METHODS TO ESTIMATE LINK LEVEL TRAVEL BASED ON SPATIAL EFFECTS

by

Prasanna Reddy Kusam

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

Charlotte

2011

Approved by:

____
Dr. Srinivas S. Pulugurtha

____
Dr. Martin R. Kane

____
Dr. Rajaram Janardhanam

____
Dr. Jean-Claude Thill

____
Dr. Mark E. Faust

ABSTRACT

PRASANNA REDDY KUSAM. Methods to estimate link level travel based on spatial effects. (Under the direction of DR. SRINIVAS S. PULUGURTHA)

Annual Average Daily Traffic (AADT) is used in several planning, roadway design, operational and safety analyses by transportation planners and engineers. Existing methods are very complex and do not adequately address the modeling needs. Errors and inaccuracies in a traditional four-step method get carried to later steps often resulting in incorrect estimates of travel demand. The primary focus of this research is to develop a systematic and simplified methodology to estimate link level travel on roadways. The proposed methodology involves scientific principles and statistical techniques, but bypasses the tedious four-step method. Two spatial methods, first one based on "spatial proximity" and second one based on "spatial weighting", are proposed to estimate link level travel. While the former method investigates to identify ideal "proximal" distance to capture spatial data, the later method involves application of "spatial weights" that decrease with an increase in distance to integrate spatial data from multiple buffer bandwidths. Generalized Estimating Equations (GEE) models are developed for both the methods using Poisson and Negative Binomial distributions with and without network characteristics to facilitate transportation planning and analysis. Validation of the developed models is carried out using Chi-Square Statistic test. The goodness of fit statistics indicates that Negative Binomial models performed better than Poisson models. Models with network characteristics performed better than models without network characteristics. Model validation results indicate that link level travel can be accurately estimated using both the spatial methods.

DEDICATION

to my
Mom Venkata Ratnam
and
Dad Balasankar Reddy

ACKNOWLEDGEMENTS

I express my sincere and heartfelt thanks to my advisor Dr. Srinivas S. Pulugurtha, for his wise and scholarly guidance, constant support, and constructive criticism and encouragement. Words fall short to express gratitude for such an inspiring mentor. I specially thank Dr. Rajaram Janardhanam for his elderly and wise counsel and support. I extend my thanks to Dr. Martin Kane for his support throughout the program, Dr. Jean-Claude Thill for his insightful guidance on the research and Dr. Mark E. Faust for his initial help with the statistical models. I thank the above cited committee members for their helpful comments, guidance and for serving on the examination committee. I would like to acknowledge Anna Gallup and Joe McLelland and several others at the Charlotte Department of Transportation for providing me with all the necessary data needed for the research. I am always indebted to the Department of Civil & Environmental Engineering and UNC Charlotte for supporting my graduate studies and providing the needed resources. My special thanks to Raghu Ariga and Vidya Sagar Ealla for teaching me programming skills needed for the research. I would like to thank the Transportation group members, and staff in the department for making it a wonderful experience. I extend my thanks to all my friends, too many to list, for making my graduate school life so memorable; I would cherish all the moments throughout my life.

Finally, I sincerely appreciate my sister Gayathri and brother-in-law Narasimha Reddy for their love and support, our family friend Kotamraju S.V.V.V.S. Sharma and my uncle Vuyyuru Basivi Reddy for inspiring me and convincing my family to send me to the United States for higher education.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER I: INTRODUCTION

Growth in population along with increased human activity and settlement in and around cities has completely changed the urban environment and its landscape over the past few decades. This had a direct impact on travel demand, traffic congestion, and associated air quality problems. Economic losses due to traffic congestion are noteworthy during the last two decades. A total delay of 36 billion hours and an associated combined delay and fuel waste cost of $959 billion (adjusted to year 2005 dollars) occurred during this period. Congestion cost due to delay and excessive fuel wastage in 2005 itself was $78 billion according to “The 2007 Urban Mobility Report”. Travelers burnt 2.9 billion gallons and wasted 4.2 billion hours on roads in 2005 due to congestion. The cost of wasted fuel and delay per traveler was $710 in 2005 compared to an adjusted $260 in 1982 (Schrank and Lomax, 2007). Also, the present transportation infrastructure capacities are becoming inadequate to cater to the current needs of the users. This is expected to worsen in the future.

Congestion on road network occurs due to an imbalance between supply and demand (i.e., when demand exceeds capacity). Understanding and identifying the factors causing travel and thus congestion helps to accurately estimate and forecast travel and its growth in the future. Accurate estimation of travel demand facilitates planners to plan, propose and prioritize infrastructure projects for capacity improvements. It helps to evaluate and select strategies to provide remedial measures to the new and existing infrastructure and mitigate congestion related problems. Strategies to reduce congestion include capacity expansions, congestion management using advanced or intelligent transportation systems, public transportation, reduced or staggered work hours, mixed and dense land use development, and improving non-motorized transportation facilities to name a few.

1.1 Background

Traffic volume data for interstates, major and minor thoroughfares (or arterials) and local roads are necessary for transportation planning, roadway design, congestion management, scenario and trend analysis, before and after studies of newly added transportation infrastructure, and in several other transportation related decision making processes (Zhao and Park, 2004; Goel et al, 2005). Traffic volume for a particular highway is defined as the number of vehicles passing a specific point on the highway in a given lane or direction during a specified time interval (Roess, Prassas and Mc Shane, 2004, page no. 106). The unit of time considered is “per day” for planning and general analysis purposes whereas “per hour” is used for detailed investigations such as traffic control and operations. Commonly known daily volume measurements are Average Annual Daily Traffic (AADT), Average Annual Weekday Traffic (AAWT), Average Daily Traffic (ADT) and Average Weekday Traffic (AWT). Of these, AADT is of prime importance for many planning and analysis purposes.

Use of AADT can be broadly divided into four transportation engineering areas, namely, planning, design, traffic operations and safety. A summary of uses of AADT in each of these areas is discussed below.

Planning and Design:

- as an indicator of roadway usage and congestion levels;
- as an important variable in geometric design;
- for planning purposes such as determining capacity (number of lanes), type of road and additional links to carry the estimated AADT smoothly and effectively and to validate planning models;
- to estimate vehicle miles traveled, and traffic flow between origins and destinations;

Operational and Safety Analysis:

- to select operational methods, intersection features and signal controls, and to conduct operational analysis for finding vehicle delays, density and level of service (LOS);
- to compute crash rates and identify high crash locations for safety analysis. For example Section 1401 of SAFETY-LU requires all the states to report top 5% of locations requiring the most severe highway safety needs. AADT estimates help analysts to evaluate the safety needs of high crash locations based on the demand;

Other Uses:

- to predict, evaluate and prioritize transportation infrastructure needs;
- to assist in air quality conformity analysis and decision making process;

1.2 Motivation

Travel demand modeling is part of Urban Transportation Planning Process (UTPP) employed by many metropolitan organizations to estimate traffic (AADT) on urban roads. Since the 1950s, the traditional sequential Four Step Modeling (FSM) process has dominated transportation planning in developed countries. FSM typically is an aggregate zonal level analysis that restricts consideration of spatial variations of factors influencing travel demand in the estimation process. Due to its nature of trip based serial step by step approach, it cannot address all the aspects of demand jointly and in the presence of congestion. The errors, assumptions and relaxations are carried over throughout the process. It lacks behavioral attention, sensitivity to policy changes, and temporal variations. Also, travelers might not follow the same sequence of steps defined in the FSM.

To address problems inherent in the FSM process and to avoid its limitations, researchers developed combined travel demand models, disaggregate behavioral/activity based models, and made estimations using traffic counts. Each model/method has its own limitations and drawbacks. Combined travel demand models are developed to account for error transfer between the steps with success in small scale applications only. Also, an aggregate model by itself, spatial variations is ignored in this modeling process. Disaggregate models require a lot of data and are typically suitable for small scale applications. Though it became computationally easier with the advancement in processing and computing technology, data collection is a very difficult problem for activity based models. Tracking each and every individual's activity (a sample is collected in general to make interpretations) and incorporating them in the modeling process for an entire metropolitan area is not an easy task. Therefore, there is reluctance among practitioners to shift entirely from FSM to these other methods.

Estimation using short term traffic counts is another easy and useful method but it does not give accurate results. AADT is estimated from short term and long term traffic counts using various seasonal and weekly factors on roadway links. But there is a large difference between the available traffic counts to the number of links and roads in a huge transportation network. It is difficult, practically impossible, and economically not feasible to conduct traffic counts on all the roadways. At the same time, estimating traffic from a very few number of counts available leads to uncertainties and inaccuracies.

