View metadata, citation and similar papers at core.ac.uk

brought to you by CORE

provided by Clemson University: TigerPrints

Clemson University
TigerPrints

All Theses

Theses

5-2017

Annual Average Daily Traffic (AADT) Estimation with Regression Using Centrality and Roadway Characteristic Variables

McKenzie Keehan
Clemson University, mkeehan@clemson.edu

Follow this and additional works at: https://tigerprints.clemson.edu/all_theses

Recommended Citation

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

This Thesis is brought to you for free and open access by the Theses at TigerPrints. It has been accepted for inclusion in All Theses by an authorized administrator of TigerPrints. For more information, please contact kokeefe@clemson.edu.

ANNUAL AVERAGE DAILY TRAFFIC (AADT) ESTIMATION WITH REGRESSION USING CENTRALITY AND ROADWAY CHARACTERISTIC VARIABLES

A Thesis
Presented to
the Graduate School of
Clemson University

In Partial Fulfillment
of the Requirements for the Degree
Master of Science
Civil Engineering

---

by
McKenzie Keehan
May 2017

Accepted by:
Mashrur Chowdry, Committee Chair
Wayne Sarasua
Eric Morris

ABSTRACT

Accurate estimation of annual average daily traffic (AADT) is critical in nearly every roadway decision, such as allocations of funding for roadway improvements and maintenance. While some roadway locations have permanent count stations capable of counting vehicles 24-hours a day throughout the entire year, they are typically only installed at selected locations on major roadways (i.e., freeways and major arterials) with high traffic volumes. On lower functional class roads and roadway segments on higher functional class roads without permanent count stations, short-term coverage counts are collected and adjusted with data from permanent count stations to estimate AADT. Short-term coverage counts are essential because they provide data from roadways of all functional classes and lane configurations, accounting for varying volumes on all roads maintained by an agency. Although necessary, coverage counts can be expensive and can exhaust resources such as investment in data collection workforce, equipment and data analysis. This study develops a strategy for estimating AADT on every roadway within a given jurisdiction using permanent count stations and short term coverage counts, while limiting the number of coverage counts needed. The goal of this thesis is to illustrate a noteworthy time and cost savings using a new centrality based AADT estimation method. A set of new deterministic variables, based on the theory of centrality, are introduced. This study revealed that estimated root mean square error (RMSE) for the new centrality based AADT method is half of the estimated RMSE in the travel demand based AADT model for the same area. Additionally, it was found that using centrality based AADT estimation model, the number of coverage count stations necessary can be reduced by more than 60% compared to the

standard factor method for AADT estimation without compromising the AADT estimation accuracy.

DEDICATION

I would like to dedicate this thesis to my mother and father, for their unconditional love and support.

ACKNOWLEDGMENTS

I would like to express so much gratitude and appreciation to my advisor, Dr. Mashrur Chowdhury, for constantly challenging me and aiding in my growth as a researcher and graduate student. During my last two years under his wing, I have seen myself evolve in ways I could have never imagined, and it has been due to his relentless support and guidance. He has helped me to exceed expectations I had previously set for myself, and has helped me to realize that there are no limitations to our potential.

I would also like to thank Dr. Wayne Sarasua and Dr. Eric Morris for serving as my thesis committee members, and providing valuable feedback throughout my research. Dr. Sarasua has also taught me in several courses and has proved to be a passionate professor who cares deeply about the success of each of his students.

I appreciate the administrative staff members from the Glenn Department of Civil Engineering, including Kristi Baker, Sandi Priddy, CJ Bolding, Karen Lanning, and Monica Hughes, for helping me with all of the managerial and organizational aspects of my position and my needs as a student.

I would like to extend my deep appreciation to Dr. Kakan Dey for being an amazing mentor through his continuous effort to improve the research quality and for always being there whenever I needed him. I would like to thank Md Mizanur Rahman, Sakib Kahn, and Sababa Islam, for being the best mentors a fellow student could ask for. Their kindness and expertise have done so much for me as I student and I would not have been able to be as successful in my student career if not for them. Moreover, I would like to thank Md Mhafuzul

Islam and Md Zadid khan for helping me with many projects, including class projects, competitions, and presentations, which were all made better due to their help and guidance.

I would like to thank the South Carolina Department of Transportation (SCDOT) for funding a lot of the work I've done throughout my graduate student career, and providing me with the data that were necessary for my research.

I cordially thank my parents, my siblings, and my entire family for being the support system for me during any critical times. They have always been my inspiration to reach my goals.

TABLE OF CONTENTS

Table of Contents (Continued)

Page

LIST OF FIGURES

Page

LIST OF TABLES

CHAPTER ONE

INTRODUCTION

1.1 Problem Statement

Annual average daily traffic (AADT) is defined as the average daily measure of the total volume of vehicles on a roadway segment over a year. Traffic volumes are the lead indication travel demand and utilization of the roadways within a specified network (*1, 2*). Therefore, accurate traffic volume estimations are critical in nearly every roadway decision.

While some roadway locations have permanent count stations capable of collecting vehicle volumes 24-hours a day throughout the entire year, they are very costly to implement and maintain, and are typically only provided at selected roadway segments. Therefore, short term coverage counts are taken at thousands of strategically placed roadway segments of all functional classes and lane configurations, accounting for varying volumes on the roads maintained by an agency (*3*). The counts at these locations are usually only collected once a year or every few years with pneumatic tubes, and the data collection period can last from 24 hours to 7 days, depending on the responsible transportation agency's policy and reporting requirements (*3, 4*). For short-term count stations, the permanent count stations serve as control counts, and are used to determine daily, monthly, and seasonal factors to calibrate the data collected at short-term count stations in estimating AADT.

However, operating and estimating AADT using coverage counts can be expensive and can exhaust a significant amount of resources in terms of manpower, equipment, and data analysis. In addition, the data provided by these counts is limited, and not always sufficient in predicting accurate AADT values. For example, a 24-hour count on one day throughout the entire year may need to be used to calculate AADT using only the one count and calculated daily and monthly factors using the factor method, as further explained in the Chapter 2: Literature Review. Any inconsistencies between the day's data and the other elements of that day and month may result in inaccurate AADT estimation.

1.2 Objective of the Thesis

This thesis aims to develop a unique means of estimating AADT on roadways within a given jurisdiction, while limiting the required amount of implementation time and effort. All of the data used in development of the AADT estimation method is readily available. The motivation for this research is to develop an AADT estimation method, which any jurisdiction can implement with minimal time and effort.

1.3 Organization of the Thesis

Chapter 2 describes the background research and literature on the currently available AADT estimation methods. Some of these methods are currently being utilized by state and local transportation agencies, while others are simply being used in academia for research purposes. Additionally, Chapter 2 discusses the theory of centrality, its applications, and its potential in AADT estimation. Chapter 3 discusses the method for applying the theory of centrality in AADT estimation used in this thesis. The steps used in the application are outlined and explained in detail.

Chapter 4 provides a detailed explanation of the linear regression model produced by this method. Additionally, Chapter 4 compares the newly developed model to an existing travel demand forecasting model for the city, both conceptually and statistically. Next in Chapter 4, a procedure is performed in order to determine if this method could reduce the number of short term count stations that the city of Greenville should use in order to produce AADT estimates for all current short term count locations. Finally in Chapter 4, a cost savings analysis is performed to determine the financial competence of this method. Chapter 5 concludes this thesis, offering conclusions and recommendations for further research.

CHAPTER TWO

LITERATURE REVIEW

2.1 Overview

Because quality AADT estimation on local roads is vital, ample amounts of research have been carried out in order for models to be developed that can adequately estimate the AADT of every roadway within a given area (*2,5,6*), which is discussed further in this chapter. It is important to note the amount of time and costs associated with these often inaccurate or inconsistent methods.

In this chapter we have reviewed AADT estimation methods and the theory of centrality, as well as its feasibility in AADT estimation.

