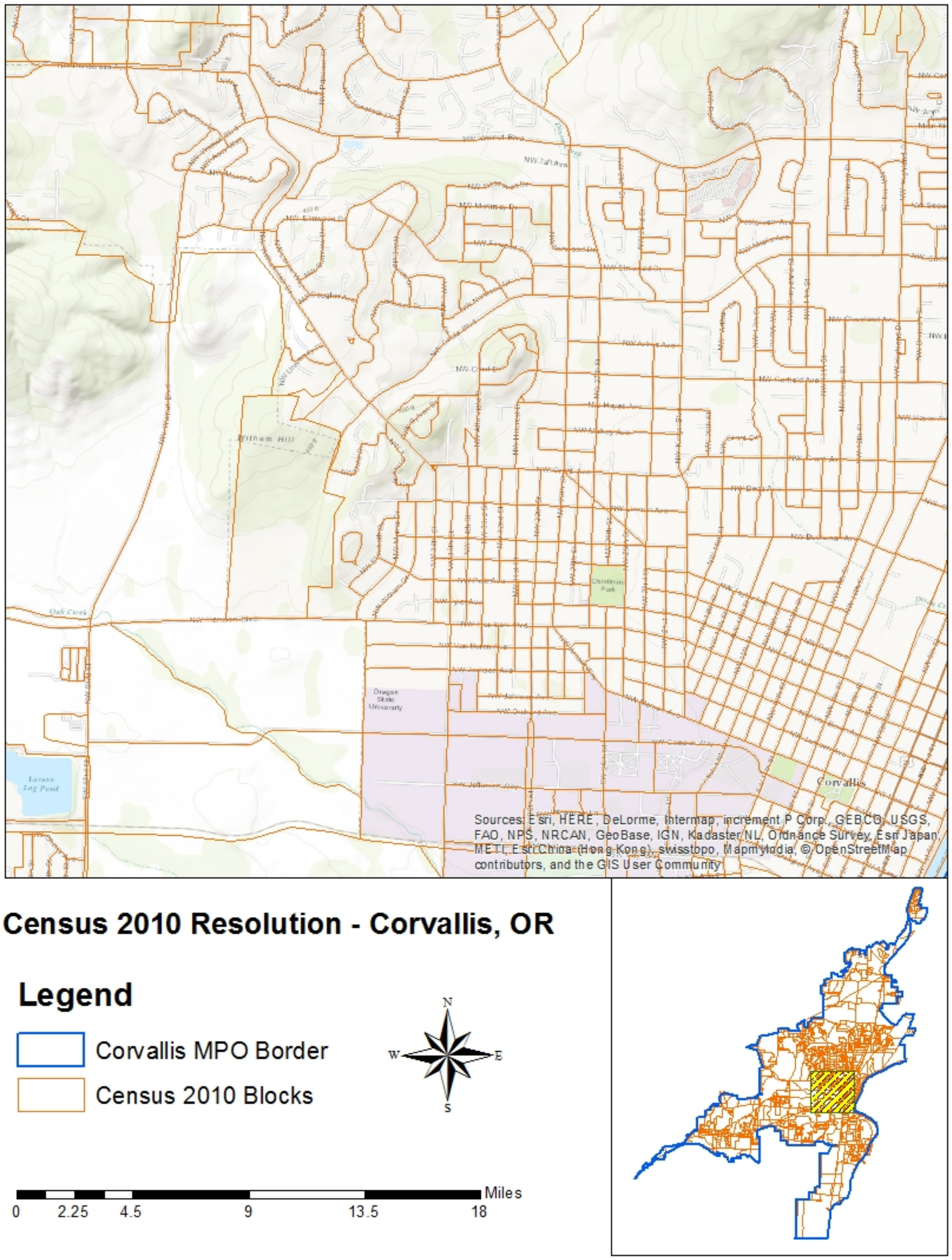

Figure 4-6: Census 2010 block sizes in Corvallis

4.1.2 American Community Survey (ACS) 2014

The American Community Survey is a survey carried out every year and contains detailed information about people and workforce in USA at block group level. The information extracted from this database helps to distribute more than $400 billion federal and state funds every year. Based on the literature review, the following data might be useful for the purpose of this project:

- Population data
- Mode of transportation to work and associated travel times
- Aggregated travel time to work
- Household data
- Income (median, per capita, aggregated, etc.)
- Poverty
- Employment

4.1.3 Vehicle Registration

Oregon DMV reports vehicle registration for vehicles that were registered at the end of each year at county level for 2012 to 2016. Table 4-3 shows the vehicle registration at the county level in Oregon in 2016. The 12 counties in Region 2 are: Lane, Linn, Benton, Lincoln, Polk, Marion, Yamhill, Tillamook, Clatsop, Columbia, southern Clackamas and western Washington counties. About 53% of the total vehicle registrations in Oregon are in these 12 counties. Multnomah and Washington Counties have the highest number of vehicles registrations in Oregon.

Table 4-3: Vehicle Registration by County in 2016

County	Passenger Vehicle	Total Registrations	County	Passenger Vehicle	Total Registrations
Baker	17,127	24,250	Lake	8,585	13,338
Benton	69,782	83,858	Lane	312,643	379,260
Clackamas	357,483	444,758	Lincoln	46,096	55,760
Clatsop	37,298	45,371	Linn	113,511	146,821
Columbia	52,808	67,104	Malheur	25,608	37,019
Coos	60,129	77,203	Marion	280,363	351,539
Crook	25,165	35,984	Morrow	11,436	16,266
Curry	24,649	31,628	Multnomah	564,483	747,430
Deschutes	184,375	233,251	Polk	69,617	83,689
Douglas	107,387	138,338	Sherman	2,386	3,856
Gilliam	2,300	3,661	Tillamook	27,912	35,946
Grant	8,188	11,966	Umatilla	69,440	94,344
Harney	7,675	11,702	Union	25,184	34,405
Hood River	25,423	31,512	Wallowa	8,418	12,271
Jackson	194,229	246,786	Wasco	26,417	33,564
Jefferson	21,397	28,634	Washington	465,791	536,812
Josephine	88,622	108,960	Wheeler	1,656	2,430
Klamath	65,115	87,094	Yamhill	93,210	113,468

ODOT provided vehicle registration information by census blocks for Marion County. This data contains the number of vehicles in Marion County in each block. Figure 4-7 depict this information in part of Marion County.

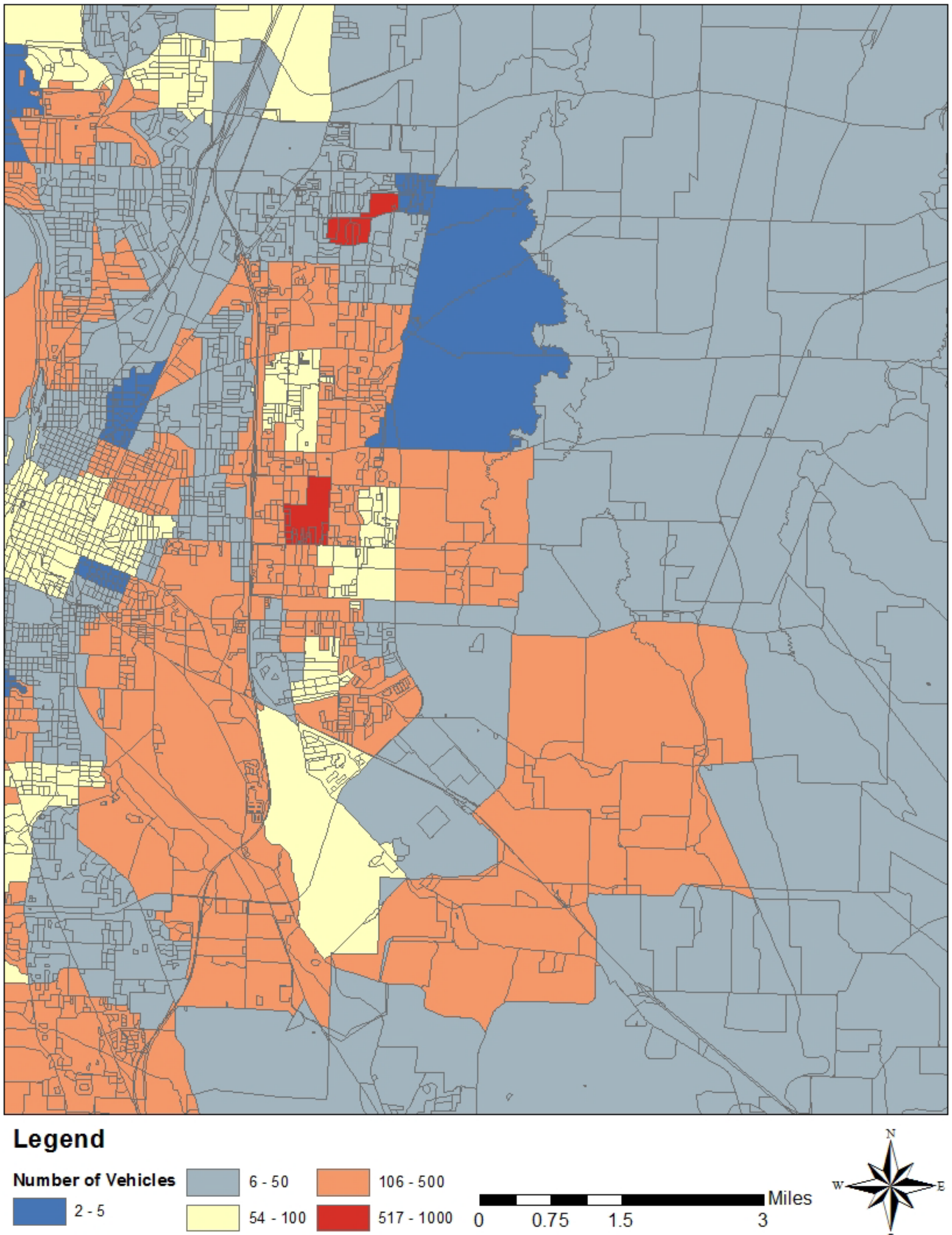

Figure 4-7: Vehicle registration by census blocks in Marion county

4.1.4 Transit Stops

Transit stops are highly correlated with urbanization and population density. The number of transit stops on a route might be related to AADT as transit stops are often located on higher volume corridors. *ODOT GIS Database* covers 59 transit services in Oregon. For example, Corvallis transit system has 387 transit stops in the City of Corvallis and Cherritos public transit system has 722 stops along Salem area. Figure 4-8 and Figure 4-9 shows the transit stops in cities of Salem and Corvallis.

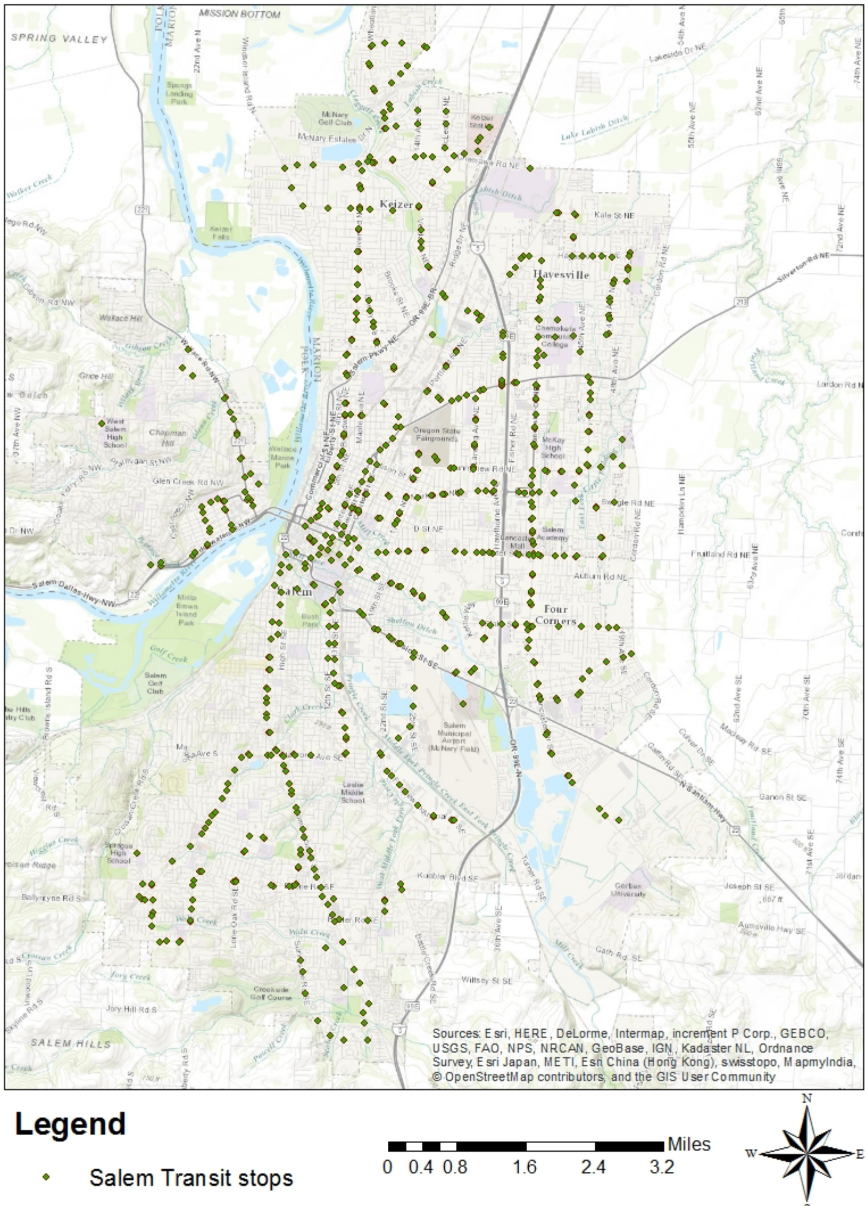

Figure 4-8: Transit stops in the City of Salem

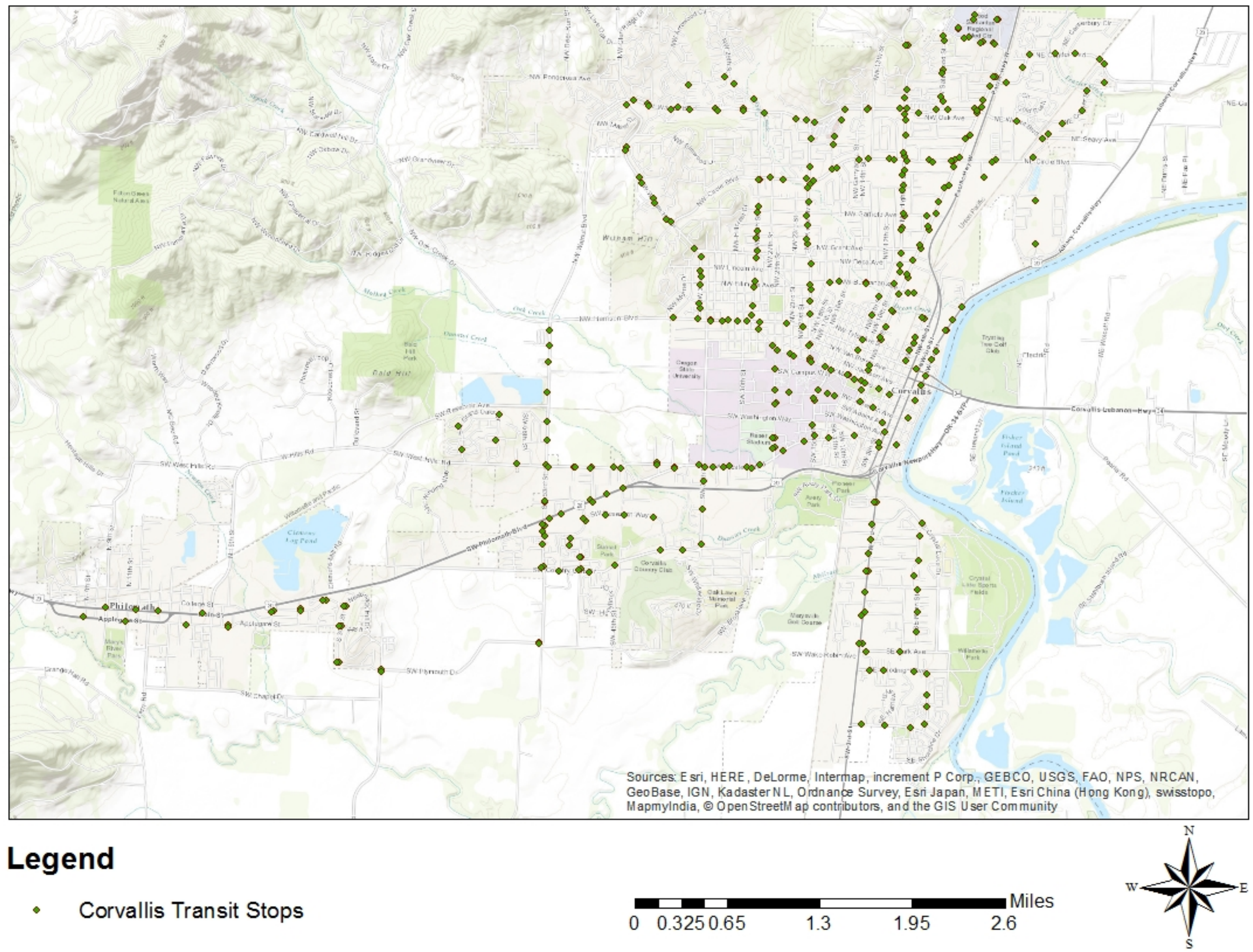

Figure 4-9: Transit stops in the City of Corvallis

4.2 VEHICLE COUNT DATA

The vehicle count data contains ATR counts, and AADT counts for non-state and state roads in State of Oregon.

4.2.1 ATRs in Oregon

Based on the *ODOT GIS Database*, 207 ATRs covered Oregon in 2015, of which 53, 59, 25, 35, and 35 of them are in ODOT region 1,2,3,4, and 5, respectively. Region 2 contains nearly 30% of the ATRs. Figure 4-10 shows the statewide coverage of these ATRs.

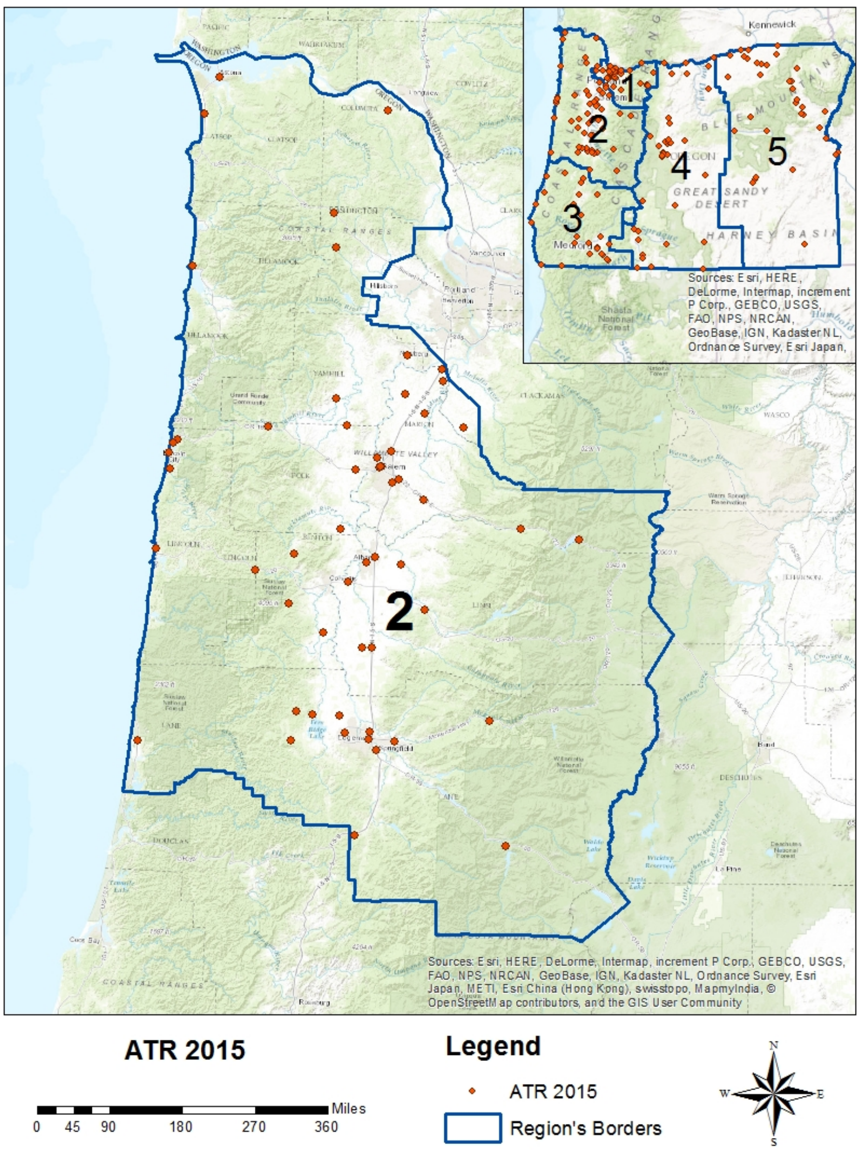

Figure 4-10: 2015 ODOT, Region 2, ATR coverage AADT on state and non-state roads

The *ODOT GIS Database* contains statewide AADT on state roads from 2010 to 2015 and on non-state roads for 2014. These data are based on 24 hours vehicle volumes, classification road tube counts, and manual classification counts. All these counts are adjusted by axle and seasonal factors. Figure 4-11 shows the number of AADT available on state and non-state roads for ODOT regions for 2014. Region 2 has 32% of the AADT counts for state roads and 45% of the AADT counts for non-state roads respectively.