To summarize, the estimation process in previous travel demand models are complex, data hungry, and time consuming. The transfer of errors is unavoidable. These models do not account for spatial variations of the characteristics influencing travel demand. Researchers in the past did not attempt to estimate link level traffic directly using the spatial characteristics of travel demand. Those that attempted did not consider the effect of spatial proximity or integrating data using spatial weights that decrease with an increase in distance. There is a need to develop, a simple and systematic methodology, to estimate link level travel bypassing the tedious FSM process. Such a methodology should use scientific principles, spatial analytical methods, and statistical techniques for accurate estimation of traffic. It should be easy to adopt at any scale, size and level.

1.3 Problem Statement

Due to rapid urbanization and sprawl, supported by huge transportation infrastructure investments and policies, urban settlements became segregated. As a result, urban citizens have to travel to work, shop, and perform various daily activities typically from their places of residence. Various modes of travel such as walking, bicycle, transit and vehicle are used for travel. While walking and biking trips in general are less than "one-half" or "one" mile; trips more than "one" mile are carried out using vehicular mode. Land development that typically takes place along the transportation corridors allows various roadways to attract traffic for certain distance (say "five" miles) depending on the level of urbanization and road network system. Various road functional classes are used in their journeys. Higher functional classes typically attract more traffic supplied by lower functional classes, in a hierarchical order in general.

Each functional class of roadways collects traffic at varying distances. Lower level functional classes such as driveways and local roads collect traffic from local land developments. These lower functional classes feed major streets and arterials that serve comparatively larger areas to move people and traffic. Major roads and arterials generally feed freeways/expressways that serve an entire region. The characteristics as well as level-of-service provided by these functional classes vary. User's expectation of level-of-service offered by the functional classes also varies. Further, higher functional classes are also given more priority in allocation of funds for improvements by transportation agencies. Therefore, there is a need to capture spatial characteristics that influence travel demand such as, demographic, socio-economic and land use data at various distances and develop models for various functional classes. Assessing models developed by capturing data at various "proximal" (accessible) distances helps to select the best model to accurately estimate traffic by road functional class.

It is well known in the literature that the effect (travel) decreases with increase in distance (one type of accessibility measure). Combining data captured at various accessible distances by applying gradually decreasing weights would improve the accuracy of the models. Therefore, spatial variations of demographic, socio-economic and land use characteristics based on accessible distances and spatial weights that decrease with an increase in distance (analogous to distance decay) should be incorporated in the estimation process.

1.4 Research Goal and Objectives

The primary goal of this research is to develop methods and assess models to accurately estimate link level travel or AADT for transportation planning, operational and safety analysis. Several objectives were identified to achieve this goal. They are:

- Identify on-network characteristics (such as number of lanes, and speed limit) and off-network characteristics (such as area type, demographic, socio-economic, employment, and land use) that explain travel demand with minimal multicollinearity.
- Develop models to estimate travel demand for the following road functional classes as a function of on-network and off-network characteristics.
  - (i.) Freeway/Expressway
  - (ii.) Major thoroughfares
  - (iii.) Minor thoroughfares
- Examine the effect of “spatial proximity” to extract and capture off-network characteristics and identify the most appropriate network-accessibility to develop models.
- Develop and assess whether models developed by integrating data from different buffer bandwidths based on “spatial weights” that decrease with an increase in distance would yield better results than models based on “spatial proximity”.
- Examine the effect of spatial dependency of network characteristics by considering the role of upstream and downstream network links in estimating traffic on a link.

- Compare outcomes with field counts and demonstrate the applicability of models to estimate traffic.

1.5 Organization

The proposed methodology in this research deals with the estimation of traffic on roadway links for various road functional classes. Various on-network and off-network characteristics data were considered as independent variables in the development of models. On-network characteristics of the study links, as well as upstream, downstream, and cross-streets network links, were considered to account for spatial dependency. In addition to the socio-economic and demographic characteristics, land use characteristics were also considered to capture the travel/traffic demand accurately.

Spatial buffer widths of varying sizes are considered to capture the off-network data and develop models. Two spatial methods are used to develop models in this research, namely, “spatial proximity” and “spatial weighting”. Incorporating network characteristics in the estimation process helps to directly estimate traffic on roadways for analysis purposes. On the other hand, models developed without network characteristics, facilitates estimation of travel demand for planning purposes. An assessment of models based on several scenarios is carried out to select the best model to estimate traffic. While comparison of models to estimate travel using various methods helps to identify the best model to accurately estimate traffic, model validation using actual traffic counts also helps to compare predictive capability among the models. The estimation of traffic on roadway links usually accomplished in the final step of the FSM process is achieved directly using the proposed methodology. The methodology is easy to adopt and can be applied universally to urban settings of any size and level.

The remainder of the dissertation consists of seven chapters. Chapter II discusses various travel demand estimation/modeling techniques such as sketch planning, O-D estimation, the traditional four step method, combined travel demand models and their merits and demerits. Various models to estimate AADT using short-term traffic counts and their limitations are also discussed. Further, various spatial principles and methods like accessibility, gravity principles and geographic weighted regression are introduced. A brief discussion of statistical analysis methods and distributions is also presented.

Chapter III describes the proposed methodology. Data processing involving data integration and measuring accessibility are explained. The spatial density concept to integrate data from different buffer bandwidths based on spatial weights that decrease with an increase in distance and its application in a GIS environment is illustrated. Statistical analysis procedure adopted is discussed leading to the data compilation, and results based on various spatial methods in the following chapters.

Chapter IV describes the selection of study area, data collection, design of sample size, and various characteristics of the data considered for the analysis. Chapter V and VI presents results for various models developed to estimate AADT based on "spatial proximity" method and the "spatial weighting" method, respectively. Poisson and negative binomial regression models results were presented for all road functional classes, freeways/expressways, major thoroughfares, and minor thoroughfares with and without network characteristics. An assessment is carried out based on the model results. Chapter VII provides model validation results. Finally, a summary of findings, conclusions, and potential for future research are presented in chapter VIII.

CHAPTER II: LITERATURE REVIEW

2.1 Travel Demand Modeling

To estimate the demand for travel and understand the travel patterns, Metropolitan Planning Organizations (MPOs) conduct traffic and planning studies. Travel demand forecasting is a part of the planning process which predicts the traffic volumes, flows on links of a road network and transit patronage in the future. The process considers socio-economic, demographic, land use and any new developments or transportation infrastructure projects in a region. The basic objective of this process is to provide comprehensive and continuing guidance for the development, evaluation and implementation of future transportation planning proposals, policies and in prioritizing projects to allocate available funds for future investments. Various policy evaluations are carried out in a planning process that would impact travel or transportation with new projects under consideration. Various methods and models used in practice in the transportation planning process are discussed next.

2.2 Simple Transportation Demand Models

2.2.1 Sketch Planning

Sketch planning is a simple to use, less data intensive model with outcomes sufficient to prioritize and test policy options, transit pricing policies and infrastructure investment programs primarily in the very initial stages of the planning process (Ortuzar and Willumsen, 2001). Zahavi's (1979) Unified Mechanism of Travel (UMOT) model based on consumer behavior and utility maximization of available opportunities, represented by total daily travel distance under travel cost (money) budget constraints, is a typical example of this kind.

2.2.2 Estimation using Traffic Counts

Traffic counts can be used to estimate origin destination matrix. Unlike the traditional interview-based surveys, traffic data using in-pavement and video sensors collect vehicle information automatically and require very minimal man power. Moreover, the traffic data collection does not disrupt travelers, unlike the conventional survey approaches. Several studies have been conducted to estimate Origin-Destination (O-D) matrices from traffic counts. Several models and algorithms were developed to estimate O-D matrix from traffic counts – Van Zuylen and Willumsen (1980) used information minimization and entropy maximization approaches. Fisk and Boyce (1983) adopted a network equilibrium based approach, Cascetta (1984) used a generalized least squares method, Reddy and Chakraborthy (1998) developed fuzzy inference based algorithms, Baek et al. (2004) used genetic algorithms to estimate multi vehicles O-D matrix.

Abrahamsson (1998) attempted to provide a literature survey classifying different models based on transportation modeling and statistical inference approaches, and gradient based solution techniques. Traffic counting stations and the number of traffic counts available in a regional network are limited, typically far less than the number of links available to obtain solutions. Traffic count locations and disappearance of traffic flow at link ends are a few problems associated with estimation using traffic counts. Associated problems and alternatives are discussed in detail later in the report.