2.2 AADT Estimation Methods

Although collecting traffic count data every day in a year is the most accurate method of calculating AADT, it is not economically feasible to install and maintain data collection systems on a widespread scale. Traditionally, for roadways without permanent count stations, AADT is calculated using the America Association of State Highway and Transportation Official's (AASHTO's) factor method. Using this method, daily and monthly factors are calculated using permanent counts stations, and typically each permanent counts station is associated with a group of short term count locations, based on clustering. 24-hour traffic counts from short term count stations are then multiplied with the adjustment factors to find AADT, as shown in the following equation:

Eqn. 1 $AADT = V_{24ab} \times DF_a \times MF_b$

where V_{24ab} = 24-hour volume for day a in month b (vehs); DF_a = daily adjustment factor for day a; and MF_b = monthly adjustment factor for month b (*1*). Another AADT estimation method, used by some states and implemented in programs such as the Traffic Count Database System (*7*), used additional factors such as a seasonal factor and an axel-correction factor, as shown in the following equation:

Eqn. 2 $AADT = V_{24ab} \times SF_{ab} \times AF_c$

where V_{24ab} = 24-hour volume for day a in month b (vehs); SF_{ab} = seasonal adjustment factor for day a and month b; and AF_c = applicable axel-correction factor for c number of axels (*7*).

However, these formulas can only be used on roadway segments that have short term count locations, which are not used to collect data on most publically maintained roadways within a given state or jurisdiction. Therefore, ample amounts of research have been carried out in order for models to be developed that can adequately estimate the AADT of every roadway within a given area (*2*,*5*,*6*).

Several studies have shown that linear regression that utilizes roadway characteristics and socioeconomic factors at short term count locations can be used to estimate AADT (*2*,*5*,*6*,*8*). Doustmohammadi et al. developed a linear regression model to calculate AADT using a variety of socio-economic factors and roadway data for small and medium sized urban communities in Alabama (*5*). The significant variables identified in the final two models (one for a small city and one for a large city) included the functional classification (FCLASS) and lane counts of roadway segments (LANE), in addition to population, retail employment (RETAILEMBUFF), and all-other employment (NONRETAILEMBUFF) all inside a 0.25 mile radius around the traffic count site. These two models are shown below:

Model 1 ($R^2 = 0.82$): AADT = -5625 + 8493 FCLASS + 219 LANE - 1.16 POPBUFF -0.58 NONRETAILEMBUFF + 11.55 RETAILEMBUFF

Model 2 ($R^2 = 0.79$): AADT = -12590 + 4479 FCLASS - 1.15 POPBUFF – 0.86 NONRETAILEMBUFF +7.91 RETAILEMBUFF

It was concluded that their AADT estimation models can accurately estimate AADT on desired roadways in cities of similar populations (*5*). However, the accuracy and means of collecting the socio-economic factors used are resource intensive and difficult for annually updating AADT. For example, the population data was obtained from the Census Department, where data is only updated every 5 years (*9*). In this study, the other two socio-economic factors - retail and all-other employment – were found from case studies that were completed as a part of long-range transportation plans recently completed by a third party organization. This socio-economic data can be time consuming to collect, and relies on sources not maintained by the DOT. Additionally, socio-economic factors determined by surveys or sampling may not be accurate or capable of validation.

Zhao et al. used regression and further statistical analysis to find factors supplemental to AADT estimation in a study conducted for Florida. The study used geographic information system (GIS) technology to investigate various factors that may be good predictor of AADT on a road (*2*). A variety of land-use and accessibility measurements were developed and tested in (*2*). The four models developed achieved R^2

values from 0.66 to 0.82. The variables for each are: i) Model 1: Lane count, functional class, access to employment centers, directness of access to expressways, and employment inside a 0.25 mile radius (R^2= 0.8180); ii) Model 2: Lane count, access to employment centers, directness of expressway access, and network distance to the mean centers of population (R^2= 0.6607); iii) Model 3: Lane count, access to regional employment centers, directness of expressway access, network distance to the regional mean centers of population, and population inside a 0.25 mile radius around a traffic count site (R^2= 0.7624); and iv) Model 4: Lane count, access to regional employment centers, directness of expressway access, network distance to the regional mean centers of population, employment inside a 0.25 mile radius around a traffic count site, and population inside a 0.25 mile radius around a traffic count site (R^2= 0.7648). During their data collection efforts, the author used employment data, which was purchased from a third party, containing the number of employees at each business location and the standard industrial classification code. In this case, the data had been purchased for a prior project, however, for most DOTs, implementing this method would not only require attaining or purchasing additional data not readily available to their agency, but also relying on data not operated and maintained by the DOT and only updated at the third party's discretion.

Wang et al. presented a means to estimate AADT for roadways through travel demand forecasting (*6*). A main factor of applying the travel demand model involved using land-use data at the parcel level to determine estimated trips produced from or attracted to each parcel. All-or-nothing trip assignment was conducted using free-flow travel times. Then, the trips were dispersed through a trip distribution gravity model at the parcel-level.

The results show that the proposed model generated 52% MAPE, which is 159% lower than the MAPE from regression models ran for the same area (*6*). While travel demand modeling methods have proven accurate in AADT estimation, they are often time consuming to develop and require a lot of data collection resources and modeling expertise. The classic four step travel demand model is complex, time consuming, and costly for many jurisdictions. This is explained in much further detail in the Chapter 4: Results and Analysis, where the newly developed model is compared to an existing Travel Demand Model.

Artificial Neural Networks are also commonly utilized as a means to find AADT on roadways. Sharma et al. developed two models for estimating AADT using an Artificial Neural Network – one using the previous 48-hour count data and one using the previous two 48-hour count data (*10*). When compared to the traditional factor model, the artificial neural network models proved to be less accurate. However, an advantage of the neural network models is that they did not require the ATRs to be grouped, unlike in the factor method. The study found that the errors of AADT estimation using the factor approach could be lowered by grouping the ATR sites appropriately and accurately assigning short-term count stations to each ATR site. For two 48-hour counts, the 95th-percentile error is between 14.14 to 16.68 percent, as compared with the range of 16.77 to 24.89 percent for a single 48-hour count. While their results are adequate, many practicing traffic engineers find the neural network approach to be more complex than the existing factor approach due to lack of expertize.

In emerging Connected Vehicle Technology where vehicles continuously send data to roadside infrastructure through a wireless communication medium potentially could reduce the needs of permanent as well as coverage count stations for collecting traffic volume data (*11, 12*). Although it will be several more years before enough connected vehicles on different road segments reduce the need for count stations for volume data collection, they potentially are considered as potentially an economically beneficial data collection strategy (*13*).

2.3 AADT Estimation through Centrality

Land-use characteristics at the parcel level have been incorporated in several AADT estimation models (*8, 14*). Centrality models, for example, seek to apply numerical values to the topological significance of each element in a network using land-use characteristics. Centrality is used in graph theory and network analysis in order to identify the level of importance of certain elements of a graph or network. A network in this case is broadly defined, and can refer to street networks, urban networks, and even social networks (*15*).

In 2012, Zhang et al. developed a model based on road network patterns and traffic analysis zones using three types of centrality - betweenness centrality, degree centrality, and closeness centrality (*14*). The betweenness centrality of a node is the count of the times that the node is intersected by all of the "shortest paths" within the network – when considering the paths from every node in the network to every other node in the network. Degree centrality of a node is the count of the number of other nodes that are adjacent to it, and with which it is, therefore, in direct contact. Closeness centrality of a node is based upon the degree to which a point is close to all other points. The greater the closeness centrality of a node is, the closer the node is to all others within the network. Zhang et al.'s study used a centrality property of the whole network, based on the node centralities calculated, to find the each type of network centrality – betweenness, degree, and closeness. It was found that network betweenness centrality was the most accurate in distinguishing and describing various Traffic Analysis Zone (TAZ) road network patterns (*14*).

Most recent approaches for estimating AADT use a modified form of stress centrality. Stress centrality is defined as the total count of the times a link would be used if one were to travel from every node to every other node via the shortest path in a network. Lowry introduced a new metric called origin-destination (OD) centrality that can be used in linear regression as a contributing variable to calculate AADT spatially (*8*). Finding OD centrality can be executed using a geographic information system (GIS) platform and less data than other methods require, including land use data and the street network. A case study of this concept yielded an R^2 of 0.95. AADT can vary greatly among roadway segments of the same functional classification, and this method can demonstrate significant variation along roadway segments of the same functional class (*8*).

There are several benefits of using centrality concepts in estimating AADT. First, the concept is simple and can be realistically applied to any given network, without the use of expensive, proprietary, and/or subjective data. Also, to derive centrality measures, a GIS program is essentially the only required tool. Not only can this entire model can be completed within a short time period, but it has the potential to reduce the number of coverage counts needed to accurately estimate AADT.

In this research, the author was motivated to investigate if the centrality method utilized on a small city could be applied to a medium size city, and determine if the model would benefit from the inclusion of additional variables, rather than the centrality variables alone (*16*).