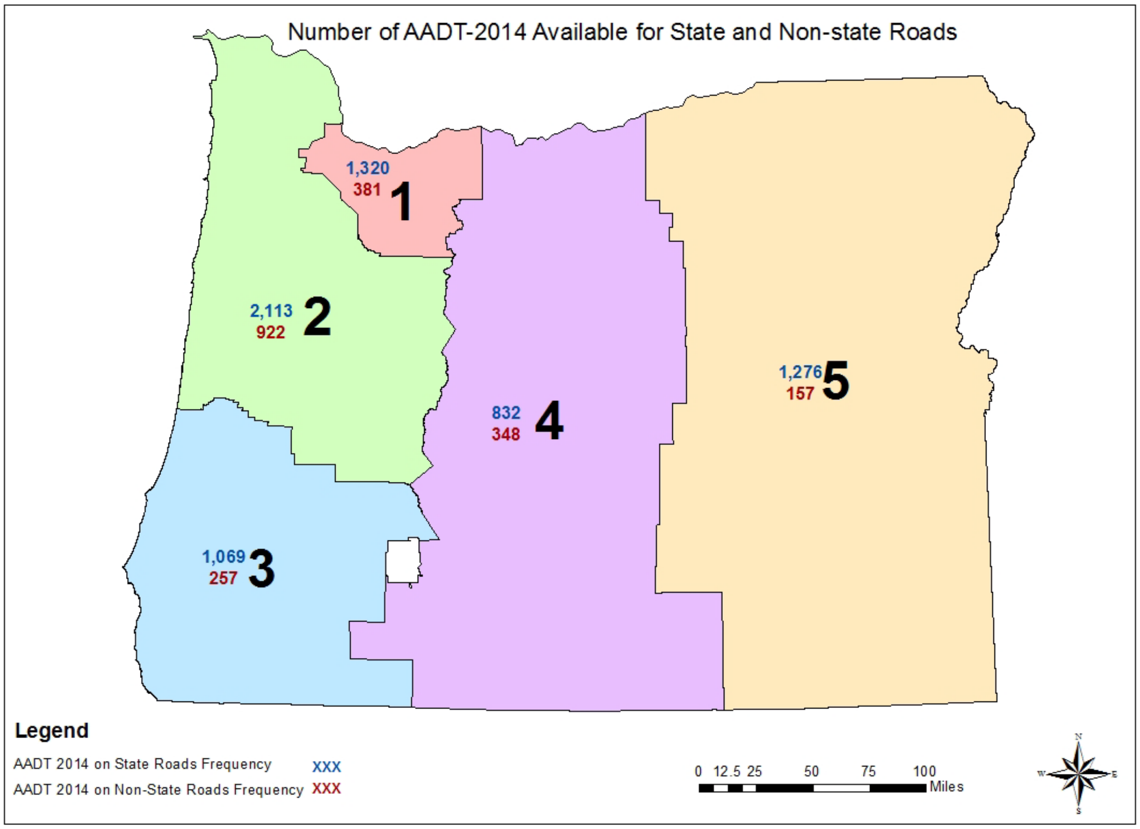

Figure 4-11: 2014 AADT counts on state and non-state roads in Oregon

Figure 4-12 represents the coverage of AADT on non-state roads in ODOT region 2 in 2014.

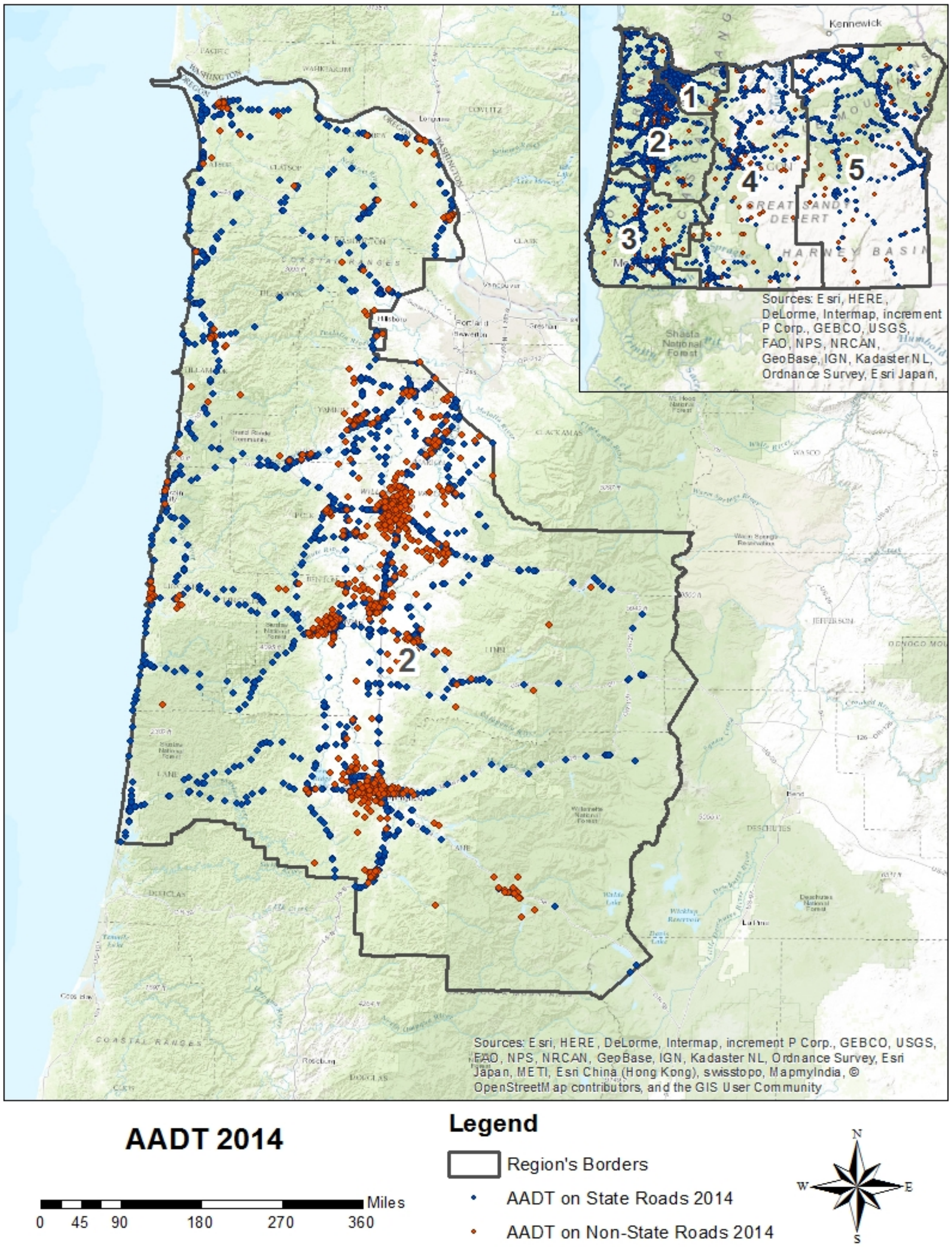

Figure 4-12: 2014 AADT coverage on state and non-State roads in ODOT Region 2

4.3 ROADWAY DATA SOURCES

4.3.1 Oregon State Roads Network

ODOT GIS Database contains statewide information on 3,259 road segments counted as state roads. Figure 4-13 shows the state roads in Oregon in 2016. Information is available on functional classification (urban interstate, rural local, rural interstate, rural major collector, urban minor collector, etc.), number of lanes, lane width, posted speed, median type (vegetation, barrier, painted, curbed, gravel, jiggle bars, etc.), right turn lane width, traffic barrier type, pavement condition (fair, good, very good, etc.) and both left and right shoulder type and width.

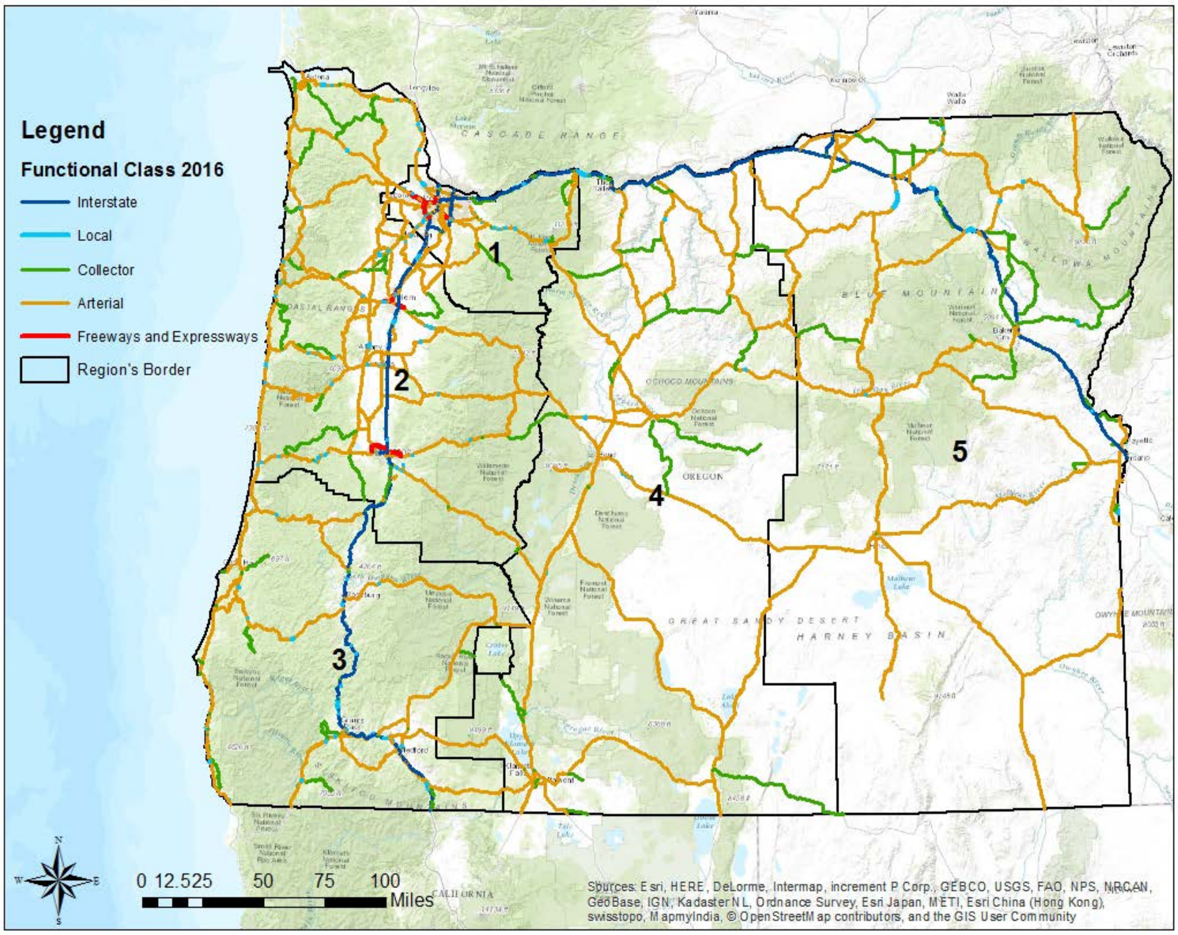

Figure 4-13: State roads in Oregon 2016

4.3.2 Oregon Non-State Roads Network

This information is available through *ODOT GIS Database* (ftp website) and *ODOT TransGIS* website. *ODOT GIS Database* contains statewide information about functional classification on

71,580 road segments that are counted as non-state roads. Figure 4-14 shows the non-state roads that are covered in this dataset for region 2. Salem has about 840 total centerline miles of streets and according to this database 316 miles of them have the functional classification available, which is about 38% of the roads (City of Salem, 2016). Note that the *ODOT TransGIS* website has information on additional roadway segments, which are not available in the *ODOT GIS Database* (ftp website).

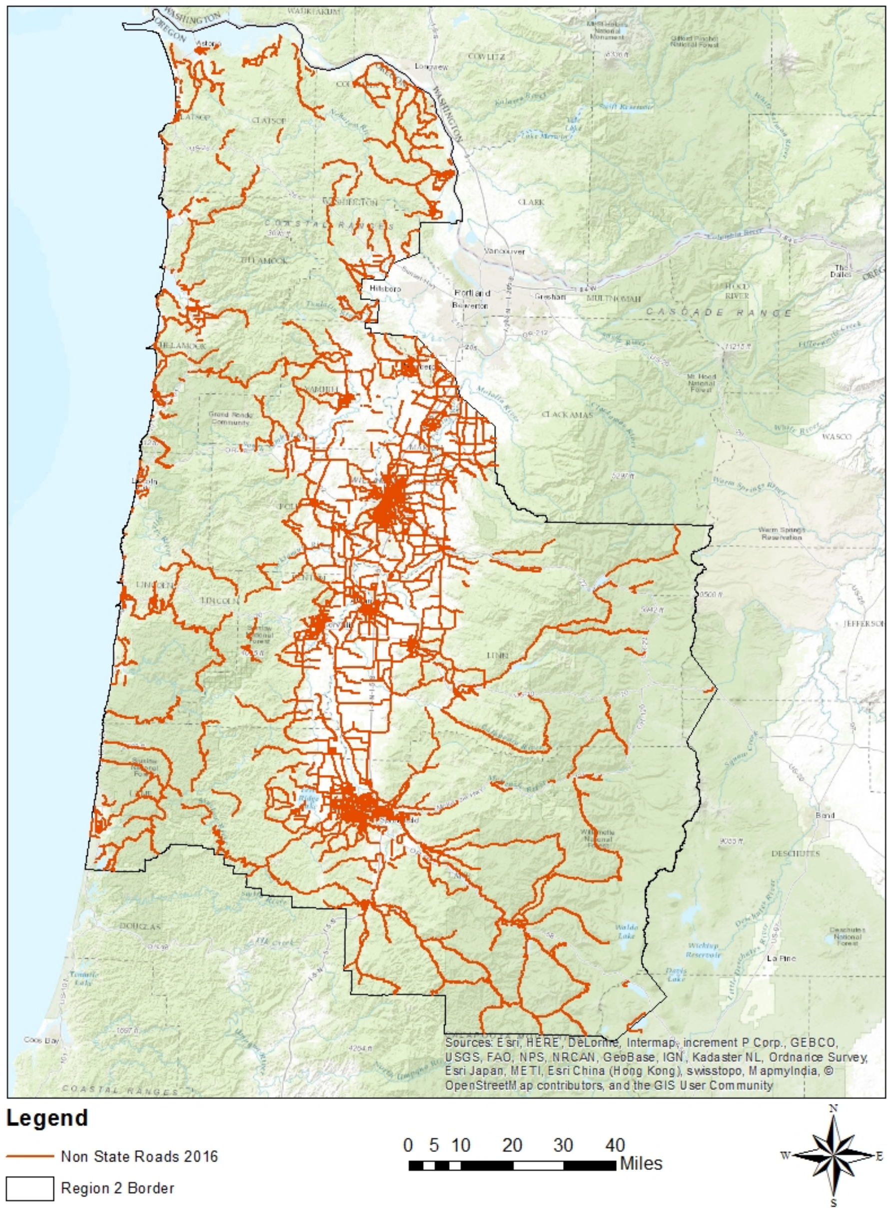

Figure 4-14: Non-state roads in Oregon 2016

4.4 ZONING, MPO AND COUNTY DATA

4.4.1 Oregon Zoning

Oregon Spatial Data library contains a comprehensive database on land use data from 169 local jurisdictions. Upon contacting the responsible party for this database, we were informed that it had to be completed and posted on their website by the end of January 2017. There are 13362, 56500, 9528, 8058, 21718 parcels at ODOT region 1, 2, 3, 4, and 5 in this data set. The land use is categorized into 60 groups: Rural commercial, industrial light, public and semipublic uses, very low density residential, exclusive farm use with more than 20 acres, prime forest with more than 80 acres, heavy industrial, parks and open space, etc.

For region 2, the smallest parcel area in this dataset is near 0 and the largest parcel is about 2,529 square mile. The average parcel area is 0.3 square mile. Residential land use type has the highest number of blocks (38,770) and farm land use types covers the largest area (12,373 square miles) in region 2. Table 4-4 shows the mean, 15th, 50th, and 95th percentiles of block sizes (areas) for region 2. Since most of the parcels in this data set are smaller than 1 square mile, Figure 4-15 gives better intuition of what is the distribution of majority of blocks.

Table 4-4: Mean, 15th, 50th, and 95th Percentiles of Block Sizes of Zoning Database for Region 2

Mean	15th Percentile	50th Percentile	95th Percentile
0.2548	0.0002	0.0004	0.2276

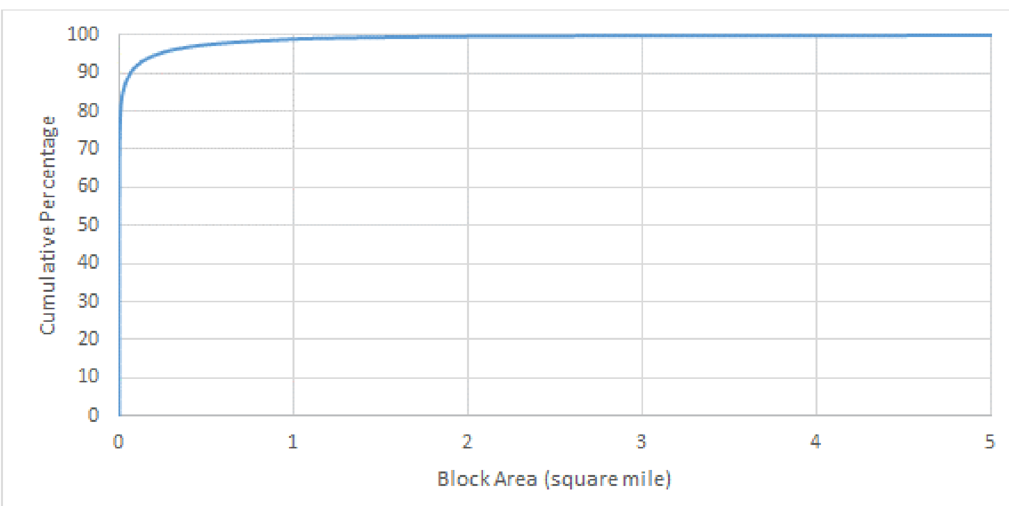

Figure 4-15: Cumulative percentage for land use by parcels for Region 2 up to 5 square miles

Table 4-5 shows the smallest, largest, average, and median parcel size and the most dominant land use type in each region in this data set. The dominant land use is presented by both number of parcels and area.

Table 4-5: Resolution of Land Use Dataset

Region	Parcel Area (Square Mile)				Most Dominant Land Use Type		2nd Most Dominant Land Use Type	
	Smallest	Largest	Average	Median	By Number of Parcels	By Area	By Number of Parcels	By Area
1	0	1,356	0.2301	0.0060	Residential	Forest and Farm Use	Mixed Used Commercial and Residential	Residential
2	0	2,529	0.2548	0.0004	Residential	Forest and Farm Use	Commercial	Rural Residential
3	0	2,529	0.2548	0.0004	Residential	Forest and Farm Use	Commercial	Rural Residential
4	0	1,596	2.2295	0.0079	Residential	Forest and Farm Use	Rural Residential	Rural Residential
5	0	8,542	1.3101	0.0079	Federal Range	Forest and Farm Use	Forest and Farm Use	Federal Range

Figure 4-16 and Figure 4-17 show the coverage of this dataset on the small areas in cities of Corvallis and Salem. Residential is the dominant land use type in both cities.

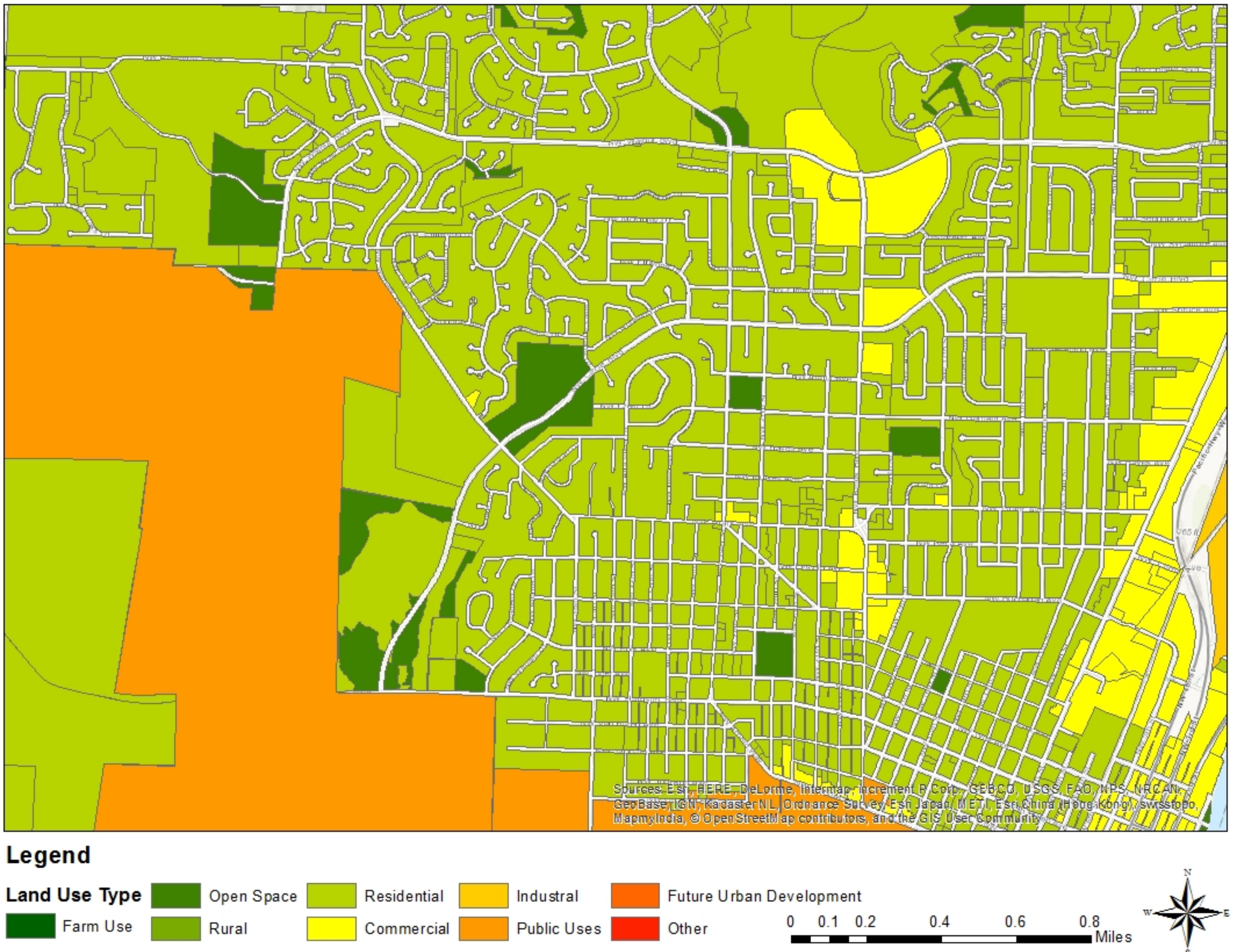

Figure 4-16: Land use data coverage in Corvallis

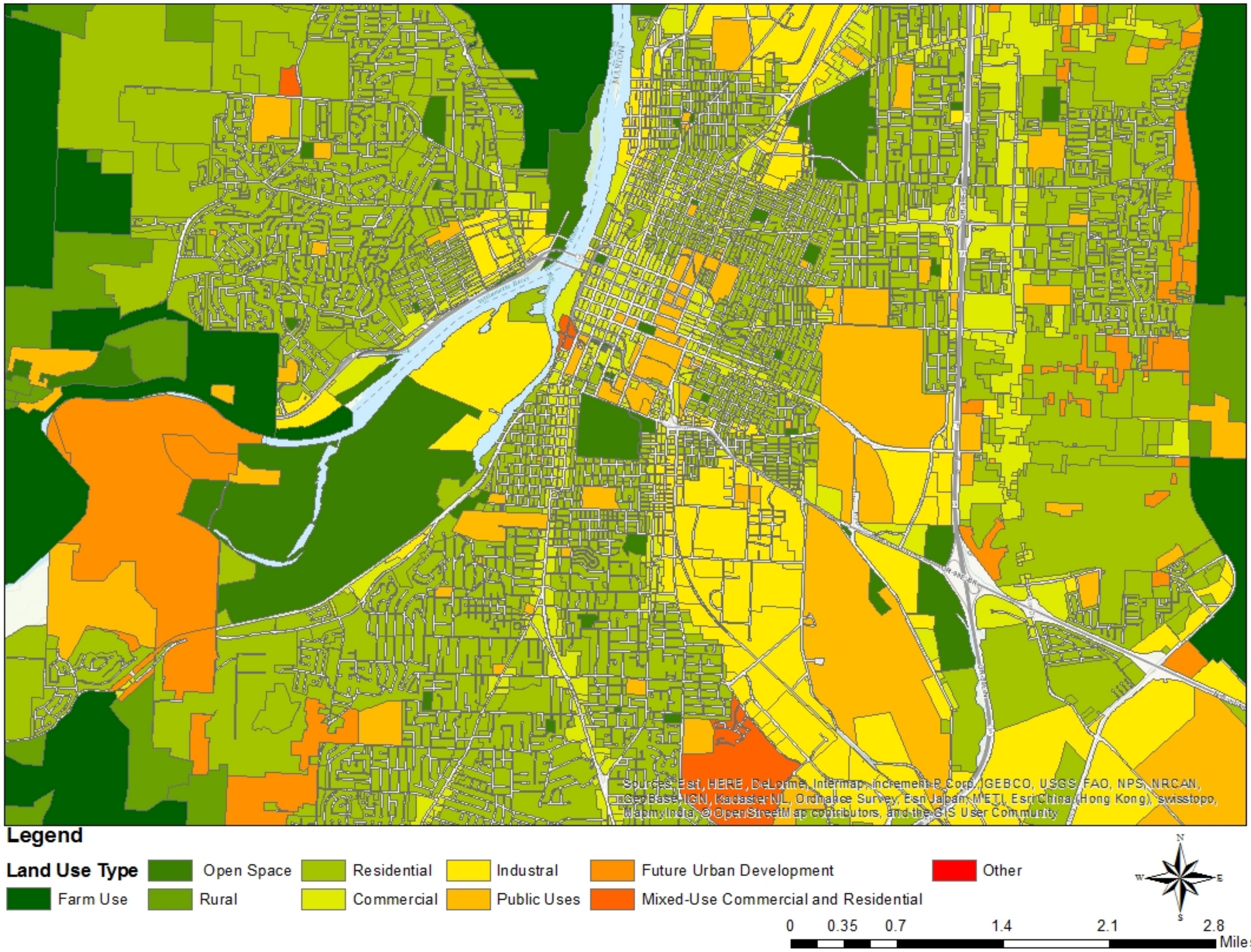

Figure 4-17: Land use data overage in Salem

4.4.2 Metropolitan Planning Organizations (MPOs)

According to the Federal-Aid Highway Act of 1962, any urbanized area with a population of greater than 50,000 are required to have an MPO. An MPO is a decision-making organization made up of representatives from local government and authorities that is funded by the federal government. MPOs are required to have a Travel Demand Model (TDM). The volumes from the TDM can be used to estimate the AADT.