2.3 Traditional Transportation Planning Models

Urban transportation planning in North America started in the year 1953 when the first transportation study was conducted in Detroit, Michigan followed by Washington D.C and Chicago, Illinois in the next two to three years. Later, many metropolitan areas in North America and developed European countries started conducting major transportation studies. The traditional sequential Four-Step Model (FSM) progressively evolved into an established methodology over the past fifty to sixty years. It is used to estimate and forecast traffic on the road network. The FSM consists of trip generation, trip distribution, mode split and traffic assignment derived from the travelers necessity to perform an activity from available choices.

The FSM requires a significant amount of data to define travel and transportation systems. The transportation network is generally represented as a graph with links and nodes. A variety of travel and activity surveys supplements the travel or activity data needed for the FSM. The information gathered by such surveys is typically aggregated to a zonal level of convenient size that represents a group of individuals or households over geographical space. The zones are typically termed as traffic analysis zones (TAZ). Defining the TAZs is an important part of modeling travel demand because the level of aggregation has a significant effect on the results and ultimately the policy measures.

The selection of TAZ size and number depends on several factors – socioeconomic, demographic and land use characteristics and also on the project objectives and the type of study conducted. The size and area of the TAZ plays an important role throughout the FSM, from calculating the number of trips generated from each zone to trip distribution and assignment. The selection of routes gets complicated with smaller TAZs (more number of internal zones). The smaller the area of a TAZ, the greater the detail captured of an area with a tradeoff between modeling complexity and computational tractability. While modeling travel demand, regional barriers such as county, city and municipal jurisdictions are typically ignored and the region as a whole is considered. The trips expected from areas out of the study area are modeled as external zones or stations, connected to the network on its periphery. The trips originating from each TAZ are loaded on to the network from the centroid of the TAZ to physical links in the network using centroid connectors. A centroid typically represents the attributes of the entire TAZ. The basic sequential FSM is presented next, though the order of sequence may not be the same in practice. A brief overview and deficiencies in each step are also discussed.

2.3.1 Trip Generation

Trip generation is a process of estimating the total number of trips produced or attracted by each TAZ (at an aggregate level) as a function of socio-economic, demographic, employment and land-use characteristics. Trip productions (P) and attractions (A), termed in a sense synonymous as the total number of trips produced (origins) should be equal to the total number of trips attracted (destinations) by each TAZ in the system (study region). Because trips produced in an origin should have a destination, all the productions in origin zones should be balanced by trip attractions in the destination zones in the study region. The number of trips produced in a TAZ commonly from home (origin) depends on population size and density, household size, structure, income levels, car ownership, and accessibility. On the other hand, trip attractions depend on employment, land use type (industrial, commercial, retail, and recreational) and floor space available. Trips are typically classified by purpose (home based work, home based other and non-home based trips) and time of day (AM, PM). The main problem with trip attractions is the data availability. While significant progress and understanding has been observed in the production models, literature documents limited research for models based on trip attractions. Data collection efforts for modeling trip attractions were also observed to be minimal in practice. Trip attractions are primarily indices of relative attractiveness of zones helpful in later stages (trip distribution) of the FSM. Trip generation (P/A) is generally carried out using two methods, linear regression analysis and the cross classification/category analysis (Ortuzar and Willumsen, 2001).

Multiple Regression Analysis:

Multiple regression analysis derives a linear relationship between the number of trips generated, a dependent variable, and different explanatory or predictor variables, typically termed as independent variables, with parameters to explain the weight of the variables in determining the dependent variable. Trips can be estimated at TAZ or household level with household level values typically aggregated to TAZ level. Regression analysis employs strict statistical assumptions like linearity, normality and homogeneity (variances of the observed data are similar). The presence of correlation between the selected independent variables is a major concern that leads to unstable regression parameter values. In the spatial context of trip generation models using regression, spatial dependency and heterogeneity properties have significant ill effects in parameter estimation (Miller and Shaw, 2001). Spatial dependency is a phenomenon in which independent variables are related to each other over space (often measured as the degree of spatial autocorrelation). Spatial heterogeneity is an inadequate representation of estimated parameters of spatial model at the local level, because variables are non-stationary over space. Problems of spatial heterogeneity can be addressed with newly developed Geographically Weighted Regression (GWR) technique (discussed in later sections of the report) that allows variation in the independent and dependent variables at local level. The parameters are estimated using weighted least squares based on local weighting functions.

Cross-classification/Category Analysis:

Regression based trip generation models dominated until the late 1960's (Ortuzar and Willumsen, 2001). Later, many preferred category or cross-classification analysis to regression models in the FSM. Various attributes of household data are grouped to classify under each category. The response of such cross-classification is estimated from the cell values of each category. The advantages of this method are (Ortuzar and Willumsen, 2001): groupings are independent of the zone system, no prior assumptions of the relationship are required (in regression analysis a linear relationship is assumed at the beginning itself) and relationships can vary in each category by the group classified. The disadvantages are that the classification is restricted by the upper and lower ends, there is no statistical goodness of fit and a large sample size is required for reliable results.

Trip production estimates are generally considered superior and comparatively well defined than attractions. Once the trip productions and attractions are estimated, they are matched using trip balancing techniques for consistency typically by a factor based on total trip productions (since production estimates are comparatively more reliable). This is necessary, especially in the subsequent trip distribution step of the traditional FSM.

2.3.2 Trip Distribution

Trip distribution attempts to model the way in which the generated or attracted trips to various zones are linked. Principles from different areas like physics (gravity model), sociology and psychology (intervening opportunity and destination choice models) were used to demonstrate the urban travel phenomena. The gravity model adapted from Newton's gravitational law of physics is commonly used in trip distribution. According to the method, the trip attractions between origin-destination TAZs diminish with an increase in the distance between the TAZs. Growth rates observed from the previous year data and experience were also generally applied for trip distribution (Stopher and Meyburg, 1975). However, the traditional gravity model does not consider temporal variations of trips, special attractions, and future developmental attractions. Moreover, travel time matrices for inter and intra-zonal trips are required for both base year and forecast year and the traffic mix by mode is undefined at this stage to predict travel time matrices accurately (Stopher and Meyburg, 1975). The limitations of this step are hence transferred to the next step; "the modal split" (if mode split follows trip distribution).

2.3.3 Mode Split

Mode split is typically performed after trip distribution, though in some cases it is performed after trip generation and before distribution. The origin destination volumes are split or distributed by available alternate modes in this phase. Trip-end models used before the trip distribution and after generation are good in small networks for preserving characteristics of individuals. However, they are not viable for large networks where different modes have various levels of influence on the choice of the mode (Ortuzar and Willumsen, 2001). Trip interchange models are based on service characteristics, like travel time, cost and accessibility to different modes. Since the travel costs are to be determined in trip interchange models, they always follow trip distribution. Problems with carrying the trip distribution step prior to mode split were discussed in the previous subsection.

2.3.4 Trip Assignment

Trip assignment is the final step of the FSM process. The trips from a given origin to a given destination on a given mode obtained in the previous step are now assigned to routes comprising of a set of links. The trips are assigned to links based on shortest path or minimum impedance (travel time) paths for a no congestion scenario. Several techniques, methods and market equilibrium theories are generally used in practice. Trip assignment or network assignment in congested networks is addressed using equilibrium principles developed by Wardrop (1952) in his seminal work on road traffic research. User optimal and system optimal principles, Wardrop’s first and second principles, respectively, are built on assumptions that the users do not have a choice to change routes to minimize cost and the entire system balances out to equilibrium (Miller and Shaw, 2001). Dynamic models to account for time variations, stochastic models for allowing user cost minimization, dynamic traffic assignment to predict and incorporate future traffic in the iterations and advanced variational inequality models entertained in the trip assignment are still left with several questions.

2.3.5 Summary of Drawbacks of Conventional FSM Process

(i.) Travelers may not follow the same sequence of steps of the FSM. Traveler behavior and choices are much more complex than the modeler's definition of sequence.

(ii.) The conventional approach lacks the single unifying rationale that would explain all aspects of demand jointly and in the presence of congestion (Oppenheim, 1995).

(iii.) Due to the sequential top down approach, each level is treated serially and independent of others. The outputs are carried to the next levels. Likewise, the errors, assumptions and relaxations applied are also carried over to the succeeding steps.

(iv.) Travel costs are dependent of travel volumes and vice versa if congestion is present on the road network. Iterative processes applied in the present travel demand estimates might not give single convergent solutions (Boyce, 2002).