CHAPTER THREE

RESEARCH METHOD

Opting to use only data that is readily available to South Carolina DOT, this study attempts to find new deterministic variables to calculate AADT on roadways in a medium size city (population roughly 65,000) based on the theory of centrality. A similar study was conducted for the small city of Mascow, Idaho (population of roughly 24,000) (*7*). Our study area, the city of Greenville, SC currently has 6 automatic traffic recorders (ATRs) and 153 short-term count locations within the city limits. The main objective of this particular research was to apply origin-destination centrality to the city as a means to find new variables to estimate AADT that are statistically significant at least 95% (p-value<0.05).

To expand off of the discussion in the Literature Review, origin-destination centrality is a type of centrality that is derived from stress centrality. Stress centrality is defined as the total count of the times a link would be used if one were to travel from every node to every other node via the shortest path in a network. In terms of calculating stress centrality, the stress centrality equation of a link within a network is given below:

Eqn. 3 $$Stress\ centrality_e = \sum_{ij \in V} \sigma_{ij}(e)$$

where V = the set of all zones in a network, σ_{ij} = the shortest route from node i to node j, and $\sigma_{ij}(e) = 1$ if link e is used in this path and $\sigma_{ij}(e) = 0$ if link e is not used in the shortest path. This stress centrality would be based on three types of travel in or through a city:

Internal-to-Internal, Internal to External, and External to External. These will be explained further in Step 5, later in this chapter.

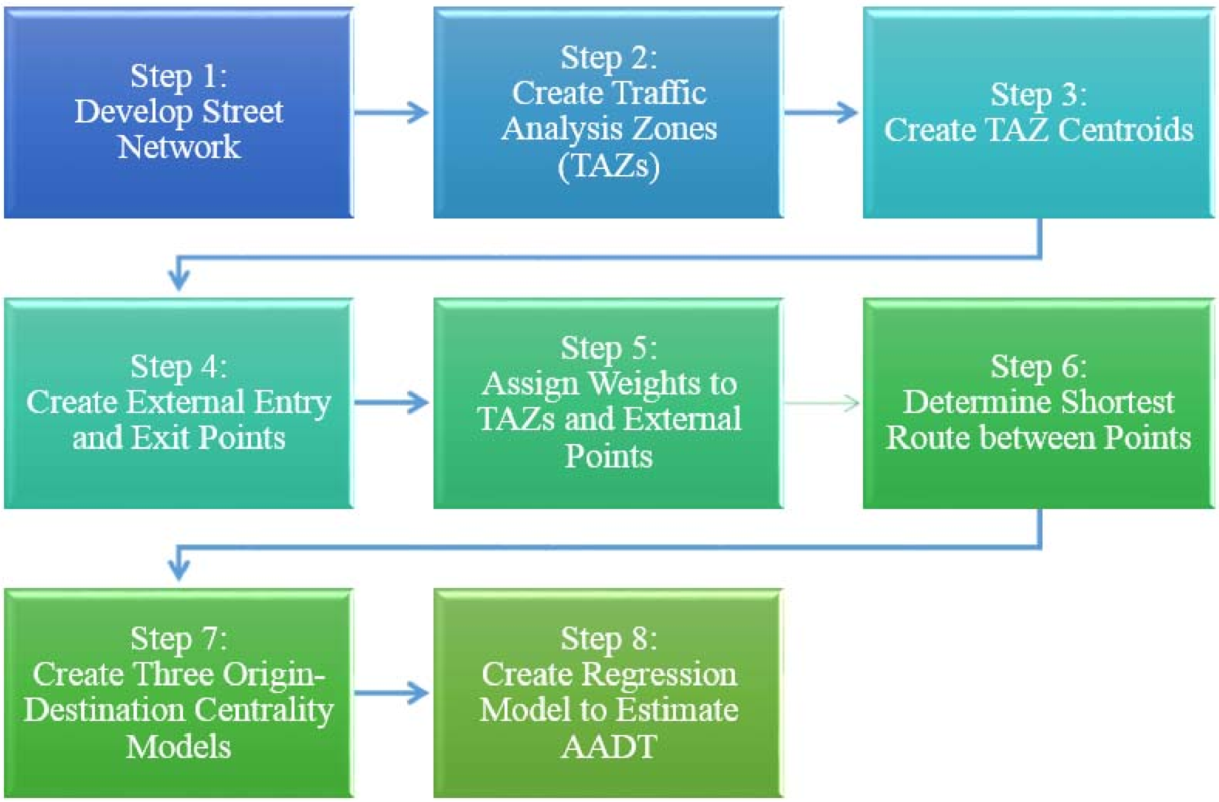

FIGURE 1 METHOD OF ORIGIN-DESTINATION CENTRALITY CALCULATION

Unlike stress centrality, the origin-destination centrality of a link uses relative weights of the TAZs and gateways in each origin-destination combination. Additionally, the theory of centrality assumes all links are of the same size, and does not take into account the capacity of each. Therefore, this O-D centrality of each link is actually significant only per lane. Therefore, we will incorporate the number of lanes into the equation, making it simply:

$$\textbf{Eqn. 4} \qquad OD\ centrality_e = N \sum_{ij \in V} \sigma_{ij}(e) W_i W_j$$

where N = is the number of lanes on link *e*, V = the set of all zones in a network, σ_{ij} = the shortest route from node i to node j, $\sigma_{ij}(e) = 1$ if link e is used in the path and $\sigma_{ij}(e) = 0$ if link e is not used in the shortest path, W_i = the relative weight of origin I, and W_j = the relative weight of destination j. A "node" in this instance will refer to either a TAZ or a gateway.

The majority of the steps (*Steps 1-5* and *Step 7*) included in this research effort were performed through a GIS software. All of the data used in the new model development are publically accessible, updated at least once annually. The base data used in the model development was collected from the City of Greenville, SC's GIS Data website (*17*) and the ITE Trip Generation Manual (*18*). This method does not require any additional data collection or cost for the purchasing of data from private organizations. The data used are geocoded in GIS shapefiles for the data sets shown in Table 1 below.

TABLE 1 Shapefiles and Attributes used in the Model Development

Shapefile	Notable Attributes
Street Centerlines	Street Name, Functional Class, Speed Limit, Number of Lanes
Parcels	Parcel Number, Number of Buildings, Building SF, Parcel Area
Zoning	Zone Number, Zoning Code
Count Station Locations	Station Number, Count Data

Step 1: Develop Street Network

First, the polyline shapefile of the street network, titled "Street Centerlines" was used to develop a street network for the City of Greenville, SC. The ESRI ArcGIS extension, "Network Analyst", has a tool called "Create Network Dataset" that can easily convert a shapefile to a network. When using ArcGIS to make routing decisions, it is essential to first convert a polyline shapefile to a network of links.

Step 2: Create Traffic Analysis Zones (TAZs)

Much like travel demand models, the stress centrality and origin-destination centrality models are composed of Traffic Analysis Zones (TAZs). The "Parcels" shapefile contains land use attributes such as land area (acre), land use type, number of buildings, etc. Each TAZ is composed of multiple adjacent parcels based on geometry, land use type, and roadway access. Within the city limits of Greenville, SC, the parcels were sorted into 1,262 zones.

Step 3: Create TAZ Centroids

The area (acre), land use type, number of buildings, building square feet, and others are all specified in the shapefile's attributes for each of the parcels. Given these attributes, combined with trip generation data from the ITE Trip Generation Manual (*18*), each parcel is assigned its own relative weight, by calculating daily trip generation rate for each TAZ, as shown in **Figure 2**. For example, if a parcel's land use was identified as a single family residence, the daily trip generation rate from the Trip Generation Manual, in this case 9.5 trips per dwelling unit, would be multiplied by the number of dwelling units within that parcel, giving that parcel a relative trip generation rate.

Centroids for each zone are established through a weighted mean analysis based on the geometry and weight of each parcel within a zone. The weighted mean is essentially the centroid of the zone taken as the mean center of all of the parcels within the TAZ, except instead of each of the parcels contributing equally to the center, some parcels contribute more than others based on relative trip generation rates.

Step 4: Create External Entry and Exit Points

Because the utilized methods of centrality use travel that can begin and/or end outside of the city limits, external points of entry and exit to the city are established by examining all potential major exit/entry roadways. Just outside of the city limits of Greenville, 32 major points of entry or exit are identified.