There are 10 MPOs in the state of Oregon: Bend, metropolitan area of Portland (Multnomah, Clackamas and Washington County), Medford, Longview/Kelso/Rainier, Salem, Albany, Middle Rogue, Walla Walla Valley, Eugene, and Corvallis.

4.4.3 County Level Data

To see if there are any data sources available on counties that are not shared anywhere else, we contacted relevant department or organization to see if they have any sort of data set for their counties that might be useful in this project. Table 4-6 shows the summary of their responses.

Table 4-6: County Responses

County	Response
Benton	provided a GIS based website on zoning data (they said they will check to see if they can give us any GIS database + ADT based on 7-day tube counts (shapefile + excel spreadsheets)
Clackamas	will send county-wide traffic counts soon
Columbia	sent an excel file on their traffic counts, GIS based website on land use
Curry	sent an excel file on their traffic counts from 1997 to 2014
Deschutes	try to count their arterials and collectors every 3-5 years
Douglas	emailed parcel level land use shapefile
Grant	land use maps but not documented
Jackson	traffic volumes from 2002 to 2016 on their roads
Josephine	land use data at parcel level
Multnomah	traffic counts on various roads
Umatilla	annual counting program on their roads
Wasco	7+ years old traffic count data

4.5 SUMMARY

The research team gained the following insights from analyzing the data sources.

- Socio-demographic data is available from the Census and ACS 2014 database at the block level.

- Land use data is available at the parcel level.

- Vehicle registration data is available at the county level for the entire state, which may be too aggregated for the purposes of this research. In Marion County vehicle registration data is available at the block level. This will be useful if the case study chosen in region 2 is in Marion County and if the vehicle registration data is available in a database that can be linked to other data sources.

- Travel demand models are available for the urban areas included in MPOs. Currently, we have not analyzed the travel demand models for their network coverage and aggregation level. However, during the region 2 case study, if the area of interest lies in an urban area, we will look at the corresponding travel demand model to study if the network aggregation level is appropriate enough to use for AADT estimation.

- The *ODOT GIS Database* (downloaded from ftp website) has roadway characteristic information for all state roads. For non-state roads, functional classification is available for roads in the *ODOT GIS Database*. Note that there are roadway segments which are not present in the *ODOT GIS Database* (downloaded from ftp website) with functional classification which is available in the *ODOT TransGIS* website. The research team is likely to need access to this additional data source.

Table 4-7 summarizes the data sources discussed in this chapter.

Table 4-7: Summary of Data Sources

Data	Source	Resolution
AADT	*ODOT GIS Database*	2065 and 6610 locations on non-state and state roads, respectively
Functional Classification	*ODOT GIS Database*	all state roads and roughly on 50% of non-state roads in Salem through the *ODOT GIS Database* (downloaded from ftp website)
Lane width	*ODOT GIS Database*	all state roads in Oregon
Number of lanes	*ODOT GIS Database*	all state roads in Oregon
Median	*ODOT GIS Database*	all state roads in Oregon
ATR	*ODOT GIS Database*	207 ATRs in Oregon in year 2015
Traffic barriers	*ODOT GIS Database*	all state roads in Oregon
Right turn lane width	*ODOT GIS Database*	all state roads in Oregon
Poverty	ACS 2014	block group level – statewide
Employment	ACS 2014	block group level – statewide
Income	ACS 2014	block group level – statewide
Population	Census 2010, ACS 2014	block group level – statewide
Households	Census 2010, ACS 2014	block group level – statewide
Land use	OR Spatial Data library	109333 parcels – statewide
Vehicle Registration	*Oregon DMV*	at county level, in Marion county at a block level

Table 4-8: Linkages between Variables found in the Literature Review and relevant Data Sources in Oregon

Category	Variables used in the Literature Review	Oregon Data Source
Sociodemographic Data	county population, distance to mean centers of population, population data aggregated based on buffer distances, population within incorporated area, population density, population accessibility index, population at Census block	Census
	automobile ownership	available at county level through *Oregon DMV* and block level in Marion County
	service employment, agricultural employment, transportation employment, accessibility to employment centers, employment data aggregated based on buffer distances, labor force, employment at census block	ACS
	total number of households	Census
	number of dwelling units, number of manufacturing houses aggregated at buffers of various distances, area of manufactured homes and mobile parks, area of major educational, medical, government, cultural, or religious institutions	-
	percentage of population change over years, county population annual growth rate	-
	median time to leave for work	-
	number of persons working outside of county	-
	number of buildings, area of each building	-
Roadway Characteristics	arterial mileage, total lane mileage of highways	*ODOT GIS Database* (all state roads and a collection of some non-state roads) + *Oregon TransGIS* (on all roads statewide)
	accessibility to state highway, accessibility to freeways, shortest distance from the count location to interstates and major US highways	*ODOT GIS Database*

number of lanes, number of lanes in downstream cross street	limited to the roads available through *ODOT GIS Database* and *Oregon TransGIS* (missing data on some minor roads)
functional classification	limited to the roads available through *ODOT GIS Database* and *Oregon TransGIS* (missing data on some minor roads)
median type	*ODOT GIS Database* (limited to state roads)
speed limit, upstream link speed limit, downstream link speed limit	*ODOT GIS Database* (limited to state roads)
whether the roadway section is a through or destination street	-
latitude, longitude	*ODOT GIS Database* and *Oregon TransGIS*
whether the cross street is a minor arterial or not, whether the cross street is a major collector or not, number of downstream cross streets	can be determined if cross-street has functional classification which is limited to the roads available through *ODOT GIS Database* and *Oregon TransGIS* (missing data on some minor roads)
presence of a right turn lane on the minor road, presence of a right turn lane on the major road, presence of left turn lane	*ODOT GIS Database* (limited to state roads)
presence of a centerline on the minor road, presence of striped edgelines on the minor road, presence of edge stripping, presence of center striping	-
indicator variable for urban minor arterial, principal arterials, local street and collector	can be determined if roadway section has functional classification which is limited to the roads available through *ODOT GIS Database* and *Oregon TransGIS* (missing data on some minor roads
connectivity to a city or town within two miles	*ODOT GIS Database*

	number of principal arterials within one mile, number of freeways within 2 miles, number of major collectors within 2 miles	ODOT GIS Database
	accessibility to Parking lot	-
	direct access to expressway/freeway, road network	*ODOT GIS Database* and *Oregon TransGIS*
	roadway curvature	can be computed state wide (all the roads are available through *Oregon TransGIS* and *ODOT GIS Database*)
	road capacity	-
	connectivity importance index	can be computed with land use information and network topology information from *ODOT GIS Database*
	presence of the link in CBD	-
Economic	personal income, percentage of people below poverty line, median household income, unemployment rate, per capita income	ACS
	retail sales	-
Land Use	area type (rural or urban)	*Oregon Spatial Data Library*
	land use type	*Oregon Spatial Data Library*
	number of agricultural farms, percentage of farms with 500 acres or more	*Oregon Spatial Data Library*
	Is the adjacent land developed or not, distance to MSA	*Oregon Spatial Data Library*
	residential properties, commercial properties	*Oregon Spatial Data Library*
Traffic	median travel time to work	-
	AADT estimates, ATR counts, current AADT, short term counts	*ODOT GIS Database* (limited to ATR and short term traffic count locations available through *ODOT GIS database*)
	daily average probe count	-
	number of vehicles observed in an image of highway, number of cars on the road, car intensity	Google Map

5.0 POINT BASED MODEL FOR NON-STATE UPPER FUNCTIONAL CLASSIFICATION ROADWAY SEGMENTS

This chapter presents the point based model developed for predicting AADT on non-state upper functional classification roadway segments as a function of the roadway, land use, and geometric characteristics. We provide descriptive statistics for the non-state database used for model development. The data collection procedure is then described. The relationships among AADT and various roadway, land use, and signage characteristics are studied followed by the description of the final model developed.

5.1 NON-STATE DATABASE DESCRIPTION

The ODOT GIS database contains estimated AADT on non-state roads for 2014 and 2015. These data are based on 24 hours vehicle volumes, classification road tube counts, and manual classification counts and are adjusted by the axle and seasonal factors. Based on discussions with the Technical Advisory Committee (TAC) and preliminary analysis of the database the following decisions were made:

- The TAC recommended the use of the latest AADT database. After reviewing the location information, the 2015 database was found to be more accurate than the 2014 database. We used the 2015 database for this analysis. The 2015 database had 1147 points originally.

- Close to 88% of the data points had AADT values lower than 10000. Based on discussions with the TAC, we considered only those roadway segments with AADT values lower than 10000.

- Roadway functional classifications corresponding to freeways and expressways were eliminated from the dataset as regular counting programs should cover them.

- Data points corresponding to local roads were excluded from this analysis. Local roads are analyzed separately in the next chapter.

The final, “cleaned”, dataset had 990 points or observations. Table 5-1 shows the overall distribution of AADT in each functional classification. Since the dataset had a high degree of skew and outliers, we recommend using the median as the measure of central tendency or the statistic which best describes the central point of the dataset. As expected, the median AADT increases from 470 for a rural minor collector to 7700 for urban principal arterials.

Table 5-1: Descriptive Statistics of AADT in each Functional Classification in Region 2, ODOT

Functional Classification	Number of Observations	Min AADT	Max AADT	Average AADT	Median AADT
Rural Minor Collector	88	2	2200	618	470
Rural Major Collector	105	60	4700	1509	1300
Rural Minor Arterial	22	540	8400	3688	3200
Urban Minor Collector	141	50	7000	1398	1100
Urban Collector	380	80	9800	2581	2200
Urban Minor Arterial	241	10	9900	5475	5800
Other Urban Principal Arterial	13	920	9800	6855	7700
All Observations	990	2	9900	2910	2100

Next, we classified ODOT Region 2 into Coastal, Mountain, MPO, and Valley-rural sub-regions. Coastal sub-region is the narrow region along the Pacific coast. The Mountain sub-region covers the Coastal range to the west of Salem and Cascade Range to the east of Salem. A topographic map was used to identify the Mountain sub-region. The Willamette Valley in Region 2 in-between the two mountain ranges were divided into MPO and Valley-rural sub-region. The MPO sub-region borders are available through the ODOT GIS database. The Valley-rural sub-region corresponds to all non-MPO areas in the Willamette Valley in Region 2. Using ArcMap, shapefiles were created for each of these sub-regions (except MPO that was available through the ODOT GIS database). These shapefiles were used to identify the sub-regions associated with each AADT data point. Figure 5-1 shows the data points in the four sub-regions.

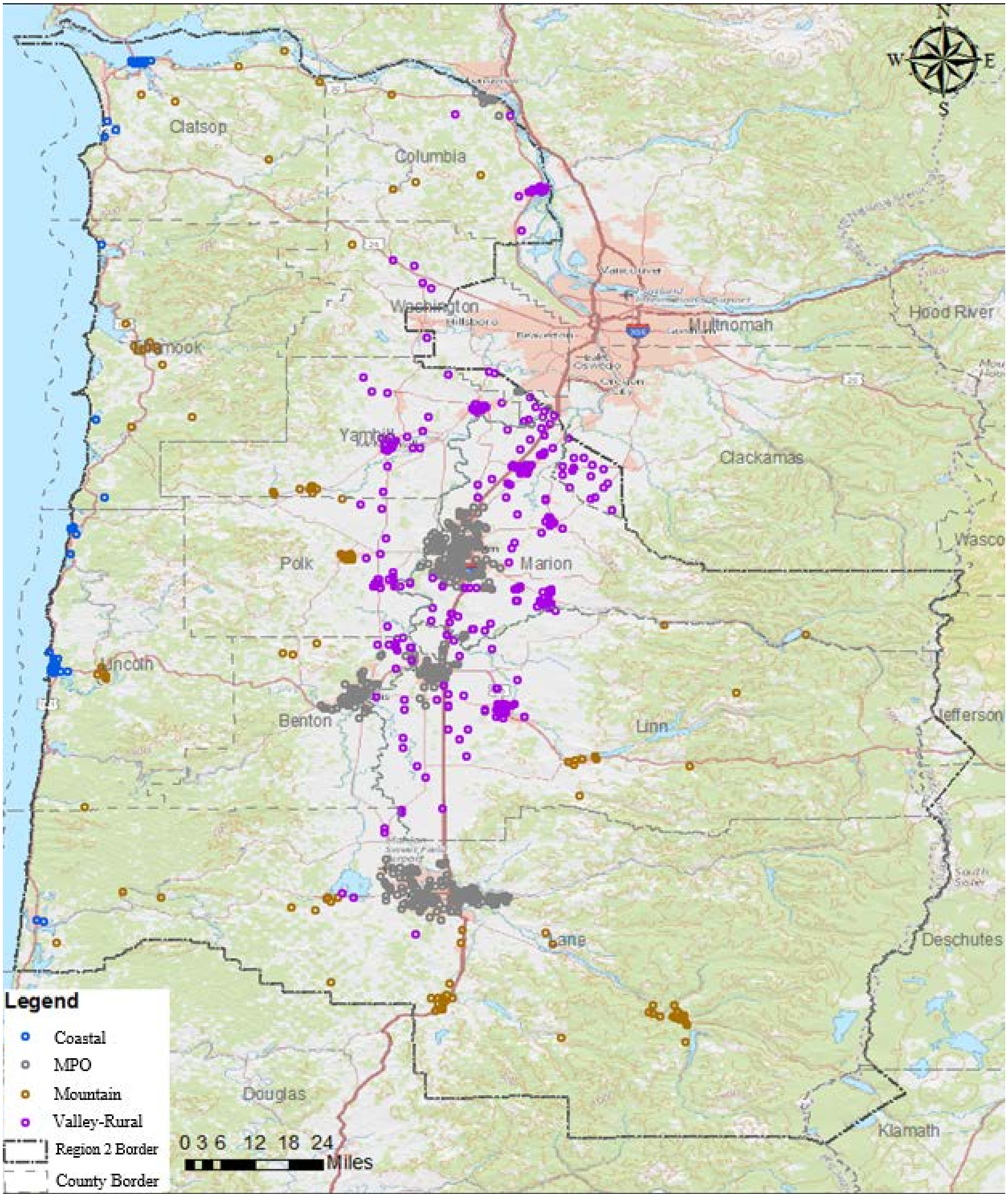

Figure 5-1: 2015 AADT data points by sub-regions

Table 5-2: Descriptive Statistics of AADT in the four Sub-regions in ODOT Region 2

Region	Number of Observations	Min AADT	Max AADT	Average AADT	Median AADT
Mountain	88	2	7500	1265	1050
Coast	46	80	9800	2062	1700
Valley-Rural	362	60	9800	2416	1800
MPO	494	50	9900	3643	3000

Table 5-2 shows the distribution of AADT in the four sub-regions. As expected, the Mountain sub-region has the lowest median AADT whereas the MPO sub-region has the highest median AADT. To get a better estimate of the default AADT values, we also determined descriptive statistics for roadway segments classified by sub-region and their functional classifications (see Table 5-3, Table 5-4, Table 5-5, and Table 5-6). The median AADT can be used as default AADT values except in the cases where the number sample points are low (< 5).

Table 5-3: Descriptive Statistics of AADT in the Coastal Sub-region of ODOT Region 2 by Functional Classification

Functional Classification	Number of Observations	Min AADT	Max AADT	Average AADT	Median AADT
Urban Minor Collector	12	80	2600	1119	1200
Urban Collector	31	80	9800	2303	1900
Urban Minor Arterial	3	2500	3800	3333	3700

Table 5-4: Descriptive Statistics of AADT in the Mountain Sub-region of ODOT Region 2 by Functional Classification

Functional Classification	Number of Observations	Min AADT	Max AADT	Average AADT	Median AADT
Rural Minor Collector	21	2	1200	328	110
Rural Minor Arterial	1	540	540	540	540
Rural Major Collector	24	120	3000	1006	700
Urban Minor Collector	14	210	1500	946	960
Urban Collector	21	180	4000	1956	1800
Urban Minor Arterial	7	10	7500	3630	3100

Table 5-5: Descriptive Statistics of AADT in the Valley-rural Sub-region of ODOT Region 2 by Functional Classification

Functional Classification	Number of Observations	Min AADT	Max AADT	Average AADT	Median AADT
Rural Minor Collector	67	70	2200	708	540
Rural Major Collector	79	60	4700	1664	1400
Rural Minor Arterial	20	1800	8400	3875	3450
Urban Minor Collector	27	140	3000	1139	870
Urban Collector	110	210	7200	2460	2150
Urban Minor Arterial	59	980	9800	5371	5100

Table 5-6: Descriptive Statistics of AADT in the MPO Sub-region of ODOT Region 2 by Functional Classification

Functional Classification	Number of Observations	Min AADT	Max AADT	Average AADT	Median AADT
Rural Major Collector	2	610	2200	1405	1405
Rural Minor Arterial	1	3100	3100	3100	3100
Urban Minor Collector	88	50	7000	1588	1400
Urban Collector	218	90	9500	2742	2200
Urban Minor Arterial	172	310	9900	5624	6000
Other Urban Principal Arterial	13	920	9800	6855	7700

5.2 DATA COLLECTION FORM

A key insight from the literature review was that roadway and geometry related variables were more important in predicting AADT than land use and socio-demographic variables. For example, *Xia et al. (1999), Zhao and Chung (2001), Pan (2008), Yang and Wang (2014), Anderson et al. (2006), Zhao and Park (2004), Eom et al. (2006), Musunuru et al. (2017), Selby and Kockelman (2013), Lowry (2014), Pulugurtha and Kusum (2012), Keehan et al. (2017),* and *Kusam and Pulugurtha (2016)* found number of lanes to be an important variable. Functional classification was used in models to predict AADT by *Xia et al. (1999), Zhao and Chung (2001), Anderson et al. (2006), Barnet et al. (2015), Eom et al. (2006), Musunuru et al. (2017), Lowry (2014), and Keehan et al. (2017).* Moreover, roadway and geometry related variables are easier to obtain than land use and socio-demographic variables. Therefore, in this research, we developed a data collection form (see Appendix B) to collect relevant geometric and roadway related characteristics. Based on discussions with the TAC, the Model Inventory Roadway Elements (MIRE) fundamental data elements were considered as a separate category. The data collection was conducted using Google Street View and the ODOT GIS database. The explanation of all the variables collected is given below:

- Land Use
    - Generator: The primary AADT generator adjacent to the road with the main access on the study road segment, 1000 feet upstream and downstream of the midpoint (gas station, hospital/medical center, school, recreational facility, shopping center, or none)
    - Dominant Land use: Primary dominant land use adjacent to the study road segment, 1000 feet upstream and downstream of the midpoint (residential, commercial, industrial, forest, farm, or other)
    - Single House: Presence of single house residences adjacent to the study road segment
- MIRE
    - Median: Type of median (undivided, one-way, single lane, vegetation, flush, raised, depressed, two-way left turn, transit, or other)
    - One/Two-Way: Whether the road segment operates as one-way or two-way
    - Number of Through Lanes: Number of through lanes
    - Access Road: Degree of access control on the study road segment (full access, partial access, or no access)
- Intersection
    - Right-Turn: Presence of right turn lane on the study road segment, 1000 feet upstream and downstream of the midpoint
    - Left-Turn: Presence of left turn lane on the study road segment, 1000 feet upstream and downstream of the midpoint
    - Traffic Signal: Presence of traffic signal on the study road segment, 1000 feet upstream and downstream of the midpoint (on any side of intersections along the main corridor of study)
- Roadway
    - Paved: Variable indicating whether the road segment is paved or not
    - Pavement Marking: Presence of pavement horizontal marking along the study road segment (lane marking, shoulder marking, both or none. Also note that bicycle lane markings are not counted as shoulder marking)
    - Shoulder: Presence of shoulder on the study road segment (paved, unpaved, or none)
    - Crosswalk: Presence of crosswalk on the study road segment, 1000 feet upstream and downstream of the midpoint (either crossing the main corridor of study or on any side of intersections along the main corridor of study)
    - Sidewalk: Presence of sidewalk along the study road segment, 1000 feet upstream and downstream of the midpoint (both sides, one side, or none)
    - Bike Lane: Presence of bike lane along the study road segment, 1000 feet upstream and downstream of the midpoint
    - Bus Stop: Presence of bus stop along the study road segment, 1000 feet upstream and downstream of the midpoint

- Parking Lot: Presence of parking lot adjacent to the study road segment, 1000 feet upstream and downstream of the midpoint (including pay to park, parking lots for schools, shopping centers, recreational facilities, hospitals, etc.)
- Calming Device: Presence of traffic calming devices along the study road segment, 1000 feet upstream and downstream of the midpoint (according to ITE, traffic calming devices include speed humps, neighborhood traffic circles, speed tables, chicanes, raised intersection, choker, closure, and center island narrowing)

- Signage
  - Cross Road Stop Sign: Presence of stop sign on the cross roads of the study road segment, 1000 feet upstream and downstream of the midpoint
  - Stop Sign: Presence of stop sign on the main corridor of study, 1000 feet upstream and downstream of the midpoint
  - Sign: Presence of signs, other than stop signs on the main corridor of study, 1000 feet upstream and downstream of the midpoint
- Obtained through the ODOT GIS Data Base
  - Distance to NHS: Distance from the point to the national highway system
  - MPO: Whether the study location is in an MPO or not
  - Distance to state arterial: Distance to the nearest state arterial
  - Distance to state highway: Distance to the nearest state highway
  - Distance to non-state arterial: Distance to the nearest non-state arterial
  - Distance to the non-state highway: Distance to the nearest non-state highway

5.3 DESCRIPTIVE ANALYSIS OF DATA

This section first describes the sampling procedure used to identify the roadway segments for data collection purposes. A descriptive statistical analysis is conducted to get a sense of the relationship between AADT and the variables collected. We only include those variables which were found to have an impact on AADT,

The functional classification had an obvious impact on the AADT. We adopted a stratified random sampling approach to ensure that all functional classifications were represented in the data. The number of data points selected in each functional classification was set to be the maximum of (i) 20% of the data points in each functional classification in the 2015 non-state AADT dataset, and (ii) 30. The roadway segments corresponding to ramps were eliminated from analysis. For certain functional classifications such as rural minor arterial, there were less than 30 data points in the original dataset, and therefore, the number of data points sampled were less than 30. We also sampled additional data points in the Mountain and Coastal sub-region to ensure adequate representation of those regions.