Figure 2.1 below outlines the traditional four-step process with feedback loops.

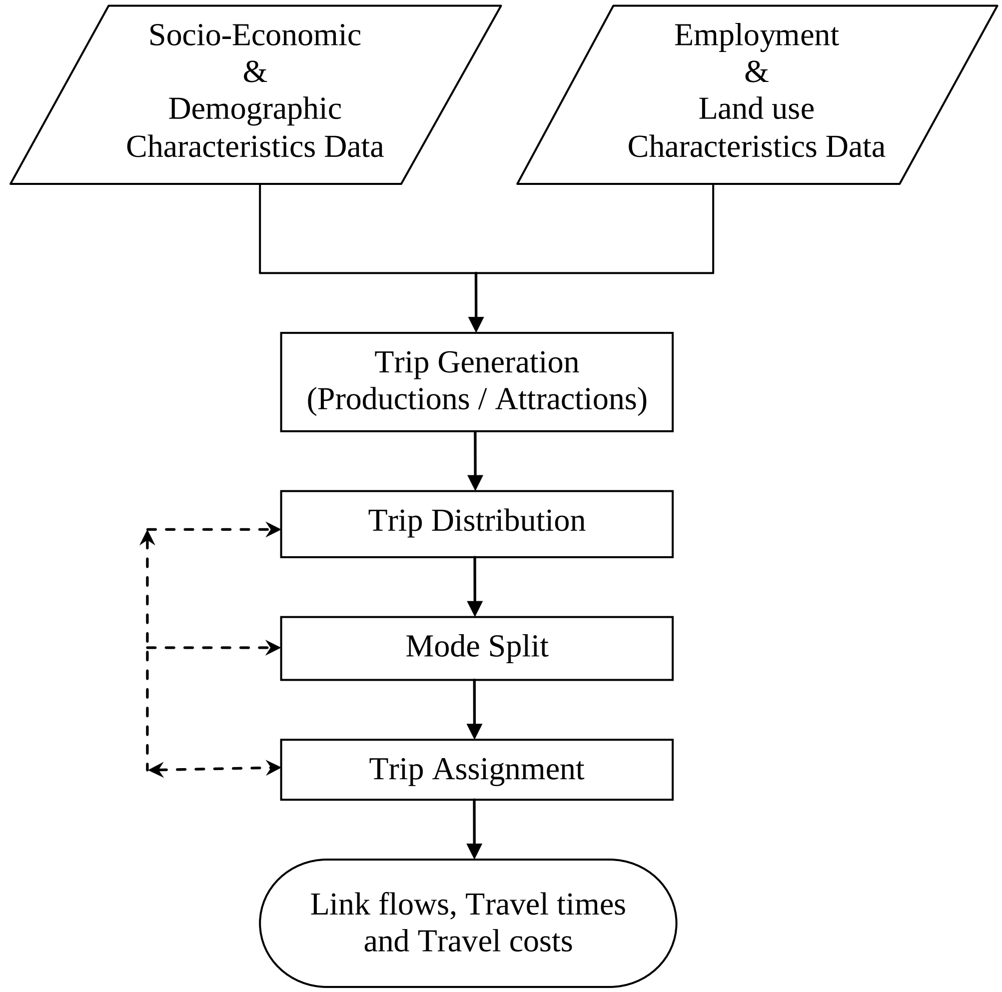


FIGURE 2.1: The Four Step Process

There are several other critiques on the traditional urban transportation planning process. Readers may want to refer to Stopher and Meyburg (1975), Ortuzar and Willumsen (2001), Oppenheim (1995), and Miller and Shaw (2001). A detailed review of travel forecasting, history, development and application of network traffic equilibrium was also discussed by Boyce (2007).

2.4 Combined Four Step Methods – Merits and Demerits

In order to reduce errors and uncertainties transferred in the sequential approach, combined models were developed to carry out all the four steps simultaneously. These include efforts by Beckman et al. (1956), Florian, Nguyen and Ferland (1975), Evans (1976), Florian and Nguyen (1978), Friesz (1981), Fisch (1985), Safwat and Magnanti (1988), Oppenheim (1995), Bar-Gera and Boyce (2003), Boyce and Bar-Gera (2003, 2004), Ho et al. (2006), and Hasan and Dashti (2007). Most recently, Zhou et al. (2009) developed a mathematical programming based variational inequality formulations for a combined model integrating all the four steps of a travel demand model.

Safwat (1982) developed a simultaneous transportation equilibrium model (STEM) and later applied it (Safwat and Magnanti, 1988) to the intercity road network in Egypt and the urban transportation network in Austin, Texas. They found that the model can have sufficient behavioral richness and computational advantages. Boyce et al. (1994) compared FSM with feedback and combined trip distribution, mode split and assignment steps with a sketch planning model for a network in Chicago, Illinois with 300 TAZs and 3,000 links and suggested the use of a combined model. Oppenheim (1995) in his book on *Urban Travel Demand Modeling* provided good theoretical foundation on combined travel demand models in congested and uncongested networks.

Hasan and Safwat (2000) compared the traffic predictions by STEM, and FSM for a small transportation network in Tyler, Texas. The results indicated that the simultaneous approach performed better when compared to the sequential approach. Siegel et al. (2006) compared the urban travel forecasts prepared with sequential and combined methods for City of Concepcion, Chile, which is a smaller network again. The study found key inconsistencies in the traditional sequential approach without applying feedbacks, particularly in congested conditions. Feedbacks reduced error propagation and similar solutions were found with respect to combined models, but important inconsistencies remained. Yang and Chen (2009), in a most recent study, conducted a gradient-based sensitivity analysis of a combined travel demand model considering five applications - identification of critical parameters, paradox analysis, access control, destination choice, and error and uncertainty analysis. They demonstrated the usefulness of the analysis for assessment of various transportation improvement policies.

Practical application of simultaneous or combined models are limited with very few exceptions of real world applications in Egypt; Riyadh, Saudi Arabia (Hasan and Al-Gadhi, 1998); Austin and Tyler, Texas (Hasan and Safwat, 2000); Economic and Social Commission for Western Asia (ESCWA) countries (Safwat and Hasan, 2004); and City of Concepcion, Chile (Siegel, 2006). Siegel et al. (2006) suggested that more testing is required with different kinds (size and operating conditions) of networks. Moreover, combined models are mathematically complex to incorporate temporal and spatial variations. It is also difficult to change the existing regional models to a new method without testing it on a wide variety of networks. The extreme hardship for a wide spread application of these models is the availability of a software and the lack of knowledge of the underlying principles of the models for a common transportation modeler. Simultaneous models are still aggregate models and do not accommodate local variations in the modeling process.

2.5 Estimation of AADT

Typical traffic count types are permanent, seasonal and short term counts, collected at portable traffic counting (PTC) sites or telemetry traffic counting (TTC) sites with little variations in the terminology used by various transportation departments. However, oftentimes, traffic counts are not available for all the roadway segments in the transportation network due to associated problems in collecting, storing and transferring data whenever needed. The lack of funds or resources to collect data at several locations continuously throughout a year is another problem.

Short term traffic counts that are less expensive and easy to collect are typically used to estimate AADT by applying appropriate adjustment factors available in the *Traffic Monitoring Guide (2001)*. Ritchie and Hallenbeck, 1986; Sharma et al, 1996a; Stamatiadis and Allen, 1997; Granato, 1998; Zhao et al., 2004; and Li et al., 2006 provide information on various types of factors applied to short duration volume counts. The AADT values for several roadways are estimated to replicate the homogenously classified roadway segments with similar temporal properties using data collected with temporary traffic recorders located along the homogenous group road segments (Goel et al., 2005). Several methods have been proposed earlier to accurately estimate AADT volumes from a sample of traffic counts available. Estimating data with a far less number of “knowns” compared to a very large number of “unknowns” is always associated with errors and inaccuracies.

Several attempts were made to estimate, predict and forecast AADT from available intermittent seasonal coverage and short duration count data. State, county and local transportation departments use growth factor, time series or linear regression analysis techniques to estimate or forecast AADT values (Park, 2004). Time series analyses assume that the past trends are repeated in the future and forecast based on the trends observed. Growth factor models simply apply annual growth rates (growth rate trends) observed over a period of history (Smith and Demetsky, 1997). Regression analysis is commonly used in explaining linear relationships between dependent and independent variables (factors expected to influence variation in AADT in this context).