FIGURE 2 Eternal Gateway Points

Step 5: Determine the Shortest Route between TAZs and Gateway Points

Once all internal TAZ centroid and external points are established, the shortest routes between each origin-destination combination are determined. **Table 2** below shows the inputs required for each routing type. Using the previously created Network in ArcGIS, the shortest distance between all of the points in a given dataset can be found through the use of the "Closest Facility" tool in the Network Analyst extension. The output of the tool produces routes between each origin-destination combinations established.

TABLE 2 Three Centrality Types and Associated Inputs

Stress Centrality Method	Inputs	
	Origins	**Destinations**
Internal - Internal	Zone Centroids	Zone Centroids
Internal - External	Zone Centroids	External Points
External - External	External Points	External Points

Travel throughout a city (i.e., jurisdiction under consideration for modeling) is based on three types of origin-destination combinations:

1) Internal-to-Internal (I-I): Trips from one travel analysis zone (TAZ) to another travel analysis zone within the city;
2) Internal-to-and-from-External (I-E): Trips from one travel analysis zone to a destination outside of the city limits, or vice versa;
3) External-to-External (E-E): Trips from an origin outside the city limits to a destination outside of the city limits, that requires traveling through the city

Step 6: Assign Weights to Each TAZ and External Point

As in determining the stress centrality for each type of origin-destination combination, finding the origin-destination centrality began with creating a street network dataset, TAZs, and external exit/entry points. In this method, each zone's weight is determined by summing the weights of all of the parcels within that zone, which were established in Step 3. The relative weight of each external point (E) is then taken as the nearest AADT value available on that roadway.

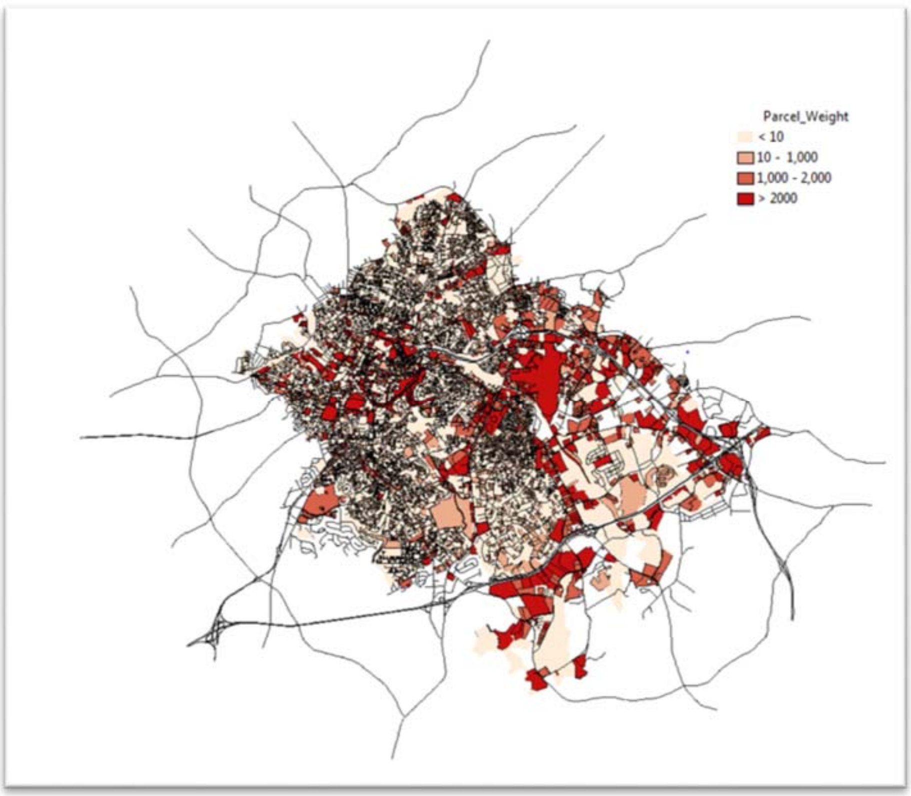

FIGURE 3 Greenville Parcels with Relative Weights

Step 7: Create Three Origin-Destination Centrality Models

The three models are created using the origin-destination travel combinations: I-I, I-E, and E-E. Their outputs incorporate the weights created in Step 5 above. The three new origin-destination centrality maps are shown in **Figure 4**, **Figure 5**, and **Figure 6.** The findings of the analysis are explained in the Chapter 4: Results and Analysis.

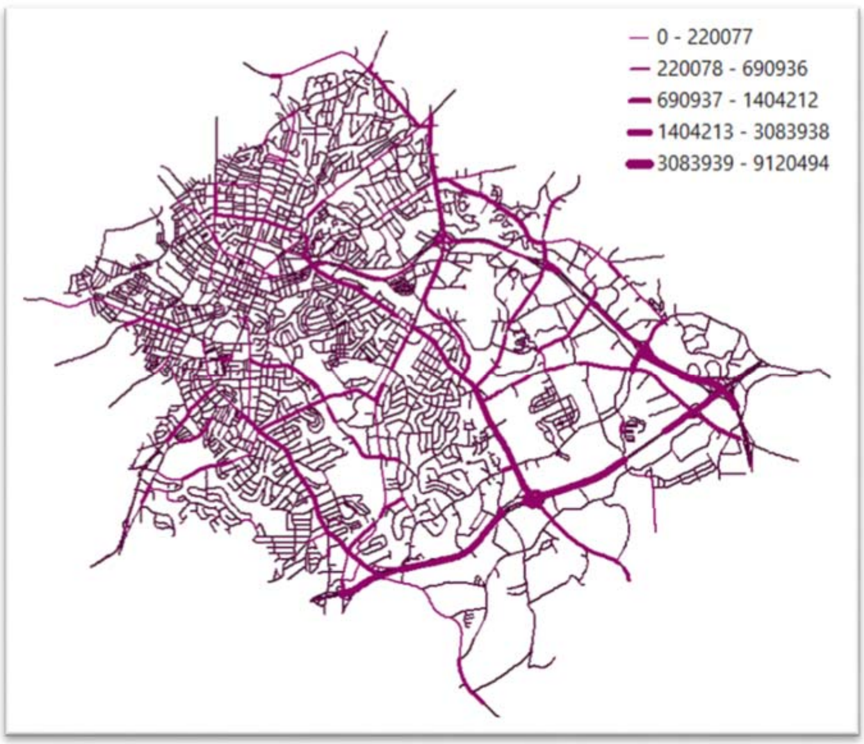

FIGURE 4 EXTERNAL-TO-EXTERNAL CENTRALITY MAP

FIGURE 5 Internal-to-External Centrality Map

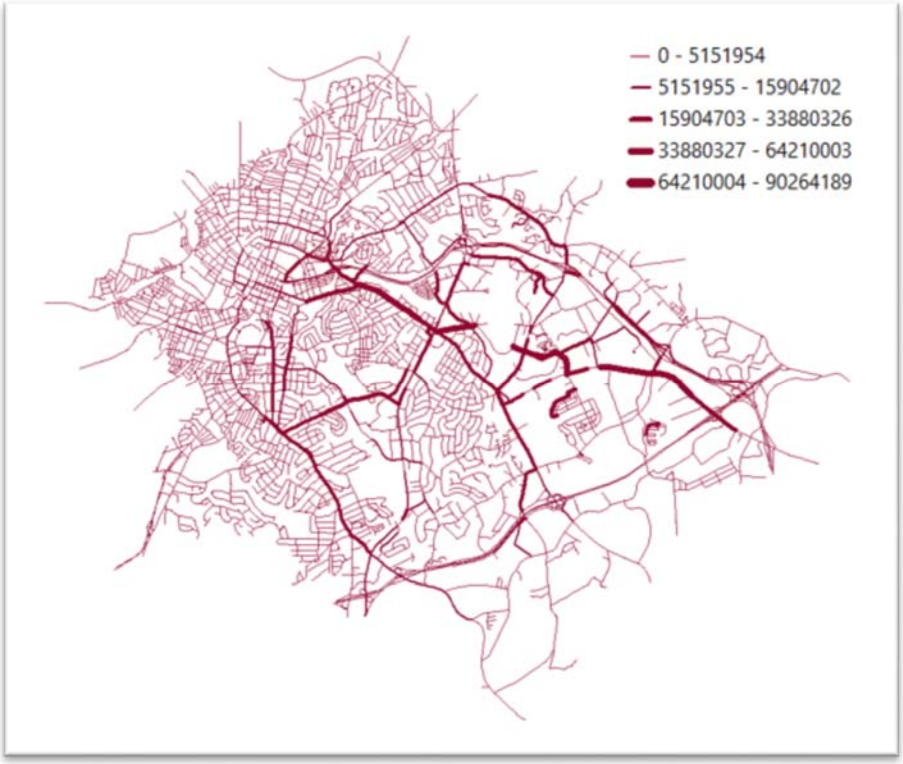

FIGURE 6 Internal-to-Internal Centrality Map

Step 8: Create Regression Model to Estimate AADT

Using the outputs of the stress centrality formula above for each link and each type of stress centrality, the three new variables were calculated as potential independent variables in AADT multiple linear regression and non-linear regression model development. Other variables were incorporated in an attempt to produce the most accurate model are speed, number of lanes, and functional class. These methods and the findings of the analysis are explained in Chapter 4: Results and Analysis.