Table 5-7: Descriptive Statistics of AADT by Functional Classification in Sampled Dataset

Functional Classification	Number of Observations	Min AADT	Max AADT	Average AADT	Median AADT
Rural Minor Collector	38	30	2200	729	630
Rural Major Collector	32	120	4700	1437	1300
Rural Minor Arterial	19	540	8400	3497	3100
Urban Minor Collector	38	140	4400	1432	1200
Urban Collector	108	80	9800	2580	2200
Urban Minor Arterial	49	880	9500	5134	4400
Other Urban Principal Arterial	7	5600	9800	7957	7150
All Observations	291	30	9800	2682	2000

Table 5-7 provides the descriptive statistics of AADT by functional classification in the sample data. As expected, rural roads have lower AADT when compared to urban roadway segments. Arterials have higher AADT than collectors. Table 5-8 shows the distribution of sample points for each AADT category. Nearly 80% of the data points have AADT values greater than 250 and lower than 5000.

Table 5-8: AADT Categories by Value

AADT Category	Number of Observations	Relative Frequency (%)
<= 250	17	5.8
251 - 2500	162	55.7
2501 - 5000	69	23.7
5001 - 7500	23	7.9
7501 - 10000	20	6.9

Next, we look at descriptive statistics and study the impact of the different variables collected on AADT.

5.3.1 Land Use Type

Table 5-9 shows the variation of AADT concerning the primary AADT generator adjacent to the road with the main access on the study road segment. As expected, the median AADT was found to increase significantly in the presence of a shopping center, gas station, or a medical center adjacent to the road.

Table 5-9: Descriptive Statistics of AADT by Generator Type

Generator Type	Number of Observations	Min AADT	Max AADT	Average AADT	Median AADT
None	177	30	9800	2047	1300
Recreational Facility	16	880	8500	2793	2400
School	29	610	8900	3158	2700
Shopping Center	54	230	9800	4151	3650
Gas Station	8	830	9100	4416	4400
Hospital/Medical Center	7	790	8000	3199	2600

Table 5-10 shows the variation in AADT with the primary dominant land use adjacent to the study road segment. The residential and commercial dominant land use types had higher AADT when compared to the forest area. A majority of the data points were in residential areas. Roadway segments with adjacent parking lots had higher AADT (see Table 5-11).

Table 5-10: Descriptive Statistics of AADT by Dominant Land Use

Dominant Land Use	Number of Observations	Min AADT	Max AADT	Average AADT	Median AADT
Forest	19	30	2700	664	390
Other	2	1100	1200	1150	1150
Farm	85	150	9200	2008	1300
Industrial	6	90	3100	1687	2050
Residential	129	80	9800	2986	2300
Commercial	50	340	9800	3991	3500

Table 5-11: Impact of the Presence of a Parking Lot on AADT

Presence of Parking Lot	Number of Observations	Min AADT	Max AADT	Average AADT	Median AADT
No	165	30	9800	1815	1200
Yes	126	230	9800	3817	3250

5.3.2 MIRE Fundamental Data Elements

Roadway segments with the two-way left turn, raised, and vegetation median types have higher AADT than those with single line median and one-way roads. Also, undivided roads have the lowest AADT (see Table 5-12).

Table 5-12: Descriptive Statistics of AADT by Median Type

Median Type	Number of Observations	Min AADT	Max AADT	Average AADT	Median AADT
Undivided	21	70	4400	1467	1300
Single-Line	243	30	9200	2392	1800
One Way	3	1900	6700	3767	2700
Vegetation	2	2600	5800	4200	4200
Two Way Left Turn Lanes	22	1300	9800	6759	7450

Nearly 98% of the data had roadway segments with two one-way lanes, were one way, and had no access road. Therefore, the other MIRE Fundamental Data Elements were not useful in understanding AADT variation.

5.3.3 Roadway Characteristics

Roadway segments with traffic signals, left turn, and right turn lanes usually have higher AADT when compared to roadway segments without these characteristics (see Table 5-13, Table 5-14, and Table 5-15). The median AADT of roadway segments with traffic signals was nearly three times those without signals.

Table 5-13: Descriptive Statistics of AADT with the Presence of Right Turn Lane

Presence of Right Turn Lane	Number of Observations	Min AADT	Max AADT	Average AADT	Median AADT
No	240	30	9500	2236	1600
Yes	51	90	9800	4783	4200

Table 5-14: Descriptive Statistics of AADT with the Presence of Left Turn Lane

Presence of Left Turn Lane	Number of Observations	Min AADT	Max AADT	Average AADT	Median AADT
No	214	30	8400	1811	1400
Yes	77	90	9800	5102	4700

Table 5-15: Descriptive Statistics of AADT with the Presence of Traffic Signal

Presence of Traffic Signal	Number of Observations	Min AADT	Max AADT	Average AADT	Median AADT
No	223	30	9400	1949	1400
Yes	68	1300	9800	5085	4700

In general, crosswalks, sidewalks, bike lanes, and bus stops are found in more urban areas with higher AADT. We found that roadway segments with crosswalks, bike lanes or bus stops have nearly three times the median AADT of roadway segments which do not have these characteristics (see Table 5-16, Table 5-17, Table 5-18, and Table 5-19).

Table 5-16: Descriptive Statistics of AADT with the Presence of Crosswalk

Presence of Crosswalk	Number of Observations	Min AADT	Max AADT	Average AADT	Median AADT
No	147	30	8400	1502	1200
Yes	144	230	9800	3887	3400

Table 5-17: Descriptive Statistics of AADT with the Presence of Sidewalk

Presence of Sidewalk	Number of Observations	Min AADT	Max AADT	Average AADT	Median AADT
None	141	30	9200	1909	1300
One side	33	140	9500	2898	2400
Both sides	117	90	9800	3552	3000

Table 5-18: Descriptive Statistics of AADT with the Presence of Bike Lanes

Presence of Bike Lanes	Number of Observations	Min AADT	Max AADT	Average AADT	Median AADT
No	226	30	9400	2090	1500
Yes	65	90	9800	4741	4100

Table 5-19: Descriptive Statistics of AADT with the Presence of Bus Stops

Presence of Bus Stop	Number of Observations	Min AADT	Max AADT	Average AADT	Median AADT
No	224	30	9400	2198	1500
Yes	67	500	9800	4299	4100

Roadway segments with pavement markings were found to have higher AADT when compared to roadway segments with no pavement markings (see Table 5-20). Roadway segments with unpaved shoulders were found to have lower AADT when compared to roadway segments with no shoulders or paved shoulders (see Table 5-21)

Table 5-20: Descriptive Statistics of AADT with the Presence of Pavement Markings

Presence of Pavement Markings	Number of Observations	Min AADT	Max AADT	Average AADT	Median AADT
No	28	70	8000	1708	1300
Yes	223	60	9800	2958	2300

Table 5-21: Descriptive Statistics of AADT with the Presence of Unpaved Shoulder

Presence of Unpaved Shoulder	Number of Observations	Min AADT	Max AADT	Average AADT	Median AADT
Yes	51	60	8000	1569	1100
No	240	30	9800	2919	2200

5.3.4 Signage Characteristics

The median AADT was higher on roadway segments with stop signs on cross roads (see Table 5-22). The presence of stop signs on the intersecting streets often indicates that the roadway segment of interest has higher AADT. The AADT was found to decrease if the roadway segment of interest had stop signs which is again reasonable as for upper functional classification roads, the presence of stop signs instead of signals might indicate lower AADT (see Table 5-23). Roadway segments with signs other than stop signs were found to have higher AADT (Table 5-24).

Table 5-22: Descriptive Statistics of AADT with the Presence of Stop Signs on Crossroads

Presence of Stop Signs on Crossroads	Number of Observations	Min AADT	Max AADT	Average AADT	Median AADT
No	61	30	9800	2082	980
Yes	230	120	9800	2841	2300

Table 5-23: Descriptive Statistics of AADT with the Presence of Stop Signs on the Roadway Segments

Presence of Stop Signs	Number of Observations	Min AADT	Max AADT	Average AADT	Median AADT
Yes	151	30	9500	1971	1600
No	140	60	9800	3449	2550

Table 5-24: Descriptive Statistics of AADT with the Presence of Signs other than Stop Signs on Roadway Segments

Presence of Signs other than Stop Signs	Number of Observations	Min AADT	Max AADT	Average AADT	Median AADT
No	77	60	6200	1276	890
Yes	214	30	9800	3188	2500

The descriptive statistics give us an idea of which roadway, geometric, and land use related variables are present in road segments with higher AADT and which are not. The trends, in general, are consistent with what we expect. The next step is to develop a model to predict AADT based on this information.

5.4 MODEL DEVELOPMENT

Chapter 2 summarized the various approaches used for predicting AADT along with their advantages and disadvantages. Regression based approaches were identified as the most promising approach. However, we found that several of the regression models were miss-specified, over fitted, or without a proper validation approach. Models which are over fitted based on variables in one region do not perform well when applied to other regions. Robust models that are parsimonious and intuitive are more likely to stand well the test of time and transferability. In this project, from our preliminary analysis, linear regression models were found to be unsuitable for AADT prediction. The statistical assumptions associated with linear regression models were not being satisfied. Moreover, the initial tests on model performance during validation were not good.

With the goal of developing a reliable and simple model, a point based system was developed to predict AADT. Based on the descriptive analysis, we identified the set of the roadway, geometric, and land use related factors which are present in roadways segments with higher median AADT. We selected a subset of these factors and assigned them one point each. The total points are calculated for a roadway segment by adding up all the points. Our model is based on the assumption that higher the points, the greater the median AADT. Hence, we developed a point based model for overall Region 2 and then specific models for each of the sub-regions. We expect that the models for the sub-regions will better predict the AADT. The point based model has two steps. In the first step, we determine the total points for each roadway segment. A default median AADT is provided for each point score. In the second step, adjustment factors are used to get a better overall fit.

5.4.1 Overall Region 2 Model

In the overall Region 2 model, we assign one point for each of the following features present in the roadway segment:

- Functional classification is an arterial
- Within city limits
- Presence of generator (Gas Station, Hospital/Medical Center, School, Recreational Facility, Shopping Center)
- Median is a two-way left turn lane
- Presence of right turn lane
- Presence of left turn lane
- Presence of traffic signal

- Presence of pavement marking
- Presence of crosswalk
- Presence of sidewalk
- Presence of bike lane
- Presence of bus stop
- Presence of parking lot adjacent to the study road segment
- Presence of stop signs on cross roads
- National highway system

The total points of the roadway segment are determined by adding up the points. Table 5-25 presents the median values to be used as the predicted AADT for each point score. For example, if the roadway segment had a two-way left turn lane, right turn lane, traffic signal, crosswalk and sidewalk, the roadway segment has 5 points, and the predicted AADT is 2100.

Table 5-25: AADT variation for each Point Score, Overall Model

Point	Number of Observations	Min AADT	Max AADT	Average AADT	Median AADT
0	3	30	80	60	70
1	28	60	1800	633	565
2	66	120	4400	1178	1000
3	33	330	6000	1922	1400
4,5	48	90	8400	2396	2100
6	15	630	7600	2641	2300
7	19	1100	8000	3058	2800
8,9	41	1300	9200	4373	4200
10	22	880	9500	5513	5450
11	9	2600	9800	6000	6600
12	7	3500	9800	7729	8700

Figure 5-2 shows the difference between predicted and actual AADT as a function of the actual AADT. The figure shows that the AADT on collectors are over predicted and the AADT on arterials are under predicted. Therefore, we develop adjustment factors to reduce the overall error and improve the fit.

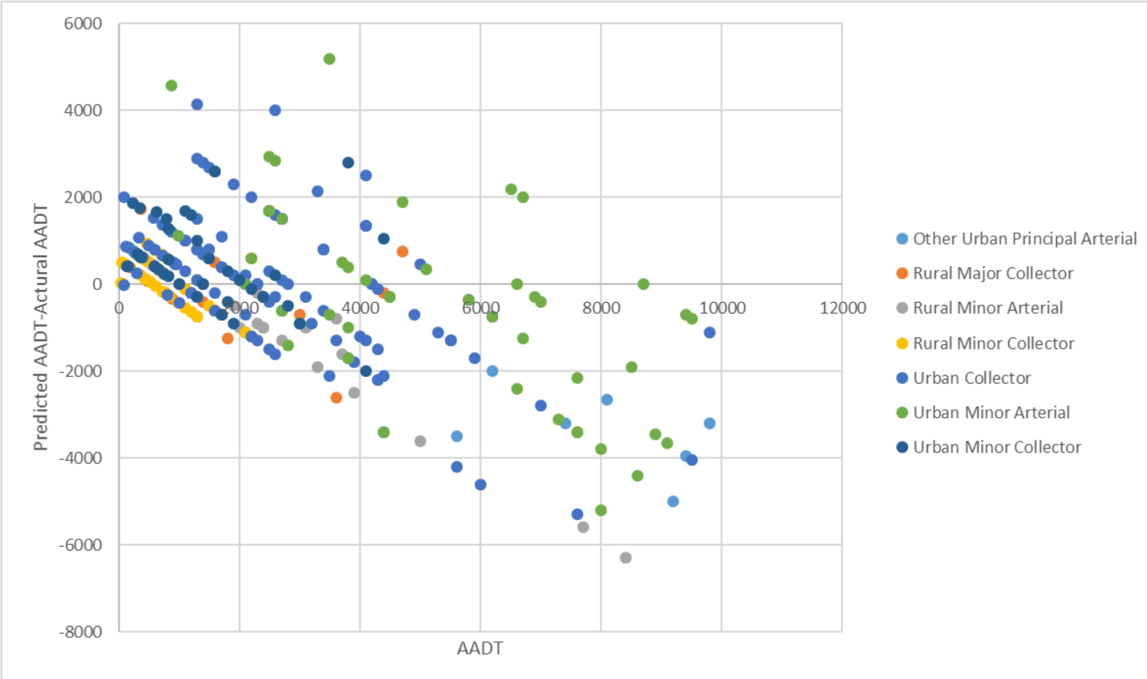

Figure 5-2: Overall model performance based on calibration data set

The idea of using an adjustment factor approach to obtain better AADT estimations can be found in the literature. For example, Figliozzi et al. (2014) proposed a new correcting function, based on analysis of AADT estimation errors, to reduce the data collection and AADT errors. In this research, we tested several adjustment factors. Since the collectors are being over-predicted and arterials are being under-predicted, the adjustment factors should be negative for the collectors and positive for the arterials. The adjustment factors were based on different percentiles of AADT and the median of the error in each functional classification. We compared the performance of the adjustment factors based on median error and average errors (see Table 5-26). Focusing on median error and average error is justified by the goal of the research that aims to develop simple and cost-effective method for missing AADT data. The median of errors in each functional classification to adjust the predictions was found to provide best performance.

Table 5-26: Analysis of Different Adjustment Factors

Percentile of AADT used as Adjustment Factor		10th	15th	20th	25th	30th	40th	50th	Half of 25th	One Fourth of 50th	Median of Error in Each Functional Classification
Adjustment Factors	Rural Minor Collector	-138	-216	-312	-393	-440	-480	-630	-196	-158	-125
	Urban Minor Collector	-302	-334	-362	-610	-700	-830	-1200	-305	-300	-1050
	Rural Major Collector	-371	-387	-430	-468	-637	-1040	-1300	-234	-325	468
	Urban Collector	-556	-743	-970	-1200	-1300	-1600	-2200	-600	-550	-25
	Rural Minor Arterial	1980	2210	2300	2350	2400	2780	3100	1175	775	1300
	Urban Minor Arterial	1720	2180	2500	2600	2760	3660	4400	1300	1100	300
	Other Urban Principal Arterial	1082	2675	4700	5750	6020	6620	7150	2875	1788	4200
Error terms	Median of Errors	-13	-54	-170	-348	-400	-525	-795	-100	-150	0
	Average of Errors	-17	29	32	-55	-92	-79	-240	-106	-178	-32

Figure 5-3 shows that the errors have reduced significantly after applying the adjustment factors. For example, if the roadway segment had a two-way left turn lane, right turn lane, traffic signal, crosswalk and sidewalk, the roadway segment has 5 points and the predicted AADT is 2100 (from Table 5-25) plus the adjustment factor. If the roadway segment was a rural minor collector, the adjustment factor is -125 (from Table 5-26). Therefore, the predicted AADT is 2100 – 125 = 1975. If the roadway segment was a rural minor arterial, the adjustment factor is 1300. Therefore, the predicted AADT is 2100 + 1300 = 3400.

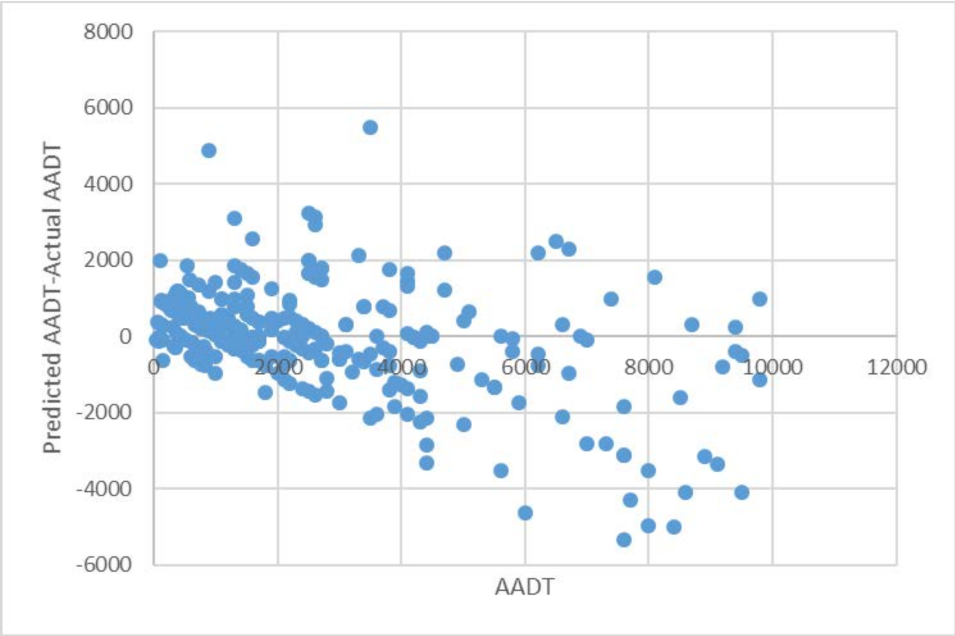

Figure 5-3: Overall Model Performance based on Calibration Data set after using Adjustment Factors

In addition to the overall Region 2 model, we also developed a similar point based model for each sub-region. The variables used in each sub-region model is described next.

5.4.2 Valley-rural Model

In the point based model for the Valley-rural sub-region, we assign one point each if the roadway segment has any of the following features:

- Functional classification is arterial
- Study area type is urban (based on functional classification)
- Within city limits
- Presence of generator (Gas Station, Hospital/Medical Center, School, Recreational Facility, Shopping Center)

- Median is a two-way left turn lane
- Presence of right turn lane
- Presence of left turn lane
- Presence of traffic signal
- Presence of crosswalk
- Presence of sidewalk
- Presence of bike lane
- Presence of bus stop
- Presence of parking lot adjacent to the study road segment

For roads in the Valley-rural sub-region, Table 5-27 provides the median AADT for each point score. After selecting the appropriate median AADT based on the scores, adjustment factors are picked based on functional classification as per Table 5-26.

Table 5-27: AADT variation for each Point Score, Valley-rural model

Point	Number of Observations	Min AADT	Max AADT	Average AADT	Median AADT
0	36	150	3600	911	740
1	20	320	5000	1792	1800
2,3,4	21	140	8400	2754	2100
5,6	15	340	3600	2300	2300
7	8	1500	8000	3713	3450
8,9	11	1400	9100	4909	4500
10,11	6	4100	8500	5733	5450

5.4.3 MPO Model

In the point based model for the MPO sub-region, we assign one point for each of the following features present in the roadway segment:

- Functional classification is arterial
- Presence of generator (Gas Station, Hospital/Medical Center, School, Recreational Facility, Shopping Center)
- Median is a Two-way left turn lane

- Presence of right turn lane
- Presence of left turn lane
- Presence of traffic signal
- Presence of crosswalk
- Presence of bike lane
- Presence of bus stop
- Presence of parking lot adjacent to the study road segment
- No stop sign on study road segment
- National highway system

For roads in the MPO sub-region, Table 5-28 provides the median AADT for each point score. After selecting the appropriate median AADT based on the scores, adjustment factors are applied based on the functional classification as per Table 5-26.

Table 5-28: AADT variation for each Point Score, MPO model

Point	Number of Observations	Min AADT	Max AADT	Average AADT	Median AADT
0	10	120	4400	877	120
1	12	590	2800	1290	970
2,3	23	90	6000	2513	2100
4,5	10	1600	7600	3440	2650
6,7	20	880	8600	4149	3750
8	12	3800	9500	6567	6400
9,10	16	2600	9800	6969	6950

5.4.4 Mountain Model

In the point based model for the Mountain sub-region, we assign one point each if the roadway segment has any of the following features:

- Functional classification is arterial
- Study area type is urban (based on functional classification)
- Presence of generator (Gas Station, Hospital/Medical Center, School, Recreational Facility, Shopping Center)

- Presence of crosswalk
- Presence of sidewalk
- Presence of parking lot adjacent to the study road segment
- Presence of stop sign on cross roads

For roads in the Mountain sub-region, Table 5-29 provides the median AADT for each point score. After selecting the appropriate median AADT based on the scores, adjustment factors are picked based on functional classification as per Table 5-26.