Neveu (1983) developed a quick response method to forecast rural highway traffic at specified locations in New York. The author developed elasticity models to estimate traffic on rural roads based on road functional classification - Interstate, principal arterial and major and minor arterials. Continuous count data (present year AADT) in rural locations modified by various town (population, housing units and households), county (population, housing units, households, automobile registrations, employment, labor force, personal income, and income per capita) and state (gasoline sales) level demographic and socio-economic (predictor) variables were used in developing the models.

Mohamad et al. (1998) developed a similar multiple regression model to estimate AADT from relevant demographic variables for county roads. The authors found that county arterial mileage, population, location and access to other roads are influential in predicting the future AADT values. Xia et al. (1999) attempted to estimate AADT for non-state roads that do not have traffic counts in Broward County, Florida. The authors used predictor variables such as roadway characteristics (number of lanes, area type and functional classification), socio-economic data variables (different types of employment, school enrollment and hotel occupancy) and accessibility to state and non-state roads.

Accessibility of non-state roads to other county roads, number of lanes, area type, functional class and auto ownership and service employment (being comparatively less predictive of all) were found to be significant predictors of AADT. Service employment was later excluded from the final AADT model developed for the same study area by Shen et al. (1999, as cited in Zhao and Chung (2001)).

Zhao and Chung (2001) extended the models used in the above two studies using larger data sets. The authors included AADTs on state roads and used new federal functional classification to develop four multiple regression models. The authors found regional accessibility to employment centers to be comparatively more predictive than other variables and included in the final model. Overall, six predictor variables were included in the final model. They are number of lanes, functional classification, direct access from a count station to expressway access point, employment, accessibility to regional employment and population in a variable sized buffer around a count station. The authors concluded that the number of lanes and functional class are the best significant predictors and models without functional class are the worst performers. They felt that the performance of the model improved significantly but was inadequate to meet the engineering design and travel demand model calibration needs. They also suggested that further examination of causes and spatial pattern of errors are needed and expressed a need for more effective land uses.

Sharma (1999) developed a neural network based multi-layered, feed forward, and backward propagation design method to estimate AADT from 48 hour traffic counts. Artificial Neural Network (ANN) and regression analysis techniques were compared by Lam and Xu (2000) to conclude that ANN performs better than regression techniques in predicting AADT (as cited in Zhao and Park, 2004). Sharma et al. (2000) made a similar comparison study that resulted in a favorable comparison between literature reported factor approach values and neural network model error values for low volume roads.

Tang et al. (2003) developed four models for the short term prediction of daily traffic flows by day of week, month and current AADT for the entire year of 1999, using time series, neural network, Non-Parametric Regression (NPR) and Gaussian Maximum Likelihood (GML) using historical (1994-1998) and current year data. They found GML to be the most promising and robust of all the models.

Zhao and Park (2004) used the GWR technique that allows local model parameter estimation instead of global parameters used in an Ordinary Least Squares (OLS) linear regression analysis. The authors investigated spatially variable parameter estimates and local R-Square from the GWR model to analyze the errors in AADT estimation. It was found that the GWR model provided accurate AADT values when compared to OLR models.

Jiang (2005) and Jiang et al. (2006) incorporated image based vehicle data with AADT estimation and found improved accuracy. Lam et al. (2006) compared NPR and GML, both being nonparametric models, to find that NPR performed better using annual traffic census data in Hong Kong.

Recently, Neto et al. (2009) used Support Vector Regression with Data Dependent parameters (SVR-DP) using 20 years of data for rural and urban roads in 25 counties in the state of Tennessee. The forecasted AADT results were compared with OLS regression results and Holt-Exponential Smoothing (Holt-ES). The SVR-DP method performed better than both the methods. Wang and Kockleman (2009) developed Kriging based methods for network and count data mining over time and space that performed better than other spatial extrapolation options.

Despite the presence of several models and studies, estimation of AADT in any of the models is neither accurate nor close to reality. There are several uncertainties in the estimation process similar to any estimation or forecasting models, especially with the limited and inconsistent data. One possible reason could be the different traffic data collection methodologies adopted by individual states. For instance, data collection schedules and methodologies to obtain AADT values for different types of roads adopted by North Carolina (NC) are different from the Commonwealth of Virginia, New York (NY) and Florida states (found on state department of transportation (DOT) traffic data web pages of NCDOT, VDOT, NYSDOT and FDOT, respectively). NCDOT traffic data collection corresponds to the *Traffic Monitoring Guide (TMG)* published by Federal Highway Administration, FHWA (TMG, 2001).

Sharma et al. (1996b), in an extension to earlier studies (Sharma and Leng, 1994 and Sharma et al., 1996a), addressed statistical accuracy of AADT estimates for seasonal traffic counts (STC) with statistical precision of short period traffic counts (SPTC) analyzed using automatic traffic recorder (ATR) data from Alberta and Saskatchewan provinces in Canada. SPTC sites were assigned to homogeneous ATR groups to calculate AADT values using respective expansion factors of the ATR group. Appropriateness of volume adjustment factors, expressed in terms of assignment effectiveness, is used in the study to represent the degree of correctness in assigning the sample sites to an ATR group. The authors stressed the need for effective assignment of count sites. They found that estimates of a properly assigned 6 hour counts proved better than the improperly assigned 72 hour count sites. Grouping of fairly similar (homogenous) ATRs with temporal commonalities in traffic for practical applications can be carried out using the most recent *Traffic Monitoring Guide (2001)*.

Consistent with previous research findings, Granato (1998) found that continuous consecutive day count improved the estimation only by 5%. Using a single ATR count station data in Cedar Rapids, Iowa, a 25% (one-quarter) error reduction in the AADT estimates was found with the application of day of week and month of year factors when compared to using continuous 24 hour counts (though the results based on single count observation is questionable, the results support previous research findings). Granato suggested using multi day traffic counts scattered across two or three weeks over consecutive-day counts.

Davis and Guan (1996) developed a Bayesian estimator of MDT to assign a short count site when an unclear situation arises to which group an ATR is to be assigned. Davis (1997) stated that the errors in short count estimation due to seasonal and day of week adjustments were neglected in the previous attempts. The estimation errors can be substantive with incorrect adjustments. The author concluded that when appropriate adjustment information is lacking, seasonal counts are preferred for accurate estimation over short term counts.

Davis and Yang (2001) attempted to understand the uncertainties associated with the estimation of total traffic volumes from a sample of daily traffic volumes based on traffic data variability equations. A computationally practical empirical Bayes approach was used to compute quantiles of predictive probability distribution of traffic totals. The median or the 50^{th} percentile of the predictive distribution using the probable ranges and their associate probability values were used to obtain total traffic. AASHTO (1992) and TMG (1995) recommends that Mean Daily Traffic (MDT), typically used for calculating AADT, can be estimated with acceptable precision by suitably adjusting 48 hour short counts (Davis and Yang, 2001). The authors identified two sources of potential errors – 1) the sampling error being unrepresentative of the entire year traffic flow, and 2) adjustment errors that arose from the applied adjustment factors to account for seasonal, day of week and month of year traffic flow variations.

With an impression that assigning SPTCs to factor groups based on count station proximity is oversimplified and prone to subjectivity, Li et al. (2004) conducted a regression analysis for estimating seasonal factor (SF)s that contribute to seasonal variations in traffic volumes. According to Li et al. (2006), typical established factor groupings are based on short count station proximity to permanent count station, functional class or engineering judgment. The authors developed a fuzzy tree construct based on known SF groupings of permanent count stations and their four land use categories to determine the SF category of a given portable count station. The land use categories used did not sufficiently represent the permanent count station locations and, with limited sample size, the traffic variations were not completely explained, due to which, ambiguity still remained in the results.

Goel et al. (2005) argued that segment traffic volumes are a result of traffic flow from origin (O) to destination (D) and AADTs estimated from coverage counts and adjustments on overlapping (O-D) segments need to be correlated. A correlation based method derived from Generalized Least Squares (GLS) that exploits correlations between 24-hour segment volumes and a traditional OLS estimation method were compared using

Monte-Carlo simulations on a small network representing intercity flows in Ohio. The authors found remarkably low errors in AADT estimation using the GLS method compared to the traditional OLS method with high correlations between segment volumes, whereas little variations were found with low correlations. In reality, the network segments have more of a variety of mix than the one considered by the authors (single coverage segment compared with single permanent ATR (PATR) segment). The results indicated that the method performed well only if correlations are high, leaving segments with low correlations in confusion without assignment to a PATR, or assigned to an improper PATR. Also, more realistic networks will have more O-D pairs sharing a highway segment and more PATRs over space near a coverage count station.