CHAPTER FOUR

ANALYSIS AND RESULTS

4.1 AADT Estimation Using Centrality and Linear Regression

Once the three new origin-destination centrality were calculated following the steps presented in Chapter 3: Research Method, multiple linear regression analysis was performed with these three and three roadway characteristic variables, in order to develop the most accurate centrality based AADT estimation model. The six independent variables considered are:

i. I-I Origin-Destination Centrality (I-I OD)
ii. I-E Origin-Destination Centrality (I-E OD)
iii. E-E Origin-Destination Centrality (E-E OD)
iv. Functional Classification (FC)
   5 - Interstate
   4 - Major Arterial Freeway/Expressway
   3 - Major Arterial
   2 - Minor Arterial, Major Collector
   1 - Minor Collector
v. Speed Limit (SL)

The centrality variables were combined with roadway characteristic variables (i.e., speed, functional class and number of lanes) with the goal to derive new variables that could prove to be independent variables for AADT estimation model. The final regression model is shown in **Table 3** below.

TABLE 3 Regression Analysis Summary Statistics

Regression Statistics	
Multiple R	0.910628468
R Square	**0.829244206**
Adjusted R Square	0.825851707
Standard Error	4828.530729
Observations	155

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	3	17096765045	5698921682	244.4346	9.9506E-58
Residual	151	3520521060	23314709		
Total	154	20617286105			

	Coefficients	*Standard Error*	*P-value*
Intercept	-18267.96215	2496.893055	1.39986E-11
I-I OD	3.12073E-06	4.95513E-07	3.12489E-09
E-E OD	0.001284391	0.000124769	3.68774E-19
Speed	712.4749964	77.96341866	3.94468E-16

Due to these results, the formula below was selected as the best formula:

AADT = 3.12073E-06* I-I OD + 1.284391E-04 *E-E OD+ 712.4749964 SL – 18267.96215

As shown, I-E OD centrality was not used in this final model. This is because, due to the nature of the three models, I-E OD centrality was too similar to E-E centrality and I-I OD centrality to provide statistical significance in the final model. The coefficient of I-I Centrality is much lower than that of E-E Centrality, because the values used for I-I centrality were much larger than those used for E-E Centrality. The coefficients for all variables used are positive, showing a positive correlation with AADT. Additionally, the p-values for each of the variables used are all significant at greater than 99% (p-value <

0.001). Finally, the intercept is a very large negative value, which is also not surprising, due to the large values of these variables.

4.2 Comparison of AADT Estimation using Centrality based Model and Traditional Travel Demand Model Output

In order to validate the accuracy of the centrality based AADT estimation model, the author compared the model's predictive ability with that of the city's existing travel demand model. First, it is important to carefully specify the differences between the steps and data necessary to use the two models.

Travel demand forecasting models consists of four major steps: Trip generation, trip distribution, mode choice, and trip assignment (*19*). Time-of-day and directional factoring is a very important step as well, but is not explicitly mentioned in four steps of the "four-step" model. For example, in the four step model, ample amount of data inputs are necessary prior to the initiation of the four steps, which can include, but are not limited to, employees, students, automobiles, and households by TAZ. In addition, there are two ends of any trip generation – trip productions and trip attractions – and all trips must have a given purpose – Home-based Work (HBW), Home-based Other (HBO), and Non Home-based (NHB). Considering both trip productions and attractions, in addition to each purpose, there are six types of formulas that must be used for each TAZ in order to calculate the overall trip production and attraction of each TAZ. Then, once the formulas are estimated and the trip productions and attractions are calculated, the number of attractions must be balanced to the number of productions (*19*). The trips for Internal-to-External and External-to-External travel are then calculated separately. **Table 4** and **Table 5** below compare more of the inputs and steps of each of the models.

TABLE 4 Comparison of Model Inputs for Centrality Method and TDF Model

Model Inputs	Centrality Method	Travel Demand Forecasting Model
Street Network	x	x
Street Data (Name, Number of Lanes, Speed, etc.)	x	x
Employees by TAZ		x
Students by TAZ		x
Automobile by TAZ		x
Households by TAZ	x	x

TABLE 5 Comparison of Model Steps for Centrality Method and TDF Model

Model Steps	Centrality Method	Travel Demand Forecasting Model
Develop Street Network	x	x
Create Traffic Analysis Zones (TAZs)	x	x
Create TAZ Centroids	x	x
Create External Entry and Exit Points	x	x
Assign Weights to TAZs and External Points	x	x
Generate Trip Productions by Purpose and TAZ		x
Generate Trip Productions by Purpose and TAZ		x
Balance Number of Attractions to Productions		x
Generate Trips for Internal-to-External Travel		x
Generate Trips by Gateway for External-to-External Travel		x
Generate Production & Attraction Trip Tables by Purpose		x
Transpose Table to Create Origin-Destination Tables by Purpose		x
Create a QuickSum Matrix		x
Direct Traffic between all TAZ pairs using Trip Assignment	x	x
Direct Traffic between all gateway and TAZ pairs using trip assignment	x	x
Create Three Origin-Designation Centrality Models	x	
Create Regression Model to Estimate AADT	x	

The Base AADT estimates used in **Figure 7** are based on observations from 149 short-term count locations in the City of Greenville. Base AADT data was calculated using adjustment factors and short term count data collected by SCDOT. The travel demand model was created by a third party for the Greenville Metropolitan Planning Organization (MPO) for year 2010. Average growth rates from the 6 ATRs located in the City of Greenville were used to estimate the 2015 volumes from the travel demand model's calibrated volumes for 2010. The comparison of the AADT estimates for short term count stations using two models (i.e., centrality based AADT method, and AADT from travel demand model) is shown in **Figure 7**. It is important to note that both the origin-destination centrality model and the travel demand model were calibrated using the estimated AADT using the factor approach and 24-hour count for each short-term count station.

The travel demand model achieved lower goodness-of-fit (i.e., R^2 value of 0.61), while the new centrality based AADT method developed in this research has higher goodness-of-fit (i.e., R^2 value of 0.77). Additionally, the root mean square error (RMSE) for the centrality-based AADT linear regression model and the travel demand based AADT model are 7352.54 and 14073.19 respectively. It could be concluded that centrality based AADT method performs better compared to AADT estimate from travel demand model in terms of RMSE and R^2.