Table 5-29: AADT variation for each Point Score, Mountain model

Point	Number of Observations	Min AADT	Max AADT	Average AADT	Median AADT
0	13	30	1800	598	150
1	7	370	1700	701	430
2,3	8	470	2600	1229	1200
4,5,6	8	630	3000	1665	1500

5.4.5 Coastal Model

In the point based model for the Coastal sub-region, we assign one point each if the roadway segment has any of the following features:

- Presence of generator (Gas Station, Hospital/Medical Center, School, Recreational Facility, Shopping Center)
- Presence of right turn lane
- Presence of left turn lane
- Presence of crosswalk
- Presence of bus stop
- Presence of parking lot adjacent to the study road segment
- Presence of stop sign on cross roads

For roads in the Coastal sub-region, Table 5-30 shows the predictions of AADT for each point. For the Coastal sub-region, use of adjustment factors was found not to improve the results. Therefore, the median values presented in Table 5-30 should be used as the predicted value for each point score.

Table 5-30: AADT variation for each Point Score, Coastal model

Point	Number of Observations	Min AADT	Max AADT	Average AADT	Median AADT
0	3	80	300	153	80
1,2	9	370	3800	1577	1300
3,4	11	230	2700	1560	1600
5,6	11	1100	4300	2618	2650
7	2	5500	9800	7650	7650

5.5 SUMMARY AND CONCLUSION

This chapter describes the models developed for predicting AADT for non-state upper functional classification roadway segments. The model is applicable to all roadway segments with AADT less than or equal to 10000 and excluding expressways, freeways, and ramps. The 2015 non-state database was used for model development. Since the dataset had a high degree of skew, we recommend using the median while selecting the representative AADT.

Based on the literature review and preliminary analysis, the functional classification was found to be an important variable. The Region 2 dataset was also classified into sub-regions (Valley-Rural, MPO, Mountain, and Coast). Roadway segments in Mountain sub-region were found to have lower AADT compared to roadway segments in MPO sub-region. Based on the sub-region and the functional classification, we arrived at recommended default median AADT values. These values can be used as a best guess if no other information is available.

A data collection form was developed focusing on relevant roadway, geometric, and land use related characteristics which were found to be effective for predicting AADT from the literature review. We adopted a stratified random sampling technique and selected 290 roadway segments while ensuring all functional classifications and sub-regions were well represented. After the data was collected, descriptive statistical analysis was conducted to identify variables which explained the variation in AADT. Roadway segments with generators, parking lot, two-way left turn median, left turn lane, right turn lane, traffic signal, crosswalk, sidewalk, bus stop, bike lane, pavement marking, and shoulder are found to have higher AADTs. Roads with no stop signs on them, stop signs on their cross roads, and with signs (other than stop signs) were found to have higher AADT compared to those without these features.

A point based model was used to predict AADT. An overall model was developed for Region 2. Since the AADT is found to be different based on the sub-region in which the road segment is located, separate models were developed for each sub-region. Table 5-31 summarizes the variables used in each of the point models. The point based model has two steps. The first step involves identifying an AADT value based on the point score received. The second stage involves adjusting the AADT using adjustment factors. The coast sub-region model does not use adjustment factors as it did not result in an improvement in prediction accuracy.

Table 5-31: Variables used in the Point System Models for Non-state Roads

Variables used in Models	Model				
	Overall	Valley-Rural	Valley-MPO	Mountain	Coast
Arterial	✓	✓	✓	✓	
Urban		✓		✓	
Within City Limits		✓			
Generator	✓	✓	✓	✓	✓
Two-Way Left Turn Median	✓	✓	✓		
Right Turn Lane	✓	✓	✓		✓
Left Turn Lane	✓	✓	✓		✓
Traffic Signal	✓	✓	✓		
Pavement Marking	✓				
Crosswalk	✓	✓	✓	✓	✓
Sidewalk	✓	✓		✓	
Bike Lane	✓	✓	✓		
Bus Stop	✓	✓	✓		✓
Parking Lot	✓	✓	✓	✓	✓
Cross Road Stop Sign	✓			✓	✓
No Stop Sign on Study Road Segment			✓		
National Highway System	✓		✓		

6.0 POINT BASED MODEL FOR LOCAL ROADS

This chapter describes the point based model developed for predicting ADT on local streets. A description of the local roads database used in the study is provided followed by the data collection procedure. The local roads database is randomly sampled, and relevant roadway, geometric, and the land use related data are collected. The point based model is developed to predict the AADT based on the data collected.

6.1 LOCAL ROADS DATABASE DESCRIPTION

The ODOT local roads database contains attributes such as street and location description, street type, last count date, county, region, and 2014, 2015, and 2016 ADT. There are 1756 local counts available statewide in this database of which 576 counts are located in Region 2. The ADT information for 2016 is available for 568 out of 576 Region 2 counts. For this study, we are using the counts which have 2016 ADT. There are two sets of latitude and longitude provided in this dataset - TCM and GIS latitude and longitude. The TCM latitude is more accurate and available for all the counts. Therefore, the TCM coordinates were used to locate the counts.

Figure 6-1 shows the location of the available counts categorized by sub-region in Region 2. There are 48, 115, 162, and 243 counts located in the Coast, Mountain, MPO, and Valley-rural sub-regions respectively. The MPO category is the result of a spatial join of counts with the MPO layer in the ODOT GIS database (see Figure 6-1).

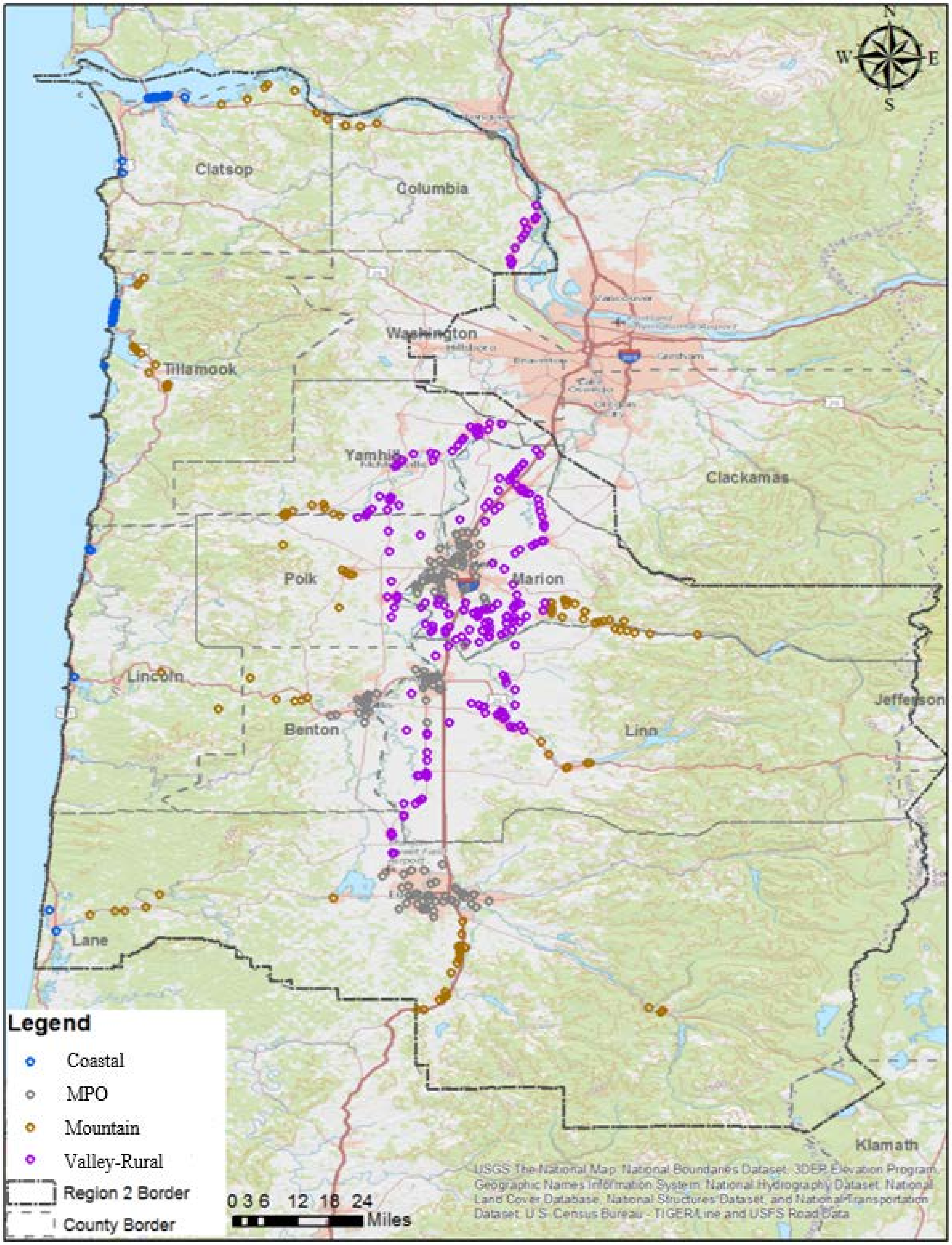

Figure 6-1: Local counts located in Region 2, ODOT

Table 6-1 shows the distribution of ADT by different sub-region and street type. As with the non-state upper functional classification analysis, the median is preferred over the average due to the presence of outliers and a high degree of skew. In the rural streets, the median ADT increases from 156 in the Coast to 267 in the MPO area. For urban streets, the Mountain sub-region has the lowest median ADT of 221. The median AADT for urban streets increases to 424 in Valley-rural sub-region. As expected, the urban local streets have higher ADT compared to rural local streets.

Table 6-1: Distribution of ADT for each Sub-region and Street Type in Region 2

Street Type	Sub-region	Number of Observations	Min ADT	Max ADT	Average ADT	Median ADT
	Coast	17	43	946	226	156
Rural	Mountain	74	2	1468	258	203
	Valley-Rural	167	4	3260	317	232
	MPO	15	28	1832	444	267
	Mountain	41	11	2515	450	221
Urban	Coast	31	48	1020	361	254
	MPO	147	11	3625	569	404
	Valley-Rural	76	31	3556	581	424

If no other information other than street type and sub-region information is available for a roadway segment, then the median ADT from Table 6-1 can be used as the best guess. Similar to the non-state upper functional classification, we developed a data collection form, sampled roadway segments from the original database, and studied the impact of roadway, geometric, and land use characteristics on ADT of local streets.

6.2 DATA COLLECTION FORM

The data collection for local roads is done through Google Maps by using the provided latitude and longitude in the database. Appendix B provides the form used for data collection on local counts. This section presents the detailed description of the variables collected:

- Land Use
    - Generator: The primary AADT generator adjacent to the road with the main access on the study road segment, looking at 1000 feet upstream and downstream of the midpoint (gas station, hospital/medical center, school, recreational facility, shopping center, or none)
    - Dominant Land use: Primary dominant land use adjacent to the study road segment, 1000 feet upstream and downstream of the midpoint (residential, commercial, industrial, forest, farm, or other)
    - Single House: Presence of single house residences adjacent to the study road segment
    - Google Street View: Availability of Google Street View adjacent to the study road segment, 1000 feet upstream and downstream of the midpoint

    - Sub-region: Whether the roadway segment lies in the Coast, Mountain, Valley-rural, or MPO
- MIRE
    - Median: Type of median (undivided, one-way, single lane, or other)
- Intersection
    - Right-Turn: Presence of right turn lane on the study road segment, 1000 feet upstream and downstream of the midpoint
    - Left-Turn: Presence of left turn lane on the study road segment, 1000 feet upstream and downstream of the midpoint
    - Traffic Signal: Presence of traffic signal on the study road segment, 1000 feet upstream and downstream of the midpoint (on any side of intersections along the main corridor of study)
- Roadway
    - Pavement Type: Variable indicating whether the road segment is paved, unpaved or gravel
    - Pavement Marking: Presence of pavement horizontal marking along the study road segment (lane marking, shoulder marking, both or none. Also note that bicycle lane markings are not counted as shoulder marking)
    - Shoulder: Presence of shoulder on the study road segment (paved, unpaved, or none)
    - Crosswalk: Presence of crosswalk on the study road segment, 1000 feet upstream and downstream of the midpoint (either crossing the main corridor of study or on any side of intersections along the main corridor of study)
    - Sidewalk: Presence of sidewalk along the study road segment, 1000 feet upstream and downstream of the midpoint (Both sides, One side, or None)
    - Bike Lane: Presence of bike lane along the study road segment, 1000 feet upstream and downstream of the midpoint
    - Bus Lane: Presence of bus stop along the study road segment, 1000 feet upstream and downstream of the midpoint
    - Parking Lot: Presence of parking lot adjacent the study road segment, 1000 feet upstream and downstream of the midpoint (Including pay to park, parking lots, and parking lots for schools, shopping centers, recreational facilities, hospitals, etc.)
    - Calming Device: Presence of traffic calming devices along the study road segment, 1000 feet upstream and downstream of the midpoint (According to ITE, traffic calming devices include speed humps, neighborhood traffic circles, speed tables, chicanes, raised intersection, choker, closure, and center island narrowing)
- Signage
    - Cross Road Stop Sign: Presence of stop sign on the cross roads of the study road segment, 1000 feet upstream and downstream of the midpoint
    - Stop Sign: Presence of stop sign on the main corridor of study, 1000 feet upstream and downstream of the midpoint

- o Sign: Presence of signs, other than stop signs on the main corridor of study, 1000 feet upstream and downstream of the midpoint

6.3 DESCRIPTIVE ANALYSIS OF DATA

This section summarizes and describes the sampling procedure and analyzes the impact of the variables described in the previous section on median ADT. We present only those variables which have a significant impact on median AADT. For this study, 200 roadway segments are selected randomly for developing models while ensuring that all sub-regions and street types are covered. As shown in Table 6-2, 96 out of the 200 counts belong to rural street type, and 104 are urban local streets.

Table 6-2: Number of Observations selected for Model Development

Sub-Region	Rural			Urban		
	Street View	No Street View	Total	Street View	No Street View	Total
Coast	2	5	7	5	6	11
Mountain	12	14	26	12	2	14
Valley-Rural	44	14	58	24	3	27
MPO	5	1	5	48	4	52

6.3.1 Google Street View

Roadway segments with Google Street View have higher median ADT (nearly 300% higher) in comparison with the roadway segments without Google Street View (see Table 6-3). This is expected as roads without Google Street View are often extremely rural and less travelled roads.

Table 6-3: Descriptive Statistics of ADT by the Availability of Google Street View

Google Street View	Number of Observations	Min ADT	Average ADT	Max ADT	Median ADT
No	48	15	245	1035	178
Yes	152	7	516	2515	407

It was not possible to accurately collect data on the roadway and geometric features for the roadway segments without Google Street View. Therefore, the rest of the descriptive analysis focuses on the 152 locations with Google Street View.

6.3.2 Land Use Type

Table 6-4 shows the variation of ADT of roadway segments with different types of generator present adjacent to the study road. As expected, roadway segments with generators such as shopping centers have higher median ADT compared to those roadway segments with no generators.

Table 6-4: Descriptive Statistics of ADT by the Generator adjacent to the Road

Generator	Number of Observations	Min ADT	Average ADT	Max ADT	Median ADT
No Generator	101	7	471	2501	344
Recreational Facility	14	135	402	883	385
Gas Station	5	221	483	914	428
School	7	141	638	1968	424
Shopping Center	25	112	733	2515	541

Table 6-5 shows the variation of ADT for the dominant land use adjacent to the road. The median ADT is higher on local streets with residential, industrial, and commercial dominant land use when compared to local streets with forest and farm land use. The majority of the counts that are located in forest and farm don't have the parking lots adjacent to them and are expected to have lower ADT. Table 6-6 shows that roadway segments with parking lots adjacent to them have higher median ADT when compared to roadway segments with no parking lots. Median ADT was found to be higher in roadway segments which have adjacent single houses (see Table 6.7).

Table 6-5: Descriptive Statistics of ADT by Dominant Land Use

Primary Land Use	Number of Observations	Min ADT	Average ADT	Max ADT	Median ADT
Forest	2	86	184	281	184
Farm	63	23	440	2501	313
Industrial	4	112	311	421	356
Residential	62	7	539	1968	403
Commercial	21	135	748	2515	555

Table 6-6: Descriptive Statistics of ADT by the Presence of a Parking Lot

Presence of Parking Lot	Number of Observations	Min ADT	Average ADT	Max ADT	Median ADT
No	67	23	379	2501	281
Yes	85	7	624	2515	459

Table 6-7: Descriptive Statistics of ADT by the Presence of Adjacent Single Houses

Presence of Single House	Number of Observations	Min ADT	Average ADT	Max ADT	Median ADT
No	57	23	441	1553	368
Yes	95	7	561	2515	425

6.3.3 Median Type

The single line median was the only type of median present in the local count dataset. There is only one location which was one way, and the rest of the counts are either undivided or having a single line median. Undivided roads have lower median ADT when compared to roads with single line divided median (see Table 6-8).

Table 6-8: Descriptive Statistics of ADT by Median Type

Median Type	Number of Observations	Min ADT	Average ADT	Max ADT	Median ADT
One-Way	1	144	144	144	144
Undivided	69	7	494	1968	394
Single-Line	82	39	547	2515	413

6.3.4 Roadway Characteristics

Roadway segments with a right turn lane, left turn lane, and traffic signals have higher median ADT when compared with the roadway segments without these characteristics. Table 6-9, Table 6-10, and Table 6-11 shows that the median ADT of the roadway segments with these characteristics are about three times higher than the ones without it.

Table 6-9: Descriptive Statistics of ADT by the Presence of Right Turn Lane

Presence of Right Turn	Number of Observations	Min ADT	Average ADT	Max ADT	Median ADT
No	148	7	491	2515	393
Yes	4	1199	1438	1553	1499

Table 6-10: Descriptive Statistics of ADT by the Presence of Left Turn Lane

Presence of Left Turn	Number of Observations	Min ADT	Average ADT	Max ADT	Median ADT
No	148	7	493	2515	393
Yes	4	1199	1363	1553	1351

Table 6-11: Descriptive Statistics of ADT by the Presence of a Traffic Signal

Presence of Traffic Signal	Number of Observations	Min ADT	Average ADT	Max ADT	Median ADT
No	150	7	508.1	2515	402.5
Yes	2	741	1104.5	1468	1104.5

Almost all of the roadway segments are paved. Therefore, the information about the pavement was not useful in predicting ADT. In general, crosswalks, sidewalks, bike lanes, and bus stops are found in urban areas and expected to have higher ADT. As Table 6-12, Table 6-13, Table

6-14, and Table 6-15 shows, the median ADT is higher in the locations that these characteristic are present.

Table 6-12: Descriptive Statistics of ADT by the Presence of a Crosswalk

Presence of Crosswalk	Number of Observations	Min ADT	Average ADT	Max ADT	Median ADT
No	128	7	478	2515	377
Yes	24	134	720	1968	506

Table 6-13: Descriptive Statistics of ADT by the Presence of a Sidewalk

Presence of Sidewalk	Number of Observations	Min ADT	Average ADT	Max ADT	Median ADT
None	19	23	448	2501	344
One side	88	7	325	714	320
Both sides	45	106	730	2515	482

Table 6-14: Descriptive Statistics of ADT by the Presence of a Bike Lane

Presence of Bike Lane	Number of Observations	Min ADT	Average ADT	Max ADT	Median ADT
No	147	7	483	2501	386
Yes	5	531	1473	2515	1468

Table 6-15: Descriptive Statistics of ADT by the Presence of a Bus Stop

Presence of Bus Stop	Number of Observations	Min ADT	Average ADT	Max ADT	Median ADT
No	148	7	503	2515	393
Yes	4	471	993	1594	953

6.3.5 Signage Characteristics

Local roads that have stop signs on the cross roads have a higher median ADT when compared to roads which do not have stop signs on the cross roads (see Table 6-16). Most of the local streets have stop signs. Local roadway segments with stop signs have marginally higher ADT than local roadway segments without stop signs (see Table 6-17). Also, roadways with signs other than stop signs were found to have higher ADT (see Table 6-18) compared to roadway segments with no signs.

Table 6-16: Descriptive Statistics of ADT by the Presence of Stop Sign on Crossroads

Presence of Stop Signs on Crossroads	Number of Observations	Min ADT	Average ADT	Max ADT	Median ADT
No	93	7	436	2501	363
Yes	59	39	642	2515	475

Table 6-17: Descriptive Statistics of ADT by the Presence of Stop Sign on the Roadway Segment

Presence of Stop Signs	Number of Observations	Min ADT	Average ADT	Max ADT	Median ADT
No	21	39	450	1468	408
Yes	131	7	527	2515	405

Table 6-18: Descriptive Statistics of ADT by the Presence of Signs other than Stop Signs on the Roadway Segment

Presence of Signs other than Stop Signs	Number of Observations	Min ADT	Average ADT	Max ADT	Median ADT
No	29	7	485	1594	383
Yes	123	23	523	2515	413

6.4 MODEL DEVELOPMENT

Similar to the non-state upper functional classification roadway systems, we developed a point based system to predict ADT. The descriptive analysis provided us with insights on the impact of roadway, geometric, and land use related factors on the median ADT of local streets. We carefully selected a subset of these factors and assigned them 1 point each. The total points for a local street roadway segment are calculated by adding up all the points. Our model is based on the logic that higher the points, the greater the median ADT.

In this project, we developed a point based model for overall Region 2 and then specific models for each of the sub-regions. We expect that the models for the sub-regions will better predict the ADT. Our point based model is applicable only for those areas with Google Street View. For the Coastal sub-region, only eight counts were present in the sampled data with Google Street View. Due to the low sample size, we recommend using overall model for the Coastal sub-region.

Note that the point based models are only applicable to those roadway segments with Google Street View. Since we relied on Google Street View to collect the data, we could not accurately determine the variables for those locations without Google Street View. Roadway segments with Google Street View have higher median ADT in comparison with the roadway segments without Google Street View (see Table 6-3). This trend was found to be consistent for each of the sub-regions (see Table 6-20). For local roads with no Google Street View, if the sub-region information is available, we recommend using the median ADT shown in Table 6-19 as predicted ADT. For local roads with no Google Street View, if the sub-region and functional classification information is available, we recommend using the median ADT shown in Table 6-21 and Table 6-22 as predicted ADT. Similar recommendations are made for local roads with Google Street View. However, if Google Street View is available, then we recommend collecting the relevant information outlined in the next section and apply the point based model.