Eom et al. (2006) suggested that spatial dependencies need to be considered to estimate AADT on lower classification roads like collectors and local roads. This is because traffic volumes at one monitoring station are often correlated with volumes at neighboring stations. The authors used a geostatistical approach called kriging to develop a spatial regression model to improve AADT estimation. Both spatial trend and correlation on non-freeway facilities in Wake County, North Carolina were taken into account. The results indicated that the predictive power of the spatial regression model is much better than the traditional method. The prediction is more reliable in urban areas when compared to rural areas. In addition, it was expected by the authors that the model could be very useful in predicting traffic volumes at locations where observed data is not available. The only concern is that the Euclidean distances between any two stations considered by the authors might not be the shortest in reality. However, better predictions are expected if the actual roadway distances are obtained.

Gadda et al. (2007) attempted to quantify all major sources of errors like factoring errors, sampling errors, misclassification, and spatial and temporal approximations present in the AADT estimation process in one piece using the datasets of Minnesota, Florida, Southern California and Austin, Texas. Errors resulting from spatial extrapolation were studied using network flow estimates from Austin's travel demand model as a function of distance from nearest sampling site. The authors found that the results were consistent across states, indicating transferability in different contexts and suggested classification by fine clustering on the basis of functional class, lane count and multiple area types. A dramatic increase in spatial errors was observed beyond a 0.5 mile buffer in urban areas and beyond a 1 mile buffer in rural areas.

2.6 Spatial Principles, Accessibility and Modeling

Several models and principles have been developed and used for many years to understand and predict the interactions between places of attractions and productions. Principles and methods related to physics, economics, and information theory has been borrowed and used to demonstrate spatial interactions of various land uses. Spatial interactions are generally modeled using the gravity model, the intervening opportunity model, entropy maximization and destination choice models. The gravity model is the most extensively used spatial interaction model.

2.6.1 Gravity Model

The original gravity model of gravitational forces between two masses separated by a distance is converted by replacing masses with population of origins and destinations for transportation planning applications. The population at origins and destinations were then replaced by trip productions and attractions at the respective places, with distance replaced by negative exponential and power functions (Stopher and Meyburg, 1975). The gravitational constant was replaced with the constant K (usually between 1 and 2), which was later split into two constants that improves the model when trip conservation rules are applied (pg. no. 141; Stopher and Meyburg, 1975). The exponent and power functions are replaced by travel time or distance decay functions that are determined from calibration. Stopher and Meyburg (1975, page no. 142) described the gravity model as follows.

$$T_{ij} = P_i B_i A_j C_j f\ (D_{ij}) \qquad [2.1]$$

Where, T_{ij} = trips from i to j,

P_i = total trips produced by zone i,

A_j = total trips attracted by zone j,

B_i, C_j = constants associated with productions and attractions respectively,

$f\,(D_{ij})$ = measure of spatial separation between zones i and j.

2.6.2 Accessibility

The term accessibility is defined in several forms in the literature based on the context in which it is used. Accessibility, in general, is the ease at which a location or facility can be reached. In transportation planning, a region is divided into TAZs and accessibility is expressed as a function of intensity of activity at a location and spatial separation and impedances in reaching other points (Hansen, 1959). In other words, accessibility is directly proportional to intensity of activity (opportunities/employment, population) and inversely proportional to the impedance function (distance, travel time or travel cost) slightly concurring with Stewart's (1948) demographic gravitation concept. Accessibility measures developed by Stewart and Warntz (1958) and Hansen (1959) are based on the weightage of the size of their locations (Pooler, 1995). Pooler (1987) reviewed geographical accessibility in terms of population potential based on Stewart's concept.

Hansen's empirical study on residential developmental patterns suggested use of accessibility for residential land use model. Neuburger (1971), in his discussion on user benefits of transport and land use plan evaluations, stated that improved transportation network increases access to places and ensues more travel. Wachs and Kumagai (1973) expressed their concern over the negligence of regional accessibility (mobility of total population in a region) in comprehensive urban transportation planning. The authors stated that regional accessibility can be an important constituent of social report, helpful in reaching the national and regional objectives of equality in opportunities. In their analysis, they found significant differences in the regional accessibility to employment and healthcare centers primarily due to socio-economic status and spatial location of communities in a region.

Savigear (1967) discussed the measures of accessibility in terms of distance from remaining zones and activity in the central area zones. This study only looked at the accessibility to the central area. It was later extended by Ingram (1971) to any area in the region. Ingram (1971) introduced concepts of relative and integral accessibility. Relative accessibility is defined as the degree of connection between two points, whereas integral or total accessibility is defined as the degree of interconnection of a point with all other points on a surface. The author stated that attraction related characteristics like employment used by Hansen (1959) are distributed unequally over space and reflect spatial variations in both degree of accessibility and attractivity. Ingram assessed various simple linear to rectangular, reciprocal and negative exponential functions based on normal and Gaussian curves. The study found Gaussian curves to be the most applicable for quantitative measurement of accessibility, but left the constant to be used in the function “unknown”, requiring further work to determine the constant.

Dalvi and Martin (1976); and Weibull (1976) went little further and added attraction characteristics to the points under consideration in an integral accessibility scenario used in Ingram’s method. Dalvi and Martin (1976) found that accessibility measure is highly sensitive to the attractor variable. They defined accessibility as a measure of ease to reach a land use activity from a location using a private transport mode. The accessibility pattern reversed between the central area and peripheries when household and population variables are replaced by employment variables. A sensitivity study of zonal configuration and aggregation showed strong sensitivities with accessibility. Similar problems identified in transportation demand literature were found in the zonal analysis. Interestingly, they did not find any effective relationship between trip generation and accessibility measure. While a few studies observed a relationship between accessibility and trip generation, a few others found accessibility to play a role in explaining trip making (Thill and Kim, 2005).

In the late 1970’s, researchers started using concepts of accessibility in transportation planning and trip making related applications. From the earlier studies it can be understood that attractions and activities (the prime factors that drive travel) can influence accessibility. Koenig (1977) and Black and Conroy (1977), as cited in Morris et al. (1979) argued that accessibility is a good criterion for transportation planning that can evaluate land use patterns and transportation systems performance. Morris et al. (1979)

described time, money and human effort as variables of travel choices borne by the communities that motivate level of service use and participation in desired activities and are consequently influenced by accessibility. They developed interrelations among accessibility indexes and underlying theories and linkages with consumer demand, evaluation, and accessibility.

Leake and Huzayyin (1979) appraised the use of accessibility measures in trip generation modeling. They developed accessibility indices for all types of modes (highway and transit) based on travel time, distance and public service frequency to represent the transportation system. Koenig (1980) developed Hansen type accessibility measures to estimate zonal trips at an aggregate level and person trips at a disaggregate level. It was believed that quality of transport and availability of attractive destinations is going to affect the number of trips made by people apart from their socio economic characteristics. Koenig believed that introducing the accessibility indicators into the trip generation modeling shall improvise and resemble equilibrium models. A study similar to Dalvin and Martin (1976) was conducted in five French cities by Koenig, who developed graphs for all the cities classified by age group, car ownership, and people employment. In all the cases, accessibility emerged as a powerful player in modeling trip generation.

Downes and Morrell (1981), in their study using household data in 1971, found a little effect of accessibility in terms of household location from the central area on travel. The authors said that their findings supported the early research of Fawcett and Downes (1977); and Downes et al. (1978) that the household trip making depended mainly on the household size and vehicle ownership. Similar conclusions were made in a study conducted by Williams (1989) to identify household travel related behavior and its relationship with residential accessibility. The socio-economic spatial structure of urban areas ought to have influence over household travel behavior rather than the residential accessibility conditions.

Hall (1983) proved that the performance of transportation infrastructure and the outcome of the destination activity are two important factors that can dictate accessibility. In other words, travel outcome and performance can affect or define a region’s accessibility to a traveler. A shopping travel demand study in Sussex, UK demonstrated that attraction and accessibility are prime determinants of variations in household shopping activity (Robinson and Vickerman, 1976). Similar findings were found between accessibility levels and shopping travel patterns but, interestingly, not with the trip frequency (Handy, 1994). In other words, no matter what the accessibility levels are or the travel distance is, the residents are going to make a certain number of trips. A higher level of local accessibility is needed for low regional accessibilities, and vice versa. A better access to regional centers lowers the impact of local activity. On the other hand, an improved or greater activity at the local level lessens the impact of good access to regional centers. Local and regional accessibility of shopping travel were assessed in the study and suggested to enhance accessibility at both the local and regional levels.