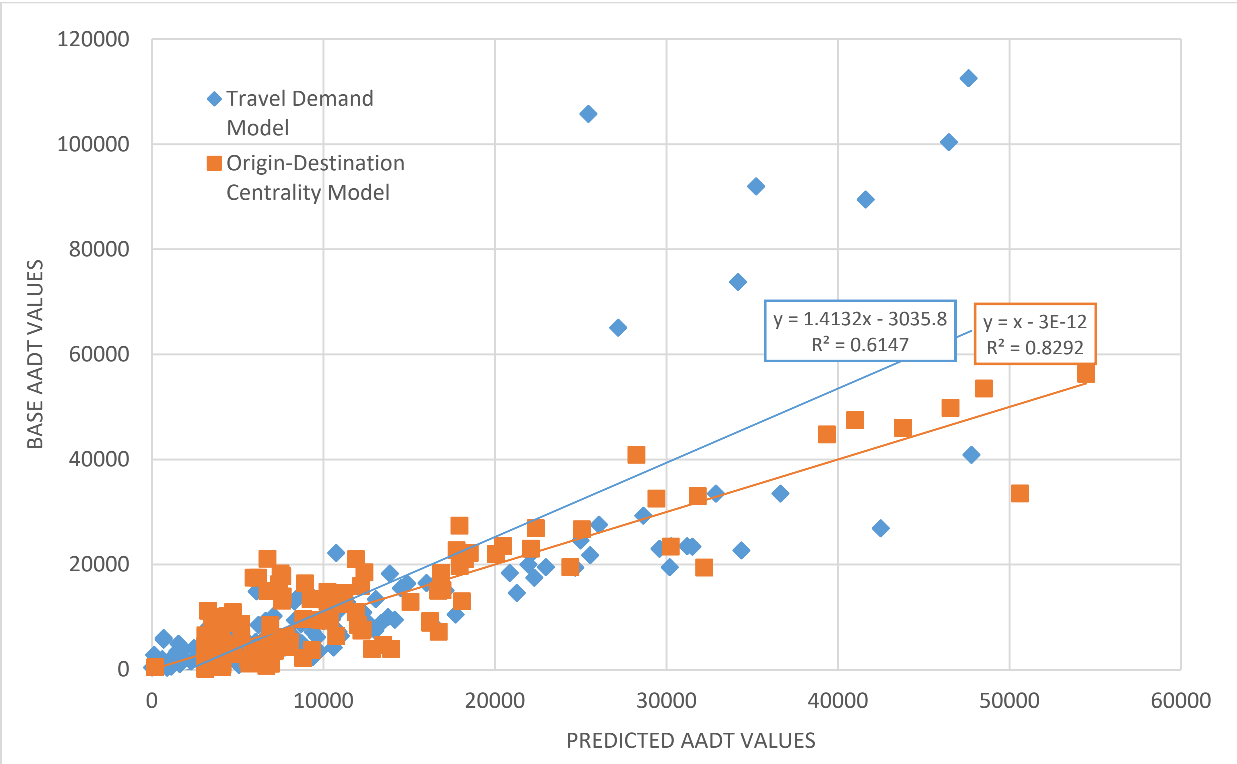

FIGURE 7 Comparison of AADT Estimation Methods

4.3 Possibility of Reduction of Coverage Counts using centrality based AADT Method

In an attempt to explore if the number of short-term count locations maintained throughout the city could be reduced applying centrality based AADT estimation method, random sub-sets of short term count were used to determine how many short-term stations were necessary to achieve same level of AADT estimation for all roads in the city. Five random sub-sets of each scenario were used, and the scenarios consisted of 100%, 80%, 60%, 40% and 20% of the total short term count stations in the city of Greenville. In each random sub-set, two thirds of the data were used for model calibration and the other third was used for model validation. This is shown in further detail in **Figure 8** below, for clarification.

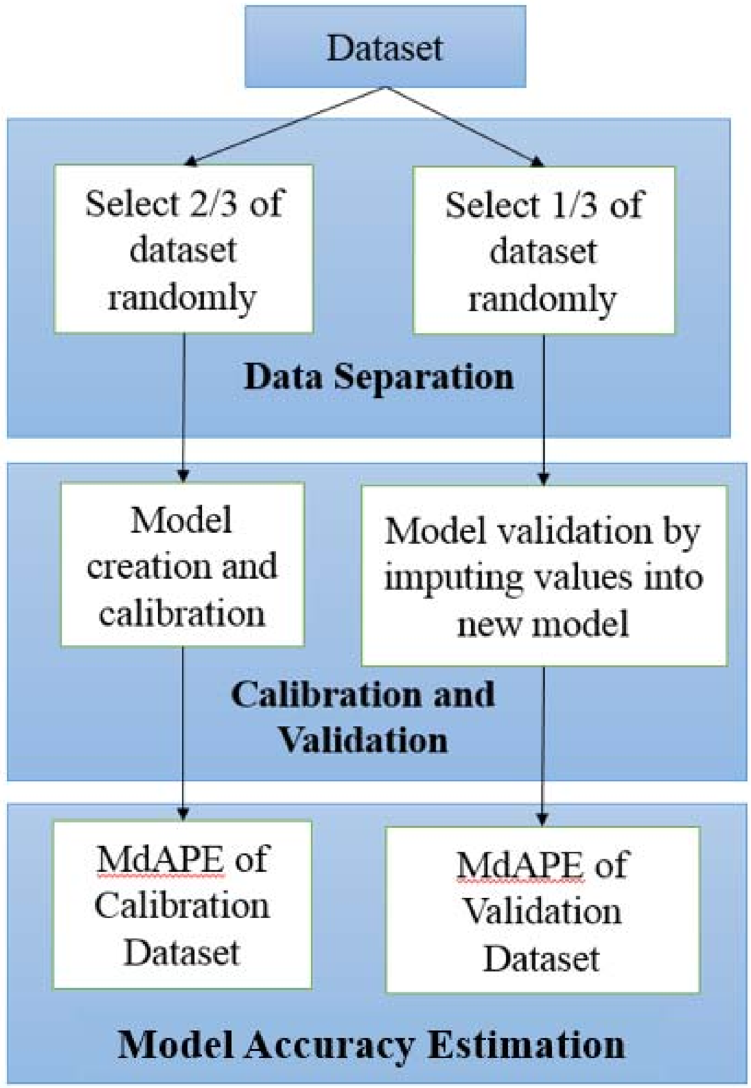

FIGURE 8 Method to Determine Coverage Count Reduction Potential

After the model of the calibration set was developed, the validation set was ran using the same linear regression model. Then the Median Absolute Percent Error (MdAPE) of the calibration and validation data sets were calculated. This process was repeated 5 times for each scenario. The results of each trial in each scenario are shown in **Table 6** below. As demonstrated in **Figure 9**, after using more than 60 short-term count stations (40% of current short-term count stations), little AADT estimation accuracy is gained in terms of MdAPE, suggesting that the number of coverage counts used in the modeling can be reduced by 60%, which could substantially reduce data collection cost for the SCDOT. The author believes this method can further save DOTs resources without compromising model accuracy. A cost savings analysis is performed later in this chapter.

TABLE 6 Trials for Reduction of Coverage Counts

Percentage of Count Stations	Trial Number	Calibration Data MdAPE	Validation Data MdAPE
	1	55.264	92.253
	2	57.391	97.678
	3	59.166	64.559
	4	59.331	66.062
100	5	64.716	62.194
	1	56.670	60.501
	2	59.967	56.790
	3	58.667	64.027
	4	64.045	47.361
80	5	58.784	70.419
	1	53.213	61.825
	2	58.629	54.766
	3	45.903	72.636
	4	63.355	53.099
60	5	65.522	64.652
	1	54.727	79.689
	2	38.788	69.292
	3	62.132	58.174
	4	59.158	51.264
40	5	60.729	111.861
	1	68.901	96.816
	2	64.788	59.709
	3	45.558	195.504
	4	38.878	90.297
20	5	67.126	57.824

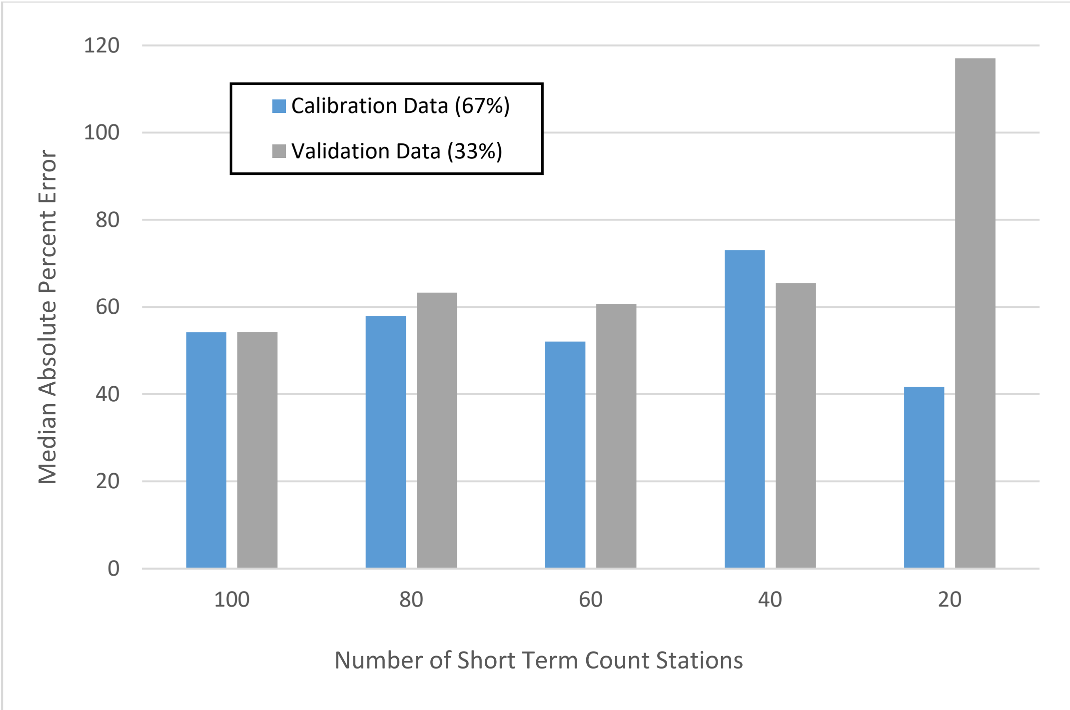

FIGURE 9 Reduction of Coverage Counts Applying the Centrality Method

4.4 Cost Savings Analysis

Because the centrality based AADT method has the potential to reduce the number of short-term count stations, a cost savings analysis was performed to estimate annual savings for the South Carolina Department of Transportation by using this method throughout the state.