Table 6-19: ADT by Sub-region for Locations without Google Street View

Sub-Region	Number of Observation	Min ADT	Average ADT	Max ADT	Median ADT
Valley-Rural	50	4	185	1088	70
Mountain	46	2	179	1106	100
MPO	13	28	175	1010	100
Coast	27	48	288	1020	158

Table 6-20: Descriptive Statistics of ADT by Google Street View and Sub-region

Sub-Region	Street Type	Number of Observations	Min ADT	Average ADT	Max ADT	Median ADT
Mountain	Google Street View	69	12	425	2515	271
	No Google Street View	46	2	179	1106	100
Coast	Google Street View	21	43	346	914	254
	No Google Street View	27	48	288	1020	158
Valley-Rural	Google Street View	193	7	456	3556	331
	No Google Street View	50	4	185	1088	70
MPO	Google Street View	149	11	591	3625	408
	No Google Street View	13	28	175	1010	100

Table 6-21: Descriptive Statistics of ADT by Google Street View and Sub-region in Rural streets

Google Street View	Street View				No Street View			
Sub-Region	Coast	MPO	Mountain	Valley-Rural	Coast	MPO	Mountain	Valley-Rural
Number of Observations	7	13	35	126	10	2	39	41
Min ADT	43	99	49	7	57	28	2	4
Average ADT	143	506	349	375	284	44	177	141
Max ADT	246	1832	1468	3260	946	60	1106	544
Median ADT	135	281	271	268	157	44	113	67

Table 6-22: Descriptive Statistics of ADT by Google Street View and Sub-region in Urban streets

Google Street View	Street View				No Street View			
Sub-Region	Coast	MPO	Mountain	Valley-Rural	Coast	MPO	Mountain	Valley-Rural
Number of Observations	14	136	34	67	17	11	7	9
Min ADT	67	11	12	52	48	37	11	31
Average ADT	448	599	503	608	290	199	193	385
Max ADT	914	3625	2515	3556	1020	1010	726	1088
Median ADT	382	440	269	428	191	119	67	119

6.4.1 Overall Model

The overall Region 2 model covers all the sub-regions. In the overall model, we assign one point each if the roadway segment has any of the following features:

- Street type is urban
- Presence of right turn lane
- Presence of left turn lane
- Presence of parking lot
- Presence of sidewalk on both sides of the street

The maximum possible points in this model is 5. The total points of the roadway segment is determined by adding up the points. Table 6-23 presents the median values to be used as the predicted ADT for each point level.

Table 6-23: Variation of ADT with Point Score for Overall Model

Point	Number of Observations	Min ADT	Average ADT	Max ADT	Median ADT
0	49	23	302	1372	228
1	27	7	416	883	368
2	35	134	611	2501	459
3	37	106	665	2515	480
4 or more	5	1199	1397	1553	1468

We conjectured that the characteristics of the roads are different in each sub-region. Therefore, we developed separate models for each sub-region. The following sections are the results of the point based system for the available four sub-regions.

6.4.2 Valley-rural Model

There are 68 location sites within the Valley-rural sub-region. The variables used for the points are:

- Street type is urban
- Presence of recreational facility adjacent to the study road segment
- Presence of school adjacent to the study road segment
- Presence of shopping center adjacent to the study road segment
- Presence of gas station adjacent to the study road segment
- Presence of left turn lane
- Presence of both lane and shoulder marking
- Presence of bus stop
- Presence of stop sign on cross roads

After finding the total points of the road segment, Table 6-24 presents the median values to be used as the predicted ADT for each point score.

Table 6-24: Variation of ADT with Point Score for Valley-rural Sub-region

Point	Number of Observations	Min ADT	Average ADT	Max ADT	Median ADT
0	21	7	165	413	151
1	20	39	293	747	210
2	20	211	566	1372	492
3 or more	7	357	891	1594	544

6.4.3 MPO Model

There are 52 counts located in the MPO sub-region in Region 2. In the MPO model, we assign one point for each of the following features present in the roadway segment:

- Presence of shopping center adjacent to the study road segment
- Median type is single-line
- Presence of crosswalk
- Presence of parking lot
- Presence of stop sign on cross street

After finding the total points of the road segment, Table 6-25 presents the median values to be used as the predicted ADT for each point score.

Table 6-25: Variation of ADT with Point Score for MPO Sub-region

Point	Number of Observations	Min ADT	Average ADT	Max ADT	Median ADT
1 or less	21	125	469	969	427
2	20	112	595	1678	518
3 or more	11	134	1074	2501	1111

6.4.4 Mountain Model

There are 52 counts located in the Mountain sub-region in Region 2. In the Mountain model, we assign one point for each of the following feature present in the roadway segment:

- Presence of shopping center adjacent to the study road segment
- Dominant land use in commercial adjacent to the study road segment
- Presence of single house residences

- Median type is single-line
- Presence of bike lane
- Presence of sidewalks on both sides of the street

The median ADT to be used as the predicted ADT for the point scores is shown in Table 6-26.

Table 6-26: Variation of ADT with Point Score for Mountain Sub-region

Point	Number of Observations	Min ADT	Average ADT	Max ADT	Median ADT
1 or less	10	61	293	661	283
2	10	86	485	1228	373
3 or more	4	706	1664	2515	1718

6.5 SUMMARY

This chapter describes the models developed for predicting ADT for local roadway segments. The ADT values in the dataset had a high degree of skew. Therefore, we recommend using the median while selecting representative ADT.

A data collection form was developed focusing on relevant roadway, geometric, and land use related characteristics which were found to be effective for predicting ADT from the literature review. We adopted a stratified random sampling technique and selected 200 roadway segments while ensuring all sub-regions and urban and rural street types were well represented. The 200 data points were classified into those with Google Street View and those without Google Street View. For roadway segments without Google Street View, default ADT values were determined based on sub-regions and street type. For the data points with Google Street View, data was collected, and descriptive statistical analysis was conducted to identify variables which explained the variation in ADT.

A point based model was used to predict ADT. An overall model was developed for Region 2. Since the ADT is found to be different based on the sub-region in which the road segment is located, separate models were developed customized for Valley-rural, MPO, and Mountain sub-regions. The number of sample points for Coastal sub-region was found to be too low to develop a point based model. Therefore, we recommend using the overall model in the Coastal sub-region. Table 6-27 summarizes the variables used in each of the point models. Unlike the point based models for non-state upper functional classification roadway segments, adjustment factors were not used in the point based model for local streets.

Table 6-27: Variables used in the Point Models for Local Roadway Segments

Variables Used in Models	Model			
	Overall	Valley-Rural	MPO	Mountain
Urban Street Type	✓	✓		
Recreational Facility		✓		
School		✓		
Shopping Center		✓	✓	✓
Gas Station		✓		
Commercial Land-use				✓
Single House				✓
Single Line Median			✓	✓
Right Turn	✓			
Left Turn	✓	✓		
Both Lane/Shoulder Marking		✓		
Bike Lane				✓
Bus Stop		✓		
Both Sides Sidewalk	✓			✓
Cross Walk			✓	
Parking Lot	✓		✓	
Stop Sign on Cross Street		✓	✓	

7.0 MODEL VALIDATION

This chapter summarizes the validation procedure and results for the point based model developed for non-state upper functional roadways with AADT less than 10000 and local roads. The with-held data technique is used for the validation. We generated the validation data set using the same procedure used for the generation of dataset used for model development.

7.1 NON-STATE UPPER FUNCTIONAL CLASSIFICATION ROADWAY SEGMENTS

The validation data points were selected randomly with respect to functional classification and sub-regions. In total, 90 points were surveyed for validation, of which 26, 32, 16, and 26 points correspond to Valley-rural, MPO, Mountain, and Coastal sub-regions respectively. The data collection form described in Section 5.2 was used for collecting data on the relevant roadway, geometric, and land use characteristics. Table 7-1 and Table 7-2 provides the descriptive statistics on the variation in AADT with respect to functional classification and sub-regions. Similar to the calibration dataset, the median AADT of urban roadway segments were found to be higher than the median AADT of rural roadway segments. Median AADT of arterials were found to be higher than the median AADT of collectors. Roadway segments in the MPO sub-region were found to have the highest median AADT.

Table 7-1: Descriptive Statistics of AADT in each Functional Classification in Region 2, ODOT (Validation Data)

Functional Classification	Number of Observations	Min AADT	Max AADT	Average AADT	Median AADT
Rural Minor Collector	10	60	681	1700	530
Rural Major Collector	10	60	1554	4500	1500
Rural Minor Arterial	3	1800	2833	4800	1900
Urban Minor Collector	14	420	1480	2900	1250
Urban Collector	36	160	2312	8200	1650
Urban Minor Arterial	14	1800	5857	9500	6150
Other Urban Principal Arterial	3	7900	8867	9800	8900
All Observations	90	60	2704	9800	1800

Table 7-2: Descriptive Statistics of AADT in the four Sub-regions in ODOT Region 2 (Validation Data)

Region	Number of Observations	Min AADT	Max AADT	Average AADT	Median AADT
Mountain	16	60	1533	7500	1200
Coast	16	630	1991	5500	1700
Valley-Rural	26	60	2454	6600	1850
MPO	32	160	3850	9800	2350

The next sections describe the validation results for the overall and sub-region models.

7.1.1 Validation Results for Overall Model

The overall point model developed in Section 5.4.1 was applied to the validation data set. First, we calculated the point score for each roadway segment in the validation data set based on the variables identified in Section 5.4.1. Depending on the point score, the AADT for each roadway segment was identified from the median AADT column of Table 5-25. This AADT was then modified using adjustment factors obtained from Table 5-26 to get the final AADT predictions. Figure 7-1 shows the validation results of the overall model. The y axis is the difference between the predicted AADT and actual AADT. The x axis is the actual AADT. The performance of the point model is found to be reasonable up to an AADT of 5000. The point based model is found to be not suitable for predicting AADT of roadway segments which have AADT closer to 10000.

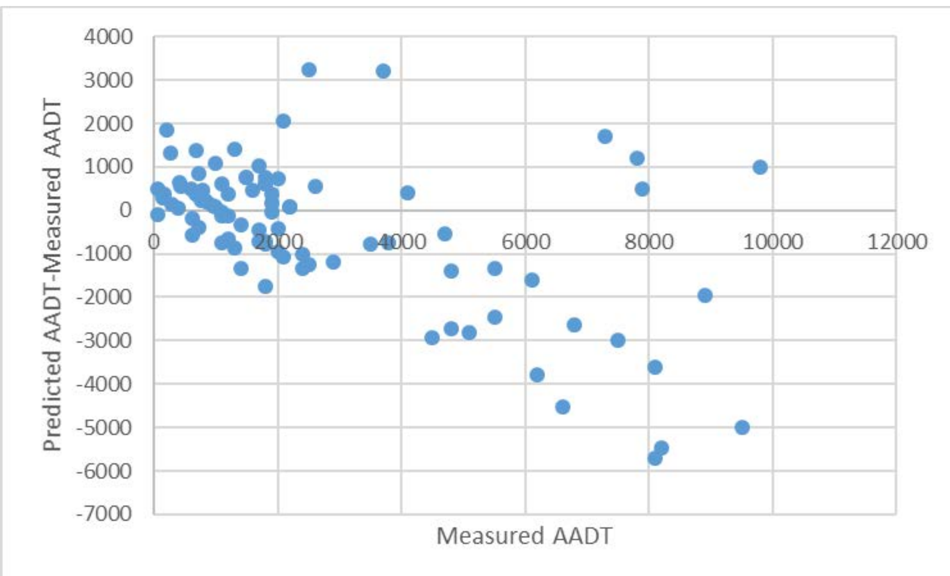

Figure 7-1: Validation results of the overall model

The error is defined as the difference between predicted and actual AADT for each data point in the validation data set. In this analysis, we studied two different metrics to characterize the accuracy of the point model: (i) median of the errors across all points in the validation data set, and (ii) average of the errors across all points in the validation data set. Table 7-3 and Table 7-4 provide an overview of the distribution of the accuracy metrics across functional classifications and sub-regions. The errors are found to be higher for urban regions compared to rural regions. On the whole, the model under-predicts the AADT as indicated by the negative average error. However, the median error is close to zero. Coastal sub-regions had lower average errors. The MPO and Valley-rural sub-region has higher average errors. This is potentially due to the presence of higher number of roadways with AADT close to 10000 in the MPO sub-region.

Table 7-3: Prediction Errors of the Overall Model with respect to Functional Classification

Functional Classification	Median of Errors	Average of Errors
Rural Minor Collector	-30	-76
Rural Major Collector	150	-150
Rural Minor Arterial	-35	-278
Urban Minor Collector	-415	-387
Urban Collector	75	-318
Urban Minor Arterial	-1175	-1111
Other Urban Principal Arterial	500	-150
Overall	13	-400

Table 7-4: Prediction Errors of the Overall Model with respect to Sub-regions

Sub-region	Median of Errors	Average of Errors
Mountain	-117	-326
Coast	125	67
Valley-Rural	-51	-566
MPO	95	-534
Overall	13	-400

7.1.2 Validation Results for Sub-regional Models

This section presents the analysis of the performance of the sub-regional point models described in Section 5.4.2, 5.4.3, 5.4.4, and 5.4.5 in predicting AADT for the validation data set.

Figure 7-2, Figure 7-3, Figure 7-4, and Figure 7-5 shows the validation results for the sub-region model for the Valley-rural, MPO, Mountain, and Coastal sub-region respectively. The difference between predicted and actual AADT increases with AADT.

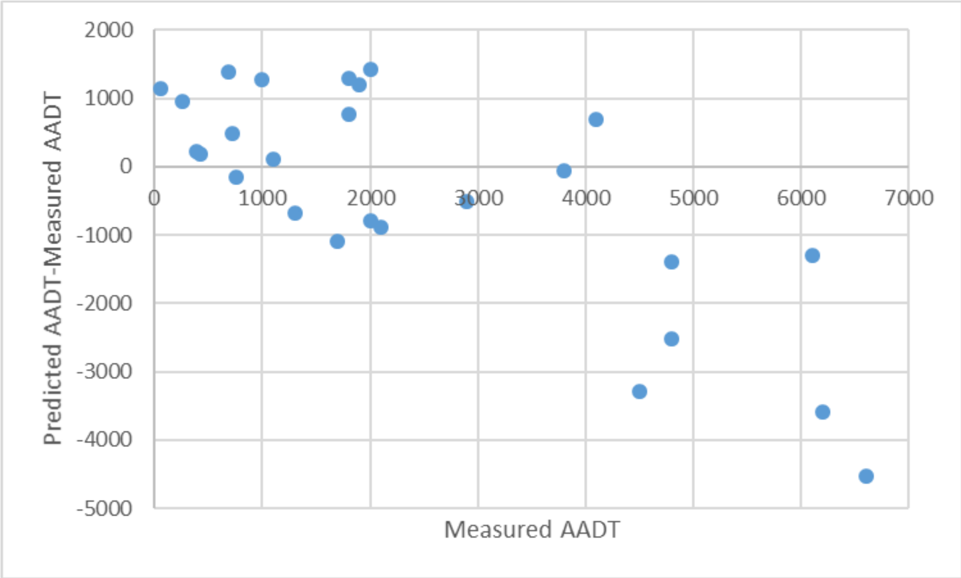

Figure 7-2: Validation results of the valley-rural model

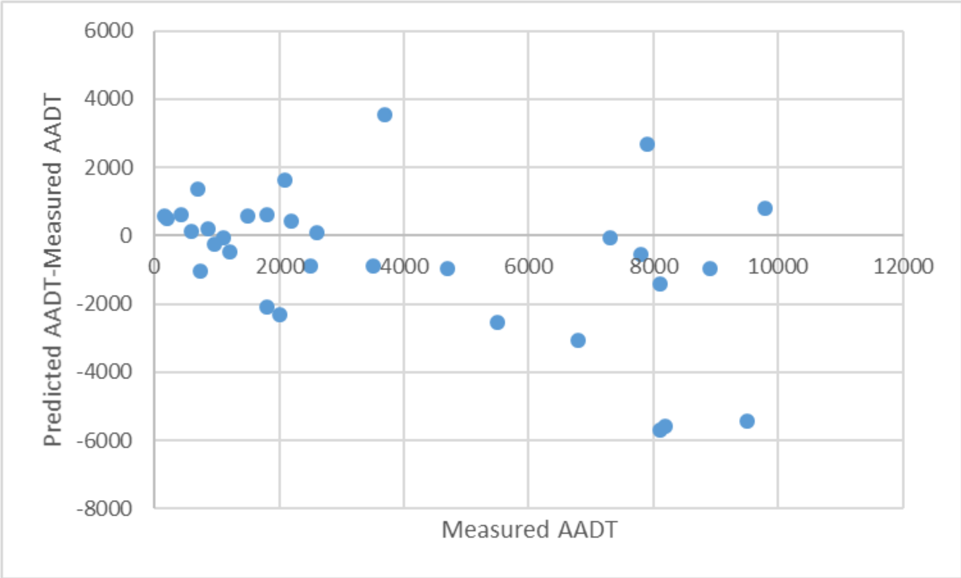

Figure 7-3: Validation results of the MPO model

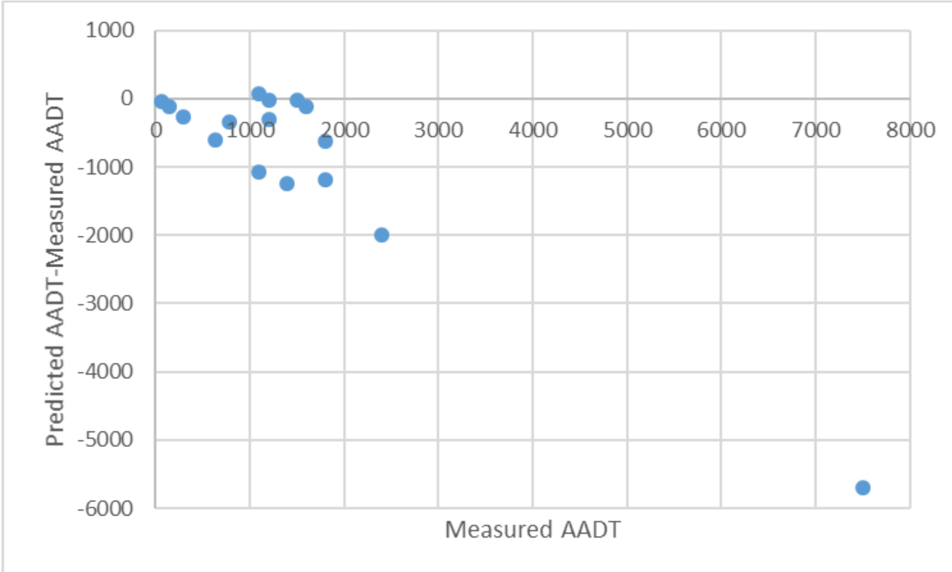

Figure 7-4: Validation results of the mountain model

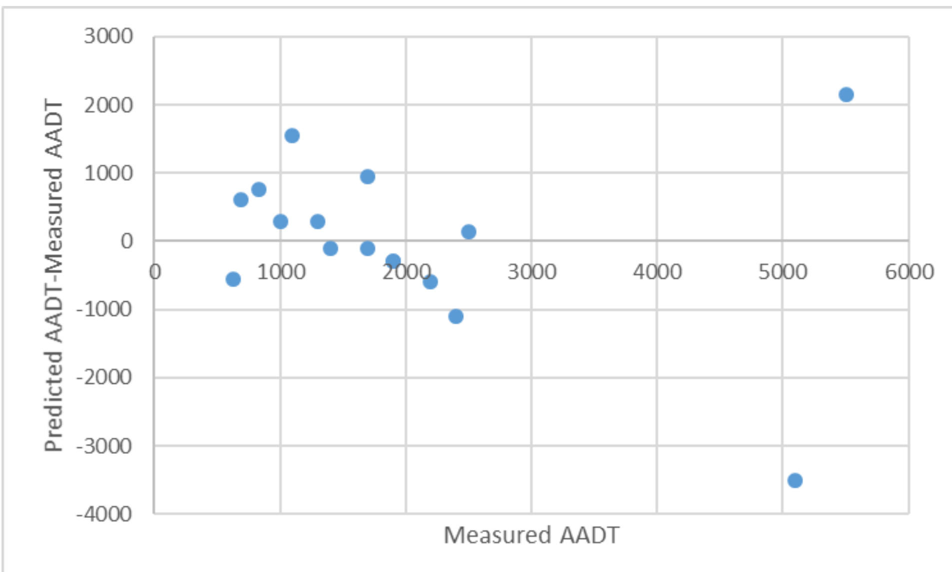

Figure 7-5: Validation results of the coastal model

Table 7-5 shows the distribution of accuracy metrics (median and average of errors) for the sub-region models applied to the four sub-regions. Significant improvement in median errors were obtained for the Coast and Valley-rural sub-region from the sub-regional model when compared to the overall model. The median error in the Coastal region from the sub-regional model was only 20% of the median error for the overall model. The median error in the Valley-rural region from the sub-regional model was only 50% of the median error for the overall model. For the other two sub-regions, there was a decrease in model performance. Therefore, the improvements in accuracy from the sub-regional model depends on the sub-region.

Table 7-5: Prediction Errors of the Sub-regional Models

Sub-region	Median of Errors	Average of Errors
Mountain	-321	-850
Coast	25	14
Valley-Rural	29	-371
MPO	-143	-639

7.2 LOCAL ROADS

This section describes the validation procedure and the results for the local roads functional classification. The data points for validation were selected randomly. We ensured that the data points covered both urban and rural local streets and the four sub-regions. The data collection form in section 6.2 was used to record relevant geometric, land use, and roadway related information. Since the point model can be applied to only those local streets with Google Street View, we focused the data collection on street segments with Google Street View. Table 7-6 provides the descriptive statistics of ADT for the local street validation data set. The minimum and maximum ADT for the validation dataset are 43 and 3625 respectively. The average and median ADT is 650 and 380 respectively. Table 7-7 provides the ADT categorized by sub-regions. As expected, the median ADT is highest for MPO. The ADT distribution is consistent with the original data set.

Table 7-6: Descriptive Statistics of ADT in each Functional Classification in Region 2, ODOT (Validation Data)

Functional Classification	Number of Observations	Min ADT	Max ADT	Average ADT	Median ADT
Rural-Local	17	43	3260	546	246
Urban-Local	22	110	3625	726	417
All Observations	39	43	3625	648	366

Table 7-7: Descriptive Statistics of ADT in the Four Sub-regions in ODOT Region 2 (Validation Data)

Region	Number of Observations	Min ADT	Max ADT	Average ADT	Median ADT
Mountain	7	176	1283	476	239
Coast	4	43	576	277	245
Valley-Rural	14	104	3260	699	380
MPO	14	162	3625	788	406

7.2.1 Validation Results for Overall Model

The overall point model developed in Section 6.4.1 was applied to the validation data set. First, we calculated the point score for each roadway segment in the validation data set based on the variables identified in Section 6.4.1. Depending on the point score, the ADT for each roadway segment was identified from the median ADT column of Table 6-23. Figure 7-6 plots the difference between predicted and actual ADT on the y axis and the actual ADT on the x axis. The errors show a strong linear relationship with AADT. There is over prediction up to an AADT of 500. For roadway segments with AADT of higher than 500, the model generally under predicts. In general, the overall point based model performance appears to be reasonable until around 1000. The prediction accuracy of the point based model appear to deteriorate significantly for streets with ADT higher than 1000.