Guy (1983) assessed access to local shopping opportunities with seven accessibility measures, using shortest distance (developed by the author) and cumulative opportunity, gravity, and Gaussian indices. Song (1996) successfully tried to statistically assess the usefulness of accessibility measures. The author evaluated nine accessibility measures based on a population density function using maximum explanatory power in a standard regression and a non-nested test on alternative pairs and concluded that gravity type accessibility measure performed better than any other measure.

A slight deviation to the conventional accessibility measurement but an emergence of promising measures took place in terms of transportation network design. Current et al. (1987) introduced a median shortest path problem (a combination of p-median and shortest path problem) to analyze cost and accessibility in designing transportation networks that deal with a completely different area of accessibility. In optimization problems like shortest coverage and maximum coverage/shortest path problem, accessibility is measured as maximum covering distance for location of facilities analysis. Interested readers may want to look at Current et al. (1984), Current et al. (1985a and 1985b), Current and Min (1986), Balinski and Spielburg (1969), ReVelle and Swain (1970), Toregas et al. (1971), Toregas and Revelle (1972), and Church and ReVelle (1974, 1976) that dealt with optimization, linear programming and location problems.

Kockelman (1997) researched on the influence of accessibility, land use mix and land use (job housing), and their relative significance on travel behavior. The author used a gravity index for accessibility, entropy for land use balance and dissimilarity index for land use mix. These measures emerged as more relevant players in predicting travel behavior than the conventional household and travel characteristics. A strong linkage between accessibility and travel behavior was found and proved to be a much better predictor of vehicle kilometers of travel and mode choice than density.

Sathisan and Srinivasan (1998) developed an accessibility index useful for transportation planning purposes. The method involves a series of steps developing an impedance network, an accessibility network, a ratio of accessibility, the extraction of demographic data and finally, a construction of accessibility index.

Thill and Kim (2005) found that both trip productions and trip attractions are significantly affected by geographic accessibility between origins and destinations, both at aggregate (TAZ) and disaggregate (household) levels. The authors used a combination of various mathematical formulations, spatial impedance functions, travel cost and attractiveness proxies and parameters by plugging in 72 metrics of accessibility in each demand model. The authors acknowledged that the significant positive results achieved in their analysis are because they did not make any a priori assumption anywhere, throughout the process of measurement of accessibility. Their research also indicated that various trips made are driven by different relationships to accessibility and compliments Kwan's (1998) conclusions that accessibility is situation or context dependent (Thill and Kim, 2005).

There is another view of accessibility based on a constraint-oriented approach implemented by Hägerstrand's space-time prisms (Miller, 1999). The author derived a space-time accessibility and benefit measure incorporating location, time and distance applicable to an urban transportation network. Kwan (1999) argued that the conventional accessibility measures ignore the complex activity of travel behavior and the role of space-time constraints that motivate an individual's accessibility experience. Kwan's study investigated individual access by gender based on the space-time construct measure of accessibility. The study found that females have significantly less access to urban opportunities than men. The differences observed according to the author are due to differences in space-time constraints. Many studies followed the Hägerstrand's (1970) seminal space-time construct and geographic framework. Lenntorp (1976) and Burns (1979) are examples of early operational formulas (Kwan, 1999). Interested readers may want to look at Kwan (1998), Miller (1991), and Recker et al. (2001) for applications of space-time constraints in accessibility measures.

Murray and Wu (2003) indicated that accessibility is a tradeoff between public access to bus stops and geographic coverage in their search for optimal spacing of bus stops along a route. Other research contributions related to transit accessibility include efforts by Murrary (2003), O'Sullivan (2000), Polzin (1999), and Pulugurtha et al. (1999). Liu and Zhu (2004) developed an integrated GIS tool for accessibility analysis called "Accessibility Analyst", capable of performing cumulative-opportunity measures, gravity-type measures, and utility-based measures.

To summarize, there are several ways to measure accessibility. Research shows that accessibility measures are very useful in various transportation, urban planning, public transport and geography related issues. Looking at several accessibility indicators discussed so far, one can understand that while one study looks at one aspect of accessibility, the other looks at accessibility from a different angle and tries to improve or suggest a different way of addressing the problem. It is not always possible to address each and every aspect in a single measure, especially when accessibility issues are concerned. Two basic elements were always found in the discussion of various accessibility measures: "impedance" and "attractiveness" (Thill and Kim, 2005).

2.6.3 Geographically Weighted Regression

National, state or local level statistics were often seen in representing or reporting various phenomena with an averaged global figure to a particular context (example, travel time index). They are typically averaged from a range or group of values giving equal weight to each individual group. Relatively accurate results can be obtained by looking at spatial maps with corresponding local variations of attributes and their relationships to approximate global figures (generally single valued).

An objective of understanding the relationships in the attribute data is to include significant variables that contribute to the strength of the relationship. Consider trip generation using regression analysis for instance. The trips produced are estimated using a variety of demographic, socio-economic, employment and land use characteristics. Trips generated at household level are aggregated to TAZ level (a global figure to that particular zone and forced to be concentrated at zonal centroid to support aggregate FSM). In reality travel preferences vary with each individual, based on several factors that make the global parameters estimated in the trip production incorrect. Irrespective of the care taken in allocating the area of a TAZ, problems continue to stay that leads to growing interest in disaggregate travel demand modeling practice. Another example is travel time index (TTI). A user is interested in travel time index of a route for his daily travel instead of a city, region, statewide or national index. City or regional TTI is useful in selecting an area or a city respectively to live. The difference between trip production and TTI examples discussed is that the former one represents spatial relationships and the later represents spatial data. To summarize, local data or local variations in the characteristics of data attributes are very useful to accurately utilize information and to develop models with relative accuracy.

There are two types of data, spatial and aspatial – spatial data is represented over geographic space and aspatial data contains only the attribute data without reference to their location. Spatial data analysis and modeling with local variations is of prime importance to the present research. Local variation deals with local disaggregation of global variables. It emphasizes differences across space and helps to search for local hot spots (Fotheringham et al., 2002). Social processes as opposed to physical processes tend to be non-stationary i.e., the models derived in one system can rarely be exactly represented in other systems. This in case of spatial social processes is termed as spatial non-stationarity. It means that the investigating processes under consideration are not constant over space and varies with the locality in which case a global statistic will be misleading locally.

The local variations of spatial dependencies of attribute values were addressed using spatial autoregressive models developed by Getis and Ord (1992) and Ord and Getis (1995, 2001). Brunsdon et al. (1998a) later applied the conceptual and theoretical models of GWR to the Ord models to demonstrate the problems associated with using global models of spatial association using an empirical example with data on owner-occupation in the housing market of Tyne and Wear in northeast England. Interested readers may want to look at Brundson et al. (1996, 1998b, 1999, 2001 and 2007); Foody (2003); for more discussion on the GWR and related research.

The GWR Concept:

The GWR allows local variations to be included in the regression model in which weighting functions define the influence of data in the region defined around the regression points on parameter estimates. Regression points is a concept used in Moving Window Regression (MWR) in which a regional grid of regression points are constructed to calibrate the regression model based on the data defined in the proximity of regression points. The parameter estimates are assigned to the location of regression points. The output results are mapped after conducting similar process for all the regression points at the end. The fundamental idea of the GWR is to use the concepts partly of MWR's regression points and kernel regression in determining parameter estimates for each location of regression point using weighting functions that gives more priority to the data points close to the regression point than to the data points away from the location. It was expected that the weighting function approach used in GWR should provide better results when compared to the MWR method that defines region around a regression point as the four cells in which the regression point is centered (Fotheringham et al., 2002).

The GWR overcomes the associated problems of non-stationarity, heterogeneity and spatial dependence in spatial modeling. Unlike the traditional regression model which estimates single set of global parameters for the entire study area, the GWR allows to estimate parameters for each individual location based on the traditional regression technique assigning gradually decreasing weights with increasing distance to data points away from the point of interest. The GWR model allows local disaggregation emphasizing differences in relationships across space. The final output can be represented in maps convenient to show spatial variations and non-stationarity over geographic space.

Selection of Spatial Weighting Function and Bandwidth:

In a global model, a unit weight is assigned to each observation. In other words, all the observations in a global model are given equal weights. However, in a Weighted Least Squares (WLS) method, weighting functions are used to give different levels of importance to the observations made over space at various distances. Observations at data points in the close proximity of a location are given more weightage compared to an observation made farther away from the location. The section follows with a weighting function employed in an MWR approach and extends to various weighting functions to address problems of discontinuity and spatial variation of influence area. Essentially the idea of GWR is to understand and predict the relationships using observed data that influences a local regression point *i* termed as 'bump of influence' using appropriate weighting functions.