SCDOT performs 12,000 short-term counts per year, each collecting 24 hour volume data. Due to the estimated reduction of over 60% short-term counts in Greenville applying centrality based AADT method cost savings analyses was performed for 40%, 50%, 60%, and 70% short-term count reduction statewide, as presented in **Table 7** below.

Each time SCDOT performs a short-term count, they use a PEEK-ADR 2000 counter and two pneumatic tubes. The counter costs an average of approximately $1000 (depending on features, etc.) and was assumed to have an average lifespan of approximately 200 uses. The tubes used have an approximate cost of $200 per pneumatic tube, with two tubes needed in every count. The average lifespan per tube is estimated at approximately 20 uses (due to tearing, breakage, and wearing out). It was assumed that every short term count requires 3 SCDOT employees for a total of 5 hours at $30/hour, which includes labor, travel, etc. The initial cost to develop the model is not included in this analysis. Additionally, according to the author's estimates, it will cost approximately $10,000 per year to pay an in-house traffic engineer to maintain the centrality model vs. maintaining the factor method model. Then, the total cost of the centrality based method at each of the four levels of count location reduction was calculated in order to calculate a percent cost savings. Therefore, the cost savings of each reduction level is simply:

Eqn. 6 Cost savings $= [\text{Cost of factor method}] - [\text{Cost of centrality method with (x\%) reduction}]$

The cost of each method is calculated using the final equation below:

Eqn. 5 $\text{Total Cost} = [\text{Cost of traffic counter}] + [\text{Cost of tubing}] + (\text{Cost of manpower}) + \text{Cost to run and update the model}$

Using this equation, the cost of the factor method was determined to be $5,700,000 annually.

TABLE 7 Cost Savings Analysis

Percent Reduction of Counts	Counts Needed	Total Cost	Total Savings
40%	7,200	$\left[\left(\frac{7200}{200}\right) * \$1,000\right] + \left[\left(\frac{7200 * 2}{20}\right) * \$200\right] + (15 * \$30 * 7200) + \$10,000$	$3,430,000
50%	6,000	$\left[\left(\frac{6000}{200}\right) * \$1,000\right] + \left[\left(\frac{6000 * 2}{20}\right) * \$200\right] + (15 * \$30 * 6000) + \$10,000$	$2,860,000
60%	4,800	$\left[\left(\frac{4800}{200}\right) * \$1,000\right] + \left[\left(\frac{4800 * 2}{20}\right) * \$200\right] + (15 * \$30 * 4800) + \$10,000$	$2,290,000
70%	3,600	$\left[\left(\frac{3600}{200}\right) * \$1,000\right] + \left[\left(\frac{3600 * 2}{20}\right) * \$200\right] + (15 * \$30 * 3600) + \$10,000$	$1,720,000

The costs illustrated in **Table 7 (Column 4)** are the costs of performing the number of count locations specified at each level of reduction. It is estimated that 12,000 counts are reduced by 4,800, 6,000, 7,200, and 8,400 counts per year for 40%, 50%, 60%, and 70% short-term count reduction scenarios, respectively. Using the Equation 5 shown, the final savings range from $1,720,000 to $3,430,000. Clearly, there is a financial benefit of utilizing this method for SCDOT. The monetary benefits could save DOTs hundreds of thousands of dollars, which could be allocated to other valuable projects, such as roadway and infrastructure maintenance.

CHAPTER FIVE

CONCLUSION AND RECOMMENDATIONS

5.1 Conclusions

Affordably estimating AADT on all of the roadways within a given street network has been a challenge that many transportation agencies face, especially in smaller cities and municipalities where resources are limited. As AADT estimation on local roads is vital, ample amounts of research have been conducted to develop models that can estimate the AADT of every roadway within any jurisdiction. For example, linear regression, travel demand modeling, and Artificial Neural Network models have all been used in an attempt to estimated AADT. The new centrality-based AADT estimation model, utilizing roadway characteristics and new derived origin-destination centrality variables based on the theory of centrality, illustrates a method that is capable of minimizing cost and effort when compared to deploying dozens of short term count locations or using models such as the travel demand model. Not only was a new variable to estimate AADT created, but the model has a lower RMSE than the city's travel demand model, and can be developed using a GIS software alone. This thesis establishes a means of estimating AADT on every roadway within a given jurisdiction, while reducing the number of coverage counts in Greenville, SC by over 60%.

This method has the potential to be used in cities that do not have the time, funding or resources to use coverage counts at many roads throughout their jurisdictions. In addition, this method uses existing and publically available data to quickly and accurately estimate AADT within a reasonable estimation threshold.

While this method was successfully applied to Greenville, SC, it has not yet been tested for transferability among other cities within and outside SC. Future research should include a study of transferability. Another limitation of this study is that the “base AADT” counts that were used and compared to this regression model were not perfectly accurate AADT values. At 142 of the 148 locations utilized were calculated counts, the AADT values compared with the ones calculated using this model were also calculated counts, and may not be entirely accurate. The base data was calculated using 24-hour short-term traffic counts and the factor method.

5.2 Recommendations

The following recommendations are made based on this research:

- The method presented in this study should be examined for other cities before a decision can be made in its broader adoption. This should include cities of various sizes to determine its applicability in various size cities.
- This method should be tested at the state level. In addition to its use at the state level, this method can be used at local level as well for low-cost AADT estimation strategy.
- This method should be tested against AADT estimations using short-term counts and the factor method at the lowest functional classifications. Currently, there is no short-term count data available for local roads in Greenville, SC.
- This method should be tested in an area with a much larger number of permanent count stations for validation. While AADT estimations using short-

term counts and the factor method are generally very accurate, they are not perfect data, and this model should be tested against actual ground truth data.

- Cost savings estimates could be conducted for a specific area based on regional costs, which can vary based on location-based manpower and equipment costs, in order to determine an exact savings value for their particular agency prior to making a decision.

REFERENCES

1. Roess, R. P., E. S. Prassas, and W. R. McShane. *Traffic Engineering*. Upper Saddle River, NJ, 2011.
2. Zhao, F., and S. Chung. Contributing factors of annual average daily traffic in a Florida county: exploration with geographic information system and regression models. Transp. Res. Rec. 1769, 2001, pp. 113–122.
3. Traffic Forecasting & Analysis. Minnesota Department of Transportation. http://www.dot.state.mn.us/traffic/data/coll-methods.html. Accessed Aug 1, 2016.
4. *Guide to LTPP Traffic Data Collection and Processing*. McLean, VA: Federal Highway Administration, Office of Infrastructure Research, Development and Technology, Turner-Fairbank Highway Research Center, 2001. http://www.fhwa.dot.gov/publications/research/infrastructure/pavements/ltpp/trfcol/trfcol.pdf Accessed Jul 2016.
5. Doustmohammadi, M., and M. Anderson. Developing Direct Demand AADT Forecasting Models for Small and Medium Sized Urban Communities. In *International Journal of Traffic and Transportation Engineering 5.2*, 2016, pp. 27-31.
6. Wang, T., G. Albert, and P. Alluri. Estimating annual average daily traffic for local roads for highway safety analysis. In *Transportation Research Record: Journal of the Transportation Research Board, No. 2398*, 2013, pp 60-66.
7. Katter, Greg. "INDOT Traffic Data Sources and Using Traffic Count Data System Provided by MS2." (2015).

8. Lowry, Michael. Spatial interpolation of traffic counts based on origin–destination centrality. In *Journal of Transport Geography 36*, 2014, pp. 98-105.

9. United States Census Bureau. *When is the census of governments conducted?* https://www.census.gov/history/www/faqs/governments_faqs/when_is_the_census_of_governments_conducted.html. Accessed on Jul 2016.