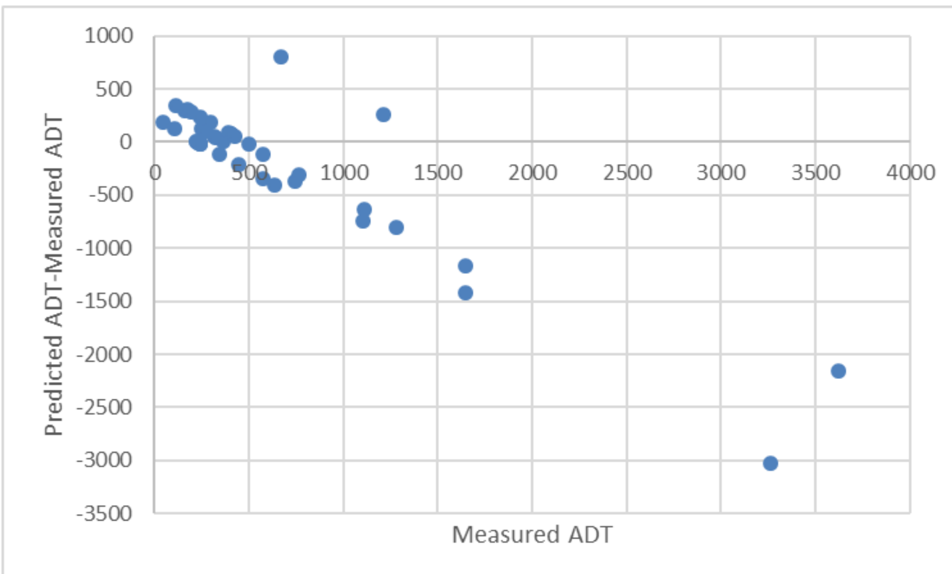

Figure 7-6: Validation results of the overall model

Table 7-8 and Table 7-9 provide the distribution of model accuracies (based on median and average of errors) for urban and rural local streets as well as the four sub-regions. The overall model under-predicts for rural local streets and over predicts for urban local streets based on median errors. Based on the absolute value of errors, the accuracy is higher for rural compared to urban local streets. The point based model performance look reasonable for all sub-regions with the lowest median error for the Valley-rural region.

Table 7-8: Prediction Error of the Overall Model by Functional Classification – Local Roads

Functional Classification	Median of Errors	Average of Errors
Rural-Local	-77	-317
Urban-Local	52	-150
Overall	-32	-223

Table 7-9: Prediction Error of the Overall Model by Sub-regions – Local Roads

Sub-region	Median of Errors	Average of Errors
Mountain	-32	-30
Valley-Rural	4	-363
MPO	-50	-221
Overall	-32	-223

7.2.2 Validation Results for Sub-regional model

This section describes the analysis of the performance of the sub-regional point models described in Section 6.4.2 and 6.4.4 in predicting ADT for the validation data set. Note that we did not develop a sub-region model for the Coastal sub-region due to the low number of data points. Figure 7-7, Figure 7-8, and Figure 7-9 illustrate the validation results of sub-regional models developed for local roads.

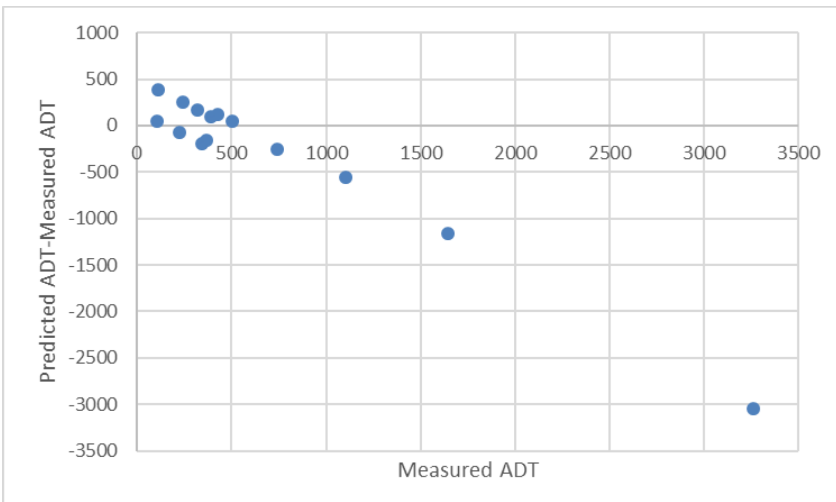

Figure 7-7: Validation results of the valley-rural model – local roads

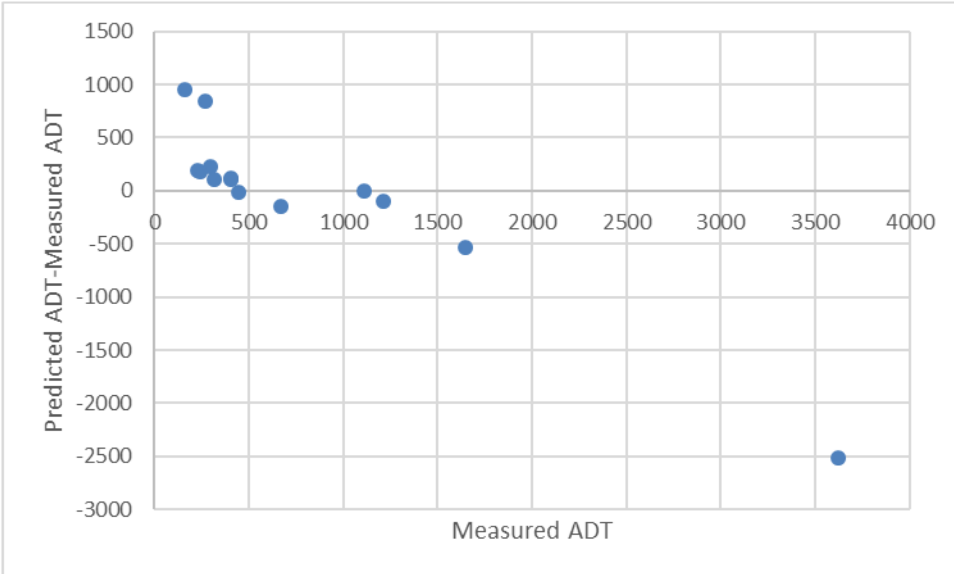

Figure 7-8: Validation results of the MPO model – local roads

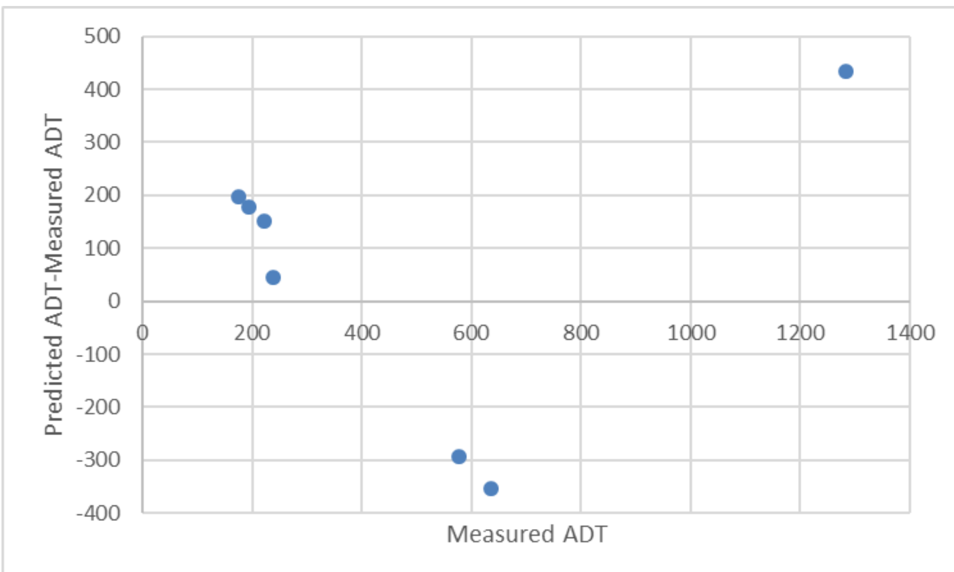

Figure 7-9: Validation results of the mountain model – local roads

Table 7-10: Prediction Error of the Sub-regional Models – Local Roads

Sub-region	Median of Errors	Average of Errors
Mountain	151	51
Valley-Rural	-16	-309
MPO	112	-42

Using the sub-regional models, the median error increased whereas there was a decrease in the average error compared to the overall model (see Table 7-9 and Table 7-10). This is potentially due to the lower samples available with Google Street View for developing the sub-regional model. Therefore, we recommend using the overall model for local streets.

7.3 LIMITATIONS

The point based model is based on the assumption that roadway segments with higher number of points will have higher median AADT. Intuitively, this assumption makes sense as roads with higher AADT often have lot more pavement markings, signage, crosswalks, sidewalks, bus stops, etc. However, we can find several examples where this assumption may not hold. For example, Figure 7-10 is an example of a rural road with only 3 points but with surprisingly high AADT. Figure 7-11 is an example of a rural road passing through farmland which has only 2 points but has a high AADT of 3600. In both these cases, we would expect the AADT to be less than 1000. The high AADT is potentially due to the fact that these roadway segments connect generators of traffic. Figure 7-12 is an example of a roadway segment in an urban area with pavement markings, sidewalks, crosswalks, bicycle lanes but with a low AADT of 880.

Therefore, it might not be possible to accurately predict the AADT just using roadway geometric, signage, and land use characteristics. However, it is important to highlight that the proposed point based methodology is appropriate for estimating the median AADT of similar roadways.

Figure 7-10: Roadway Segment with AADT = 4500, POINTS = 3

Figure 7-11: Roadway Segment with AADT = 3600, POINTS = 2

Figure 7-12: Roadway Segment with AADT = 880, POINTS = 10

7.4 SUMMARY

This chapter focuses on validating the point based models for non-state upper functional classification roadway system (developed in Chapter 5.0) and non-state local roads (developed in Chapter 6.0). The performance of the overall model for Region 2 and sub-region models in terms of prediction accuracy were compared. In this analysis, we studied two different metrics to characterize the accuracy of the point model: (i) median of the errors across all points in the validation data set and (ii) average of the errors across all points in the validation data set .

First, we looked at the overall model and sub-regional models for non-state upper functional classification roadway segments. A total of 90 data points were used for validation. These data points were randomly selected from the original non-state database while ensuring that all functional classifications and sub-regions were adequately covered. The overall Region 2 model had a median error of 13. The average error was negative which indicates that the model slightly under predicts the AADT. The sub-region models provided significantly lower median errors compared to the overall Region 2 model for the Coastal and Valley-rural sub-regions. In general, the model errors were found to be reasonable on roadway segments with an AADT of up to 5000.

For the local roads, the validation dataset had 39 randomly selected counts which covered both urban and rural functional classification as well as the four sub-regions. The overall model had a median error of -32 which indicates that the model slightly under-predicts the ADT. The overall model has the lowest median error of 4 for the Valley-rural sub-region. The errors in the other sub-regions are consistent. The sub-region model delivers marginally improved performance. However, for the local roads, we recommend using the overall model as the gains in accuracy are minimal.

It is important to highlight that most studies surveyed in the literature review did not follow a proper validation approach.

8.0 CURRENT DATA REQUIREMENTS AND FUTURE DATA SOURCES

This chapter evaluates data requirements for predicting AADT using the point based models. First, we analyze the relative importance of each variable used in the model. Later we briefly describe future (potential) mobile phone based data sources which can be used to improve AADT prediction models in the future.

8.1 DATA REQUIREMENTS FOR POINT BASED MODELS

One way to characterize the importance of variables is to see how many times they appear in different models. Table 8-1 and Table 8-2 show the number of times each variable appears in the models for non-state and local roads, respectively. For non-state roads, the following are the more important variables:

- Arterial – if the roadway segment is an arterial or not
- Generator – Presence of generator (Gas Station, Hospital/Medical Center, School, Recreational Facility, Shopping Center) adjacent to the roadway segment
- Right Turn Lane – Presence of a right turn lane on the roadway segment
- Left Turn Lane – Presence of a left turn lane on the roadway segment
- Crosswalk – Presence of a crosswalk on the roadway segment
- Bus Stop – Presence of a bus stop on the roadway segment
- Parking Lot – presence of a parking lot adjacent to the roadway segment of interest

For local roads, the following are the more important variables:

- Shopping Center – Presence of a shopping center adjacent to the roadway segment
- Urban – if the roadway segment is in urban area
- Single line median – presence of a single line median
- Left Turn Lane – Presence of a left turn lane on the roadway segment
- Both sides Sidewalk – Presence of a sidewalk on both sides of the roadway segment
- Parking Lot – Presence of a parking lot adjacent to the roadway segment of interest

- Stop Sign on Cross Street – Presence of a stop sign on cross street

Table 8-1: Number of times each Variable appear in the Non-state Models.

Variables used in models	Model					Number of times the variable appears in different models
	Overall	Valley-Rural	MPO	Mountain	Coast	
Arterial	1	1	1	1		4
Urban		1		1		2
Within City Limits		1				1
Generator	1	1	1	1	1	5
Two-Way Left Turn Median	1	1	1			3
Right Turn Lane	1	1	1		1	4
Left Turn Lane	1	1	1		1	4
Traffic Signal	1	1	1			3
Marking	1					1
Crosswalk	1	1	1	1	1	5
Sidewalk	1	1		1		3
Bike Lane	1	1	1			3
Bus Stop	1	1	1		1	4
Parking Lot	1	1	1	1	1	5
Cross Road Stop Sign	1			1	1	3
No Stop Sign on Study Road Segment			1			1
National Highway System	1		1			2

Table 8-2: Number of times each Variable appear in the Local Model

Variables Used in Models	Model				Number of times the variable appears in different models
	Overall	Valley-Rural	MPO	Mountain	
Urban Street Type	1	1			2
Recreational Facility		1			1
School		1			1
Shopping Center		1	1	1	3
Gas Station		1			1
Commercial Land-use				1	1
Single House				1	1
Single Line Median			1	1	2
Right Turn	1				1
Left Turn	1	1			2
Both Lane/Shoulder Marking		1			1
Bike Lane				1	1
Bus Stop		1			1
Both Sides Sidewalk	1			1	2
Cross Walk			1		1
Parking Lot	1		1		2
Stop Sign on Cross Street		1	1		2

All the variables in the table can be obtained from two sources – ODOT GIS database and Google Street View. TransGIS is a web mapping tool provided by ODOT, which represents layers and layers of data into an interactive map format and it serves as the standard foundation for ODOT web mapping applications. For non-state roads, the variables "Arterial", and "Urban" can be obtained from the functional classification, which is available through ODOT GIS database. Variables "Within City Limits" and "National Highway System" are also available through the ODOT GIS database. For local roads, none of the variables are available through ODOT GIS database. All other variables can be easily surveyed and determined using Google Street View.

8.2 FUTURE DATA SOURCES

The research team also evaluated alternate data sources which have the potential to improve AADT estimation. Specifically we focused on traffic data based on cell phone traces. Several private vendors have combined mobile phone GPS traces with other relevant data such as census and traffic counters to produce traffic volume, speed, and origin-destination demand estimates.

HERE technologies is a leader in providing agencies and corporations with mapping data and related services. HERE uses data collected from vehicle sensors, portable navigation devices (PNDs), road sensors, and connected cars to provide up-to-the-minute industry-leading traffic services. HERE technologies informed the research team that their estimations cover 100% of the roads in the 63 markets served.

Streetlight uses cell phone and GPS tracking data from vehicles to estimate O-D matrices, 2016 AADT, trip purpose, average travel time and travel time distribution, and commercial and personal travel vehicle comparisons. Streetlight's estimations are based on combining location-based services (LBS), GPS data, census, and a set of well-validated loop counters. Streetlight's 2016 AADT estimations cover roadway segments in both urban and rural areas. Streetlight's AADT predictions for the Virginia Department of Transportation look promising (R-squared of 0.87 and average error of 22%), but they are limited to roads with AADT of above 500 (*StreetLight's 2016 AADT Metric – Methodology and Validation Overview*). Recently *Turner et al. (2017)* compared StreetLight's AADT prediction with actual AADT data and found that errors were higher at lower volume roadways due to lower number of samples. The study recommended more detailed evaluation and further research for better volume estimations.

Another major vendor that has worked with ODOT in the past is INRIX. INRIX calculates the speed on each segment each minute and makes that data available as raw data files. INRIX predicts volume profile for the typical week, time and day in 15 minute bins. INRIX can predict traffic volumes for roads down to the minor collector level. INRIX bases their data by fusing information from commercial fleets, GPS, cell towers, mobile devices, and cameras. IdealSpot uses INRIX data to provide up-to-date traffic data down to the block level and segmented by time of day, day of week, and average speed. There are several other vendors similar to IdealSpot such as Miovision and Be-mobile.

Table 8-3 summarizes information on all of these private vendors. Since Be-mobile primarily operates in Europe and Miovision in Canada and Europe, we have not further reviewed these companies and their products.

Table 8-3: Summary of Companies using Mobile Phone Traces to Predict Traffic

Vendor	Website	Input data	Output Data (Type of Traffic Prediction?)	Coverage
Here	https://www.here.com/en/products-services/products/here-traffic/here-traffic-overview	vehicle sensor data, smartphones, PNDs, road sensors and connected cars; data updated every minute	up-to-the-minute industry-leading traffic service	100% of the roads in the 63 markets they serve
Streetlight	https://www.streetlightdata.com/	cell phone and GPS tracking data from vehicles; combines both Location-Based Services (LBS) and GPS data, as well as the census and a set of well-validated loop counters.	origin-destination matrices, select link analyses, 2016 AADT , trip purpose, average travel times, travel time distributions, commercial and personal travel vehicle comparisons	all type of roadway segments in both urban and rural areas.
INRIX	http://inrix.com/	commercial fleets, GPS, cell towers, mobile devices, and cameras	speed at a minute level aggregation, volume profile in 15 minute time bins	both urban and rural roads down the about the minor-collector level
IdealSpot	https://www.idealspot.com/	INRIX Data	up-to-date traffic data down to the block level and segmented by time of day, day of week, and average speed.	same as INRIX

All of the private vendors mentioned above rely on mobile phone traces and location information. We tried to estimate the usefulness of such services to estimate AADT in the various sub-regions of Region 2 based on the coverage of cellphones in the study region. All major cellphone companies in Oregon (Verizon, AT&T, T-mobile, and Sprint) provide up to date coverage maps on their website (T-mobile coverage map, Verizon coverage map, AT&T

coverage map, Sprint coverage map). The coverage maps indicate that AADT prediction based on mobile phone data might be adequate in MPO, valley-rural, and coastal sub-region. However, there is limited coverage in the mountain sub-region. Note that all of the coverage maps have been obtained from the websites of the respective cellphone companies. To date there has been no research on validating the accuracy of the coverage particularly outside of metro regions.

8.3 SUMMARY

This chapter reviewed the data requirements for the point based models for AADT estimation. In general, all the variables used were based on specific roadway, geometric, and signage characteristics which can be easily obtained from ODOT TransGIS database and Google Street View. Recently, there has been an increasing interest in estimating traffic data using location based services which primarily rely on GPS traces from mobile phones. We surveyed several private vendors which offer such services. HERE technologies and Streetlight appear to cover all roadway categories in the markets served whereas INRIX and IdealSpot provide data on roadway segments up to minor collector levels. We were not able to evaluate the usefulness of mobile phone data to improve AADT estimations. Though, we used cell phone coverage as a proxy as almost all of these services rely on cell phone traces. In general, we expect these methodologies to be unreliable in mountain sub-region due to spotty mobile phone coverage.

9.0 CONCLUSIONS

The objective of this research is to develop reliable and simple methods to predict AADT on non-state roadway segments with AADT of less than 10000. As a first step, a detailed literature review was conducted. The five most common methodologies used for estimating missing AADT information include regression, travel demand modeling, geospatial, machine learning, and image processing based approaches.

Regression based approaches were the most widely used. However, we found that most of the regression models were miss-specified, over fitted, or without a proper validation approach. Models which are over fitted based on variables in one region do not perform well when applied to other regions. Robust models that are parsimonious and intuitive are more likely to stand well the test of time and transferability. In this project, from our preliminary analysis, linear regression models were found to be unsuitable for AADT prediction. The statistical assumptions associated with linear regression models were not being satisfied. Moreover, the initial tests on model performance during validation were not good.

The research team also surveyed other state DOTs to identify any best practices for AADT estimation. Most state DOTs which responded to the survey have ad hoc procedures for AADT estimation or tend to use default values obtained from links with similar characteristics (functional classification, speed etc.) or adopt a county level average.

The research team performed two separate analysis. The first analysis focused on non-state upper functional classification roadways with AADT lower than 10000 excluding expressways, freeways, and ramps. The 2015 ODOT Region 2 AADT database was used as the basis for this analysis. The second analysis focused on non-state local roads. The 2016 ODOT Region 2 ADT database was used as the basis for this analysis.

The AADT was found to exhibit high levels of skew and many outliers. Therefore, we recommend using the median AADT instead of average AADT while selecting representative AADT. The AADT information was found to vary depending on the location. Therefore, we categorized the data based on functional classification and sub-regions (Coast, Mountain, Valley-rural, and MPO) and developed default values for AADT for non-state upper functional classification roadway systems. For local roads, we categorized the segments based on functional classification (urban, rural) and sub-regions and availability of Google Street View and developed default ADT values for each category.

A data collection form was developed focusing on variables which were found to be important in the literature review and other factors which were expected to affect AADT. We developed a stratified random sampling procedure to select roadway segments ensuring that all sub-regions and functional classification were adequately covered. The relevant variables were collected using Google Street View and ODOT GIS database.

A point based system was developed to predict median AADT for non-state upper functional classification roadway systems and median ADT for non-state local roads. Based on the descriptive analysis, we identified the set of the roadway, geometric, and land use related factors which are present in roadways segments with higher median AADT. We selected a subset of these factors and assigned them one point each. The total points are calculated for a roadway segment by adding up all the points. Default median AADT values (non-state upper functional classification roadway segments and ADT values for local roads are provided for each point score. Adjustment factors were used to reduce the error in the case of non-state upper functional classification roadway segments.

An overall Region 2 model as well as separate sub-regional models were developed for each of the sub-regions. It is important to highlight that most studies surveyed in the literature review did not follow a proper validation approach. We applied the stratified random sampling procedure to select roadway segments while ensuring all sub-regions and functional classification was adequately represented for both model development and validation. The prediction accuracy of the models was tested on separate validation data. For the non-state upper functional classification roadway systems, 290 points were used for model development and 90 points were used for validation. For the local roads, 152 points were used for model development and 39 points were used for validation.

Two different metrics were studied to characterize the accuracy of the point model: (i) median of the errors across all points in the validation data set and (ii) average of the errors across all points in the validation data set. For the non-state upper functional classification roadway system model, the overall region 2 model slightly under-predicts median AADT values. The sub-regional models provided significantly lower median errors for the Coastal and Valley-rural sub-regions. In general, the model errors were found to be reasonable on roadway segments with an AADT less than 5000. For local roads, the overall model had a median error of -32 which indicates that the model slightly under-predicts the ADT. The overall model has the lowest median error of 4 for the Valley-rural sub-region. The gains in accuracy by using the sub-region model are not high. Therefore, we recommend using the overall (regional) model. Overall, the developed point based methodology is appropriate to estimate the median AADT for a given type of roadway.