In a global regression model, the unit weightage of observation points is expressed as $w_{ij} = 1$ for all (i, j) where, j represents observation point in space and i represents the regression point. A simple way to exclude observation points beyond a certain distance in the model parameter estimation is giving a weighting function as below (MWR approach).

$$w_{ij} = 1 \text{ if } d_{ij} < d, j = 1, 2, \ldots, n$$
$$= 0 \text{ otherwise}$$

A problem of discontinuity is encountered when such weighting functions are applied. By assigning a zero weight to the observation points beyond a particular distance a sudden change in the estimated parameters is expected when a regression point is changed causing observed points move in and out of a window. A continuous weighting function is proposed to address this problem that allows assigning a decreasing weight with an increase in distance d_{ij}. The two most commonly used methods are (i) Gaussian weighting function and (ii) Bi-square function.

i. $w_{ij} = \exp [-1/(d_{ij}/b)^2]$

ii. $w_{ij} = [1- (d_{ij}/b)^2]^2$ if $d_{ij} < b$

$= 0$ otherwise

where, dij is the Euclidean distance between observation points i and j;

b is the distance decay function also referred to as bandwidth.

In the former case, the weight w_{ij} = 0.5 implies that the contribution of data point j is only half the weight in the calibration process as that of the regression point *i* by itself. The weight for observation points farther away from the regression point *i* is almost nearly zero. This means that all the observation points in the study area should be considered in the calibration process irrespective of its influence at regression point. This concern is addressed with a continuous near-Gaussian weights assigned to observation points up to a distance b from the regression point *i* and zero weights beyond b. The bi-square function is used by an alternate kernel, both the kernels being fixed in their shape and magnitude spatially.

GWR in Transportation:

Given the benefits of underlying concepts of local variations in the regression analysis with geographical weights there is a growing interest in the use of GWR in transportation field. Zhao and Park (2004) research on estimating AADT for the first time used GWR to address transportation related problems. They analyzed AADT estimation errors after investigating spatial variation of parameter estimates and local R^2 from the GWR models. Relationship between AADT and number of lanes, regional accessibility to employment, population, and employment in the buffer area of a count station and direct access to expressways was studied for a regional data in Broward County, Florida, USA. The authors found that GWR models predicted more accurately when compared to OLS regression models. Other research includes GWR in transit ridership models (Chow et al., 2006; and Zhao et al., 2005), relationship between out-commuting and socio-economic variables (Lloyd, 2005), intermodal freight transportation and regional accessibility (Lim and Thill, 2008), land use – transport models (Paez, 2006), car ownership models (Clark, 2007) and trip end models of local rail demand (Blainey, 2010).

2.7 Statistical Methods and Distributions

Since it is practically impossible to collect entire population data (for example, traffic on all types of roads present in the study area) for planning purposes, it is a standard practice to collect samples (a subset of the population), conduct statistical analysis and make inferences about the population. Regression analysis has dominated in the transportation literature (safety and planning) to estimate or predict trips, traffic and crashes. It is used to interpret the relationship between the covariates (predictor variables) and the dependent variable. The variables considered in the analysis could be discrete (integer) or continuous (real). A discrete random variable can take any integer value from zero to infinite, whereas a continuous random variable can take any value on a number line without interruptions.

A probability distribution is associated with any kind of variable that tell the distribution of the population. It is a mathematical formula, that gives the probability of each value for a discrete random variable and a curve that specifies the areas under the curve, and the probability of a continuous variable falls within a particular interval of the curve (Everitt, 1995). Since traffic counts are integers and are non-negative they can be considered as discrete counts. In general, counts are modeled as a Poisson distribution, a generalization of which is the “Negative Binomial distribution”. More about the distributions and various types of models are discussed next.

2.7.1 General Linear Models

Linear regression or ordinary least squares regression models are predominantly used in the trip generation, traffic estimation and other related estimation purposes. A simple linear model looks like the equation (Long, 1997) given below.

$$y = \beta_0 + \beta_1 x + e \qquad [2.2]$$

where, y is the dependent variable,

x is the independent variable,

e is the error term, and

β_0 is the intercept and β_1 is the slope coefficient of the independent variable.

A multiple linear regression equation is formed when more predictor variables are added to the equation on the right hand side.

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 \ldots + \beta_p x_p + e \qquad [2.3]$$

Assumptions of Linear Regression Models:

The four basic assumptions found in standard textbooks (Tabachnick and Fidell, 1996; Long 1997; Hocking, 2003 and Freedman, 2005) on regression modeling are (i) Linearity, (ii) Normality, (iii) Homoscedasticity and (iv) Multi-collinearity, discussed in detail next.

(i.) Linearity:

It is assumed in the regression analysis that the dependent and independent variables have a straight line relationship. Practically, linearity is needed for regression modeling because the ordinary least squares determine the regression line (equation) that optimally selects the minimum sum of the squared deviances (residuals) of all possible straight line relations between the variables. It is evident from equations 2.2 and 2.3 that the dependent variable is linearly related to the independent variable(s) through the parameters β. Often times, the relation between the variables may not be linear. For example, it has been extensively mentioned in the literature that traffic is non-linearly (exponentially) related to the predictor variables in many cases (especially in safety research). Techniques to address non-linearity have been addressed in Berry (1993), Fox (1991), and Tabachnick and Fidell (1996). A common way of establishing a linear relationship for nonlinearly related variables is to transform the variables (by applying square root, logarithm or inverse). Such transformations results in the loss of integrity of the data and also make it hard to interpret the outcomes.

(ii.) Normality:

The most common distribution used in the regression analysis is the normal distribution. Its density function (Everitt, 1995) is given by

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}}\left[-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}\right] \qquad [2.4]$$

where, 'x' is the outcome variable and μ and σ^2 are mean and variance of x, respectively. It has a bell-shaped symmetric distribution about its mean and takes all real values. The distribution is standard normal at mean '0' and variance '1' (Freedman, 2005). According to the central limit theorem, with a large sample size (n), the sample means tend to have an approximately normal distribution (Everitt, 1995). The normal distribution has been extensively used in simple and multiple linear regression analysis when large samples are available to model the data.

However, the problems with assumption of normality have been discussed elsewhere (Tabachnick and Fidell, 1996 and Hocking, 2003). When samples are overly dispersed or negatively or positively skewed it is not recommended to use the normal distribution. Most of the times traffic counts are positively or negatively skewed, which means that the basic assumption of symmetrical distribution over the mean is violated. In such cases a normal distribution may not be a good fit for the traffic counts. Also since the traffic counts can only take discrete positive integer values count models are preferred. Developing a model with normal distribution may result in negative predicted values which is not suitable for traffic counts.

(iii.) Homoscedasticity:

Homoscedasticity is yet another assumption in the linear regression models that assumes a constant variance of errors across all the variables (Tabachnick and Fidell, 1996). It is very likely that the assumption is violated which leads to heteroscedasticity of variance. It leads to distorted estimated variance of the regression coefficients (β), deflates the estimated standard errors of β that in turn inflates t-values relative to the true values (Gardner et al, 1995). This potentially leads to Type I errors (Elhai et al, 2008).

(iv.) Multicollinearity:

Multicollinearity is a problem observed when the independent variables are highly correlated to each other. In other words, the independent variables might have a linear relationship with each other or a linear combination of variables represents a variable. It might also occur from the data collected or represented in which a variable is a sum of the other variables. For instance, total employment could be a sum of all the employment categories in the data (employment categories could be redundant for the model). Due to multicollinearity, error terms are highly inflated ending up with none of the coefficients being significant. Since multicollinearity deals with independent variable(s), it is observed in almost all types of models (including Poisson and Negative Binomial).

2.7.2 Count Models

As the name indicates the count models deals with counted data, the data that represents the number of times something occurred. Count data always takes non-negative integer values. Linear regression models are used to estimate counted data treating them as if they are rather continuous than discrete that results in inefficient, inconsistent and biased estimates (Long, 1997). Count regression models are best suited for analyzing skewed (non-normal) data with non-linear relationships (Elhai et al., 2008). Also the independent variables could be continuous, binary or a mixture similar to normal linear regression model independent variables.

Poisson Models:

Poisson model is the basic and simplest count regression model (Long, 1997 and Elhai et al., 2008). Unlike the normal linear regression models in which the dependent variable is transformed (logarithmic in