10. Sharma, S., P. Lingras, F. Xu, and G. Liu. Neural networks as alternative to traditional factor approach of annual average daily traffic estimation from traffic counts. In *Transportation Research Record: Journal of the Transportation Research Board, No. 1660*, Transportation Research Board of the National Academies, Washington, D.C., 1999, pp. 24-31.

11. Khan, S. M., Dey, K., and Chowdhury, M., "Real-time Traffic State Estimation with Connected Vehicles," *IEEE Transactions on Intelligent Transportation Systems,* doi: 10.1109/TITS.2017.2658664, (2017).

12. Ma, Y., Chowdhury, M., Sadek, A., and Jeihani, M., "Real-Time Highway Traffic Condition Assessment Framework Using Vehicle-Infrastructure Integration (VII) with Artificial Intelligence (AI)," *IEEE Transactions on Intelligent Transportation Systems,* Vol. 10, No. 4, pp 615-627, (2009).

13. He, Y., Chowdhury, M., Ma, Y., and Pisu, P., "Merging Mobility and Energy Vision with Hybrid Electric Vehicles and Vehicle Infrastructure Integration," *Energy Policy, Modeling Transport (Energy) Demand and Policies*, Vol. 41, pp 599-609, (2012).

14. Zhang, Yuanyuan, X. Wang, P. Zeng and X. Chen. Centrality characteristics of road network patterns of traffic analysis zones. In *Transportation Research*
15. Freeman, Linton C. "Centrality in social networks conceptual clarification." *Social networks* 1.3 (1978): 215-239.
16. Keehan, M., Chowdhury, M., and Dey, K., "AADT Estimation with Regression Using Centrality and Roadway Characteristic Variables," Accepted for presentation in the 96th Annual Meeting of the Transportation Research Board, Washington, D.C., (2017).
17. City of Greenville, South Carolina. *GIS Data*. http://www.greenvillesc.gov/364/Access-GIS-Data. Accessed Jul 2016.
18. ITE (Institute of Transportation Engineers), 2012. Trip Generation Manual, 9th Edition, ITE, Washington, DC.
19. U.S. Department of Transportation, Federal Highway Administration, Introduction to Urban Travel Demand Forecasting, Instructor Guide, NHI Course No. 152054, Publication No. FHWA-NHI-08-061, May 2008.

APPENDIX A

SAMPLE DATA

Streets _Shapefile

FID	OBJECTID	RECNUM	ST_LABEL	ROADNUM
0	1	3	RUTHERFORD RD	S-23-21
1	2	4	RAYFORD LA	
2	3	5	N PLEASANTBURG DR	SC-291
3	4	6	TILBURY WY	
4	5	7	WEDGEWOOD DR	
5	6	8	BROUGHTON DR	
6	7	9	INGLEWOOD DR	
7	8	10	TAMBURLAINE CT	
8	9	11	WEDGEWOOD DR	
9	10	12	MEADOW CREST CIR	
10	11	13	WEDGEWOOD DR	
11	12	15	SUMMIT DR	
12	13	16	SUMMIT DR	
13	14	17	DEARSLEY CT	
14	15	18	BRENTWOOD DR	
15	16	19	RUTHERFORD RD	S-23-21
16	17	20	GREEN MEADOW LA	
17	18	21	BRENTWOOD DR	
18	19	22	BRENTWOOD DR	
19	20	23	INGLEWOOD DR	
20	21	24	SUMMIT DR	
21	22	25	TILBURY WY	
22	23	26	WEDGEWOOD DR	
23	24	27	BROUGHTON DR	
24	25	28	STONE LAKE DR	
25	26	29	N PLEASANTBURG DR	SC-291
26	27	30	SUMMIT DR	
27	28	31	GOBLET CT	
28	29	32	VENNING CT	
29	30	33	RUTHERFORD RD	S-23-21
30	31	34	COOL SPRINGS DR	

LADD1	LADD2	RADD1	RADD2	PREFIX
1241	1299	1240	1298	
1	99	2	98	
2001	2099	2000	2098	N
1	11	2	12	
401	499	400	498	
201	299	200	298	
13	99	12	98	
1	99	2	98	
305	399	304	398	
1	99	2	98	
301	303	300	302	
1001	1009	1000	1008	
1011	1041	1010	1040	
1	99	2	98	
31	41	30	40	
1221	1239	1220	1238	
1	99	2	98	
0	0	0	0	
13	29	12	28	
1	11	2	10	
927	999	926	998	
51	99	50	98	
201	299	200	298	
101	199	100	198	
1	99	2	98	
1901	1999	1900	1998	N
1043	1051	1042	1050	
1	99	2	98	
1	99	2	98	
1201	1219	1202	1218	
1	99	2	98	

NAME	TYPE	SUFFIX	ALTNAME
RUTHERFORD	RD		
RAYFORD	LA		
PLEASANTBURG	DR		
TILBURY	WY		
WEDGEWOOD	DR		
BROUGHTON	DR		
INGLEWOOD	DR		
TAMBURLAINE	CT		
WEDGEWOOD	DR		
MEADOW CREST	CIR		
WEDGEWOOD	DR		
SUMMIT	DR		
SUMMIT	DR		
DEARSLEY	CT		
BRENTWOOD	DR		
RUTHERFORD	RD		
GREEN MEADOW	LA		
BRENTWOOD	DR		
BRENTWOOD	DR		
INGLEWOOD	DR		
SUMMIT	DR		
TILBURY	WY		
WEDGEWOOD	DR		
BROUGHTON	DR		
STONE LAKE	DR		
PLEASANTBURG	DR		
SUMMIT	DR		
GOBLET	CT		
VENNING	CT		
RUTHERFORD	RD		
COOL SPRINGS	DR		

RDCLASS	MAINTENANC	SPEED	MAJRDS	LOW_STREET
3	2	40	Y	1240
1	1	30	N	1
2	2	40	Y	2000
1	1	30	N	1
1	1	30	N	400
1	1	30	N	200
1	1	30	N	12
1	1	30	N	1
1	1	25	N	304
1	1	30	N	1
1	1	30	N	300
1	1	35	N	1000
1	1	35	N	1010
1	1	30	N	1
1	1	30	N	30
3	2	40	Y	1220
1	1	30	N	1
1	1	30	N	0
1	1	30	N	12
1	1	30	N	1
1	1	35	N	926
1	1	30	N	50
1	1	30	N	200
1	1	25	N	100
1	1	30	N	1
2	2	40	Y	1900
1	1	35	N	1042
1	1	30	N	1
1	1	30	N	1
3	2	40	Y	1201
1	1	30	N	1

HIGH_STREE	STREET_ID	SECT_NO	EDITORNAME	LASTUPDATE
1299	5600	2000000002		
99	2644	0000167600		
2099	2574	2000000004		
12	5251	0000167700		
499	3709	0000165800		
299	446	0000164400		
99	1629	0000164700		
99	5252	0000168000		
399	3709	0000165700		
99	2117	0000167500		
303	3709	0000165600		
1009	3105	0000163500		
1041	3105	0000163400		
99	5563	0000164200		
41	390	0000164800		
1239	5600	2000000018		
99	1349	0000168100		
0	390	0000165100		
29	390	0000164900		
11	1629	0000164600		
999	3105	0000163600		
99	5251	0000167900		
299	3709	0000165500		
199	446	0000164500		
99	3081	0000168200		
1999	2574	2000000028		
1051	3105	0000163300		
99	5401	0000164100		
99	5288	0000164300		
1219	5600	2000000032		
99	775	0000167400		

CSOURCE	FACILITYID	Shape_STLe
LIB	8954	527.1766826
LIB	8955	1150.473647
LIB	8956	614.3874936
LIB	8957	415.4401001
LIB	8958	186.6661026
LIB	8959	735.1952526
LIB	8960	308.0473068
LIB	8961	235.2168267
LIB	8962	263.642934
LIB	8963	645.0751654
LIB	8964	165.4496045
LIB	8965	202.4614411
LIB	8966	697.1660385
LIB	8967	140.9242893
LIB	8968	297.5336242
LIB	8969	752.6313822
LIB	8970	404.4736069
LIB	8971	396.117093
LIB	8972	68.30224875
LIB	8973	412.1011087
LIB	8974	287.5025614
LIB	8975	387.9861478
LIB	8976	548.1792152
LIB	8977	1409.610659
LIB	8978	295.3921772
LIB	8979	883.2929854
LIB	8980	317.2860422
LIB	8981	118.7954023
LIB	8982	416.1107824
LIB	8983	400.3369536
LIB	8984	586.6512895