During the model development phase, the research team ensured that the models avoid subjective variables which are difficult to transfer from one location to another. All of the variables used in the model are highly objective and can be easily obtained from ODOT GIS database or Google Street View.

9.1 RECOMMENDATIONS FOR FUTURE RESEARCH

This research can be extended in multiple directions. To apply these models in other ODOT regions, as a first step we recommend identifying sub-regions and generating default values based on sub-regions, functional classification, and availability of Google Street View. The principle of point based modeling can be transferred to other regions but the default median values for each point and the adjustment factors can potentially vary. However, we do believe that for rural regions with similar socio-demographic, land use, and economic characteristics as the Coastal, Mountain, and Valley-rural sub-region as region 2, the default values and the point system should more easily transfer.

In addition to the roadway, geometric, and land use characteristics, the AADT of a roadway segment is also a function of the traffic generators alongside the roadway segment. While we have accounted for those in our point based model, we only looked at a few thousand feet upstream and downstream of the segment of interest. However, for rural roadway segments, sometimes these generators may be located 10 to 20 miles away from the segment. A roadway segment passing through extremely rural areas might have higher than expected AADT because it might be the only route between two small towns. Characterizing such effects in an objective manner is not an easy task and will require a more detailed study.

New data sources based on cell phone traces provide an intriguing way to estimate AADT. However, there is limited information on the coverage of these data sources particularly as you go outside the city limits.

10.0 REFERENCES

4G LTE Coverage Map | Check Your 4G LTE Cell Phone Coverage | T-Mobile. (n.d.). Retrieved March 15, 2018, from https://www.t-mobile.com/coverage/coverage-map

American Association State, H., & Transport. (1992). *AASHTO guidelines for traffic data programs*. Washington: AASHTO.

Anderson, M. D., Sharfi, K., & Gholston, S. E. (2006). Direct demand forecasting model for small urban communities using multiple linear regression. *Transportation research record.*(1981), 114-117.

AT&T Maps. (n.d.). Retrieved March 15, 2018, from https://www.att.com/maps/wireless-coverage.html

Barnett, J. S., Portland State, U., Department of, C., & Environmental, E. (2015). On the Estimation of Volumes of Roadways: An Investigation of Stop-Controlled Minor Legs.

Bingduo, Y., Sheng-Guo, W., & Yuanlu, B. (2014). New Efficient Regression Method for Local AADT Estimation via SCAD Variable Selection. *IEEE Trans. Intell. Transport. Syst. IEEE Transactions on Intelligent Transportation Systems, 15*(6), 2726-2731.

Castro-Neto, M., Jeong, Y., Jeong, M. K., & Han, L. D. (2009). AADT prediction using support vector regression with data-dependent parameters. *Expert Systems with Applications Expert Systems with Applications, 36*(2), 2979-2986.

Chowdhury, R. Quarterly Research Progress Report. Research Project No. SPR 717, Clemson University, South Carolina, 2015.

City of Salem. (2016). *Street System Element* (Publication). Salem, OR: City of Salem. Retrieved from https://www.cityofsalem.net/CityDocuments/tsp-street.pdf .

Cleveland, M. (2003). *The effects of errors in annual average daily forecasting: study of highways in rural Idaho*. Available from https://apps.itd.idaho.gov/apps/research/Completed/RP167.pdf .

Dixon, K., Oregon, Department of, T., Research, S., Oregon Transportation, R., Education, C., . . Federal Highway, A. (2012). Calibrating the future Highway safety manual predictive methods for Oregon state highways.

Eom, J. K. (2006). Improving the prediction of annual average daily traffic for nonfreeway facilities by applying a spatial statistical method. *Transportation research record.* (1968), 20-29.

Figliozzi, M., Johnson, P., Monsere, C., & Nordback, K. (2014). Methodology to Characterize Ideal Short-Term Counting Conditions and Improve AADT Estimation Accuracy Using a Regression-Based Correcting Function. *JOURNAL OF TRANSPORTATION ENGINEERING, 140*(5), 04014014.

Introducing StreetLight Volume: 2016 AADT Metrics. (2018, April 10). Retrieved March 15, 2018, from https://www.streetlightdata.com/introducing-streetlight-volume-2016-aadt-metrics

Jayasinghe, A., & Sano, K. (2017). *Estimation of Annual Average Daily Traffic on Road Segments: Network Centrality-Based Method for Metropolitan Areas.*

Jiang, Z., McCord, M. R., & Goel, P. K. (2006). Improved AADT Estimation by Combining Information in Image- and Ground-Based Traffic Data. *JOURNAL OF TRANSPORTATION ENGINEERING, 132*(7), 523-530.

Keehan, McKenzie, "Annual Average Daily Traffic (AADT) Estimation with Regression Using Centrality and Roadway Characteristic Variables" (2017). *All Theses*. 2644. https://tigerprints.clemson.edu/all_theses/2644

Lam, W. H. K., & Jianmin, X. (2001). Estimation of AADT from short period counts in Hong Kong-a comparison between neural network method and regression analysis. *Journal of Planning Literature, 15*(3).

Lowry, M. (2014). Spatial interpolation of traffic counts based on origin-destination centrality. *JTRG Journal of Transport Geography, 36*, 98-105.

Lowry, M., & Dixon, M. (2012). *GIS Tools to estimate Average Annual Daily Traffic* (Publication No. KLK725). Washington, D.C.: US Dept of Transportation. Retrieved from https://rosap.ntl.bts.gov/view/dot/24942/dot_24942_DS1.pdf.

Membah, J., Shamo, B., & Asa, E. (2015). Linear Spatial Interpolation and Analysis of Annual Average Daily Traffic Data. *J. Comput. Civ. Eng. Journal of Computing in Civil Engineering, 29*(1).

Mohamad, D., Sinha, K. C., Kuczek, T., & Scholer, C. F. (1998). Annual Average Daily Traffic Prediction Model for County Roads. *TRANSPORTATION RESEARCH RECORD*(1617), 69-77.

Musunuru, A., Wei, R., & Porter, R. J. (2017). Predicting Day and Night Traffic Volumes on Rural Roads for Statistical Road Safety Modeling. *Transportation Research Record: Journal of the Transportation Research Board Transportation Research Record: Journal of the Transportation Research Board, 2659*, 192-203.

ODOT GIS Database. (n.d.). Retrieved January 26, 2017, from ftp://ftp.odot.state.or.us /tdb/trandata/GIS_data .

Oregon Department of Transportation. (n.d.). *ODOT TransGIS* [Map]. Retrieved January 26, 2017, from https://gis.odot.state.or.us/transgis/

Oregon DMV Vehicle Registration Statistics. (n.d.). Retrieved February 1, 2017, from https://www.oregon.gov/ODOT/DMV/pages/news/vehicle_stats.aspx

Oregon Spatial Data Library. (n.d.). Oregon Zoning. Retrieved January 26, 2017, from https://spatialdata.oregonexplorer.info/geoportal/catalog/search/resource/details.page?uuid={9966F34D-71E7-4BD5-A91E-0D2757C91EBF}

Pan, T. (2008). *Assignment of estimated average annual daily traffic on all roads in Florida.* University of South Florida, [Tampa, Fla]. Available https://scholarcommons.usf.edu/cgi/viewcontent.cgi?article=1441&context=etd

Pulugurtha, S. S., & Kusam, P. R. (2012). Modeling Annual Average Daily Traffic with Integrated Spatial Data from Multiple Network Buffer Bandwidths. *Transportation Research Record: Journal of the Transportation Research Board, 2291*(1), 53-60. Doi: 10.3141/2291-07

Pulugurtha, S. S., & Kusam, P. R. (2016). Spatial Proximity and Dependency to Model Urban Travel Demand. *J. Urban Plann. Dev. Journal of Urban Planning and Development, 142*(2).

Seaver, W. L., Chatterjee, A., & Seaver, M. L. (2000). Estimation of traffic volume on rural local roads. *Transportation research record.* (1719).

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

Sharma, S., Lingras, P., Xu, F., & Liu, G. (1999). Neural Networks as Alternative to Traditional Factor Approach of Annual Average Daily Traffic Estimation from Traffic Counts. *Transportation Research Record: Journal of the Transportation Research Board, 1660*, 24-31. Doi: 10.3141/1660-04

Sharma, S., Lingras, P., Liu, G., & Xu, F. (2000). Estimation of Annual Average Daily Traffic on Low-Volume Roads: Factor Approach Versus Neural Networks. *Transportation Research Record: Journal of the Transportation Research Board, 1719*, 103-111. doi:10.3141/1719-13

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

Sprint Coverage Map. (n.d.). Retrieved March 15, 2018, from https://coverage.sprint.com/IMPACT.jsp

Staats, W. N., Souleyrette, R. R., University of, K., College of, E., University of, K., & Department of Civil, E. (2016). Estimation of annual average daily traffic on local roads in Kentucky.

Sun, X., Das, S., Louisiana Transportation Research, C., Louisiana, Department of, T., Development, . . . Federal Highway, A. (2015). *Developing a Method for Estimating AADT on all Louisiana Roads*.

Tang, Y. F., Lam, W. H. K., & Ng, P. L. P. (2003). Comparison of Four Modeling Techniques for Short-Term AADT Forecasting in Hong Kong. *JOURNAL OF TRANSPORTATION ENGINEERING*, *129*, 271-277.

Turner, S., Koeneman, P., Minnesota, Department of, T., Research, S., Library, . . . Institute, M. T. (2017). *Using mobile device samples to estimate traffic volumes*. St. Paul, Minn.: Minnesota Department of Transportation, Research Services & Library.

United, S., & Federal Highway, A. (2013). Traffic monitoring guide.

Verizon coverage - OpenSignal. (n.d.). Retrieved March 15, 2018, from https://opensignal.com/networks/usa/verizon-coverage

Wang, T., Gan, A., & Alluri, P. (2013). Estimating Annual Average Daily Traffic for Local Roads for Highway Safety Analysis. *Transportation Research Record: Journal of the Transportation Research Board, 2398,* 60-66. Doi: 10.3141/2398-07

Wang, X., & Kockelman, K. (2009). Forecasting Network Data. *Transportation Research Record: Journal of the Transportation Research Board, 2105,* 100-108. Doi: 10.3141/2105-13

Xia, Q., Zhao, F., Chen, Z., Shen, L., & Ospina, D. (1999). Estimation of Annual Average Daily Traffic for Nonstate Roads in a Florida County. *Transportation Research Record: Journal of the Transportation Research Board, 1660,* 32-40. Doi: 10.3141/1660-05

Zhao, F., & Chung, S. (2002). Contributing factors of annual average daily traffic in a Florida county: exploration with Geographic Information System and regression models. *Transportation research record.* (1769), 113-122.

Zhao, F., & Park, N. (2004). Using geographically weighted regression models to estimate annual average daily traffic. *Transportation research record.* (1879), 99-107.

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

APPENDIX A

The most common way to express motor vehicle traffic counts is to estimate the annual average daily traffic (AADT). This value represents the annual average 24 hour two-way count on a facility. Because it is impractical to count everywhere on a continuous basis, most counts are short duration. Since traffic has temporal and seasonal variation methods have been developed to estimate AADT from the short duration counts. This chapter provides a review of factoring methods and identifies existing approaches to factoring (i.e. the adjustments and statistical analysis that must be introduced in order to estimate AADT values from short duration counts).

A.1 VARIABILITY OF NON-MOTORIZED TRAFFIC

Counts vary over time, for example Table A-1 provides data from a major commute freeway (Interstate 84 half mile east of Interstate 5). For motorized counts, employing both day of the week and monthly seasonal factors is usually sufficiently accurate to estimate AADT volumes using short-term counts

Table A-1: Percent AADT by Month and Vehicle Type

MONTH OF YEAR	MOTOR VEHICLES I-84
January	96%
February	99%
March	100%
April	102%
May	102%
June	103%
July	103%
August	101%
September	100%
October	101%
November	96%
December	96%

A.2 MOTOR VEHICLE FACTORING METHODS

For motor vehicles, there are well established factoring methods and practices. Section 3 of the Traffic Monitoring Guide (TMG) provides guidelines for data collection and monitoring of motor vehicle traffic (FHWA 2013).

The TMG framework for counting and data collection consists of a set of permanent continuous counting sites and a complementary short duration count program, usually collected in durations of one day to one month. The primary interests for collecting motor vehicle data are to determine average annual daily traffic (AADT) volumes and to meet the data reporting demands of the Highway Performance Monitoring System (HPMS) which allocates federal funds (FHWA 2013).

The AADT is an estimate of the average daily traffic that occurs over the year over a section of a facility. Due to the temporal (seasonal, weekly, daily) variation in traffic demand, the 24-hour count of any one day will likely overestimate or underestimate the actual annual average, i.e. estimates always have an associated error. Thus, to reduce AADT estimation errors all short-term traffic counts are corrected by day of the week (DOW) and monthly adjustment factors. The permanent count sites are needed for establishing temporal trends and estimating adjustment factors for short duration counts. Accordingly, the TMG recommends developing factor groups for motor vehicles that include vehicle type, day of week, seasonal adjustment, axle correction, and growth factors from long term continuous sites in order to estimate traffic volumes from short-term counts at other locations.

A.3 MOTOR VEHICLE FACTORING METHODS FROM CONTINUOUS COUNTS

There are two primary procedures for calculating AADT from permanent, 365 days - 24 hour counting stations, also referred to as automated traffic recorders (ATR); one is a simple sum of all daily volumes for one year divided by 365 days and the other is an average of averages (FHWA 2013). The AADT calculation for averages of averages from continuous counts comes from the AASHTO Guidelines for Traffic Data Programs, prepared in 1992 (AASHTO 1992) One outcome of the method to calculate the average of averages is estimates for day of week (DOW) and monthly seasonal factors. The procedure for the AASHTO method of determining AADT using continuous counts are as follows:

1. Calculate the average for each DOW for each month to derive each monthly average DOW.

2. Average each monthly average DOW across all months to derive the annual average DOW.

3. The AADT is the mean of all of the annual average DOW.

The formula for the AASHTO method for determining AADT is:

$$\boldsymbol{AADT = \frac{1}{7}\sum_{i=1}^{7}\left[\frac{1}{12}\sum_{j=1}^{12}\left(\frac{1}{n}\sum_{k=1}^{n} VOL_{ijk}\right)\right]} \tag{A-1}$$

Where:

VOL= daily traffic for day k, of day of the week *i*, and month *j*

i = day of the week

j = month of the year

k = index to identify the occurrence of a day of week *i* in month *j*

n = the number of occurrences of day *i* of the week during month *j*

It is preferred to use at least one year of continuous data for determining AADT and corresponding factors. Multi-year data are better for to account for growth trend impacts. Estimates are then used to extrapolate estimated AADT values from short-term counts at similar or nearby locations (AASHTO 1992). Multi-year data produce better factors to estimate AADT.

ODOT's Transportation Systems Monitoring Unit uses similar methods to AASHTO for determining AADT. The procedure for the ODOT method of determining AADT using continuous counts are:

4. Calculate the average for each DOW for each month to derive each monthly average DOW
5. Average the monthly average DOW for each month to derive the annual average day of the month
6. The AADT is the mean of all of the annual average days of the month

The formula for the ODOT method of determining AADT is given as:

$$\boldsymbol{AADT = \frac{1}{12}\sum_{j=1}^{12}\left[\frac{1}{7}\sum_{i=1}^{7}\left(\frac{1}{n}\sum_{k=1}^{n} VOL_{ijk}\right)\right]}$$

(A-2)

Where:

VOL= daily traffic for day *k*, of day of the week *i*, and month *j*

i = day of the week

j = month of the year

k = index to identify the occurrence of the day of week *i* in month *j*

n = the number of occurrences of day *i* of the week during month *j*

Essentially, the AASHTO procedure of determining the AADT is to average the volumes for each DOW in each month, then average each DOW across all months. Lastly, take the average of the seven annual averages of the DOW. The ODOT procedure switches the last two steps of the AASHTO procedure by averaging all the average DOW for each month to develop an average day of the month and then average all twelve of the monthly averages to determine the AADT. The AADT using the AASHTO or ODOT procedure yields the same results.

A.4 AADT ESTIMATION FROM SHORT DURATION COUNTS

For short duration count locations, AADT must be estimated. Because the short duration count only captures the traffic in one particular season, month, week, day, or hour, this short-term count must be adjusted. To estimate AADT using short-term counts, axle counts are converted to AADT using the following equation from the TMG:

$$\boldsymbol{AADT_{est\,hi} = VOL_{hi} * M_h * D_h * A_i * G_h}$$

(A-3)

Where:

$AADT_{est\,hi}$ = the estimated annual average daily travel at location i of factor group h

VOl_{hi} = the 24-hour axle volume at location i of factor group h

M_h = the applicable seasonal (monthly) factor for factor group h

D_h = the applicable day-of-week factor for factor group h (if needed)

A_i = the applicable axle-correction factor for location i (if needed)

G_h = the applicable growth factor for factor group h (if needed)

No specific method is given for determining seasonal, DOW, growth, or axle correction factors. However, the TMG does recommend the AASHTO method for determining monthly factors for motor vehicles (FHWA 2013). The monthly factor for each long term ATR is the ratio of the AADT to MADT. Once it has been verified that the ATR station has been running reliably, then the AADT should be determined using AASHTO formula (AASHTO 1992).

A.5 QUALITY CONTROL FOR USING MOTOR VEHICLE COUNTS

Quality control is also an important part of counting programs. When data records are missing or suspect due to machine malfunction or atypical traffic periods, the above procedures must be adapted or modified. There are different methods for validating permanent and short-term count data. Methods may vary depending on each unique situation and missing data. If long term, historical data exists, missing count data may be estimated using historical data. If other count sites are nearby or have similar patterns, this data can also be used to make adjustments and estimations. If directional data is collected and only one data collection device fails, then the data from the other direction can help determine estimates for missing data (AASHTO 1992).

APPENDIX B

B.1 DATA COLLECTION FORM FOR NON-STATE ROADS

Table B-1 shows the data collection used in surveying the selected points for non-state roads:

Table B-1: Data collection form for non-state roads

	AADT Estimation Survey Form	**Choose one of the below options or write the answer for each question**
Land Use	What type of AADT generators exist adjacent to the road with the main access on the study road segment? *	Gas Station Hospital/Medical Center School Recreational Facility Shopping Center None
	Primary dominant land use adjacent to the study road segment*	Residential Commercial Industrial Forest Farm Other
	Are there any single houses adjacent to the study road segment? *	Yes No
MIRE	Type of median	Undivided One-Way Single-Line Vegetation Flush Raised Depressed 2-Way-LT Transit Other
	Does the road segment operate as one-way or two-way?	One-way Two-way
	Number of through lanes	
	Degree of access control	Full Access Partial Access No Access
Intersection	Is there a right-turn lane along the study road segment? (on either direction) *	Yes No
	Is there a left-turn lane along the study road segment? (on either direction) *	Yes No
	Are there any traffic signals along the study road segment? *	Yes No
Roadway	Is the study road segment paved? *	Yes No
	Presence of pavement horizontal marking along the study road segment* (Note that bicycle lane markings are not counted as shoulder marking)	Lane Marking Shoulder Marking Both None
	Presence of shoulder on the study road segment*	Paved Unpaved None
	Are there any marked crosswalks along the study road segment? * (either crossing the main corridor of study or on any side of intersections along the main corridor of study)	Yes No
	Presence of sidewalk along the study road segment*	Both sides One side None

	Presence of bike lane along the study road segment*	Yes	No
	Are there any bus stops along the study road segment? *	Yes	No
	Are there any parking lots** adjacent to the road with an access to the study road segment? *	Yes	No
	Type of traffic calming device*** present along the study road segment*	Yes	No
Signage	Are there any stop signs on the downstream and upstream intersections? * (on the cross roads, not the study road segment)	Yes	No
	Are there any stop signs along the road? * (looking at the main corridor of study and not the intersecting approaches)	Yes	No
	Are there any signs, other than stop sings, along the road? * (looking at the main corridor of study and not the intersecting approaches)	Yes	No
ODOT GIS Data	Distance from the midpoint to the nearest state and non-state Highway		
	Distance from the midpoint to the nearest state and non-state arterial		
	Is the road segment located in an MPO?	Yes	No
	Distance from the midpoint to the National Highway System (NHS)		
*1000 feet upstream and downstream of the midpoint			
**Including pay to park, parking lots, and parking lots for schools, shopping centers, recreational facilities, hospitals, etc.			
***Traffic calming devices include Speed humps, Neighborhood traffic circles, speed tables, Chicanes, Raised intersection, Choker, Closure, and center island narrowing (according to ITE)			

B.2 DATA COLLECTION FORM FOR LOCAL ROADS

Table B-2 shows the data collection used in surveying the selected points for non-state roads:

Table B-2: Data collection form for local roads

ADT Estimation Survey Form		Choose one of the below options or write the answer for each question
Land Use	What type of ADT generators exist adjacent to the road with the main access on the study road segment? *	Gas Station Hospital/Medical Center School Recreational Facility Shopping Center None
	Dominant land use adjacent to the study road segment*	Residential Commercial Industrial Forest Farm Other
	Are there any single houses adjacent to the study road segment? *	Yes No
	Is Google Street View available on the study road segment?	Yes No
	What is the topology of the area in which the study road segment is located?	Coast Mountain Valley-Rural MPO
MIRE	Type of median	Undivided One-Way Single-Line Other
Intersection	Is there a right-turn lane along the study road segment? (on either direction) *	Yes No
	Is there a left-turn lane along the study road segment? (on either direction) *	Yes No
	Are there any traffic signals along the study road segment? *	Yes No
Roadway	What type is the road segment pavement? *	Paved Gravel Unpaved None
	Presence of pavement horizontal marking along the study road segment* (Note that bicycle lane markings are not counted as shoulder marking)	Lane Marking Shoulder Marking Both None
	Presence of shoulder on the study road segment*	Paved Unpaved None
	Are there any marked crosswalks along the study road segment? * (either crossing the main corridor of study or on any side of intersections along the main corridor of study)	Yes No
	Presence of sidewalk along the study road segment*	Both sides One side None
	Presence of bike lane along the study road segment*	Yes No
	Are there any bus stops along the study road segment? *	Yes No

	Are there any parking lots** adjacent to the road with an access to the study road segment? *	Yes	No
	Presence of traffic calming devices*** present along the study road segment*	Yes	No
Signage	Are there any stop signs on the downstream and upstream intersections? * (on the cross roads, not the study road segment)	Yes	No
	Are there any stop signs along the road? * (looking at the main corridor of study and not the intersecting approaches)	Yes	No
	Are there any signs, other than stop sings, along the road? * (looking at the main corridor of study and not the intersecting approaches)	Yes	No
*1000 feet upstream and downstream of the midpoint			
**Including pay to park, parking lots, and parking lots for schools, shopping centers, recreational facilities, hospitals, etc.			
***Traffic calming devices include speed humps, neighborhood traffic circles, speed tables, chicanes, raised intersection, choker, closure, and center island narrowing (according to ITE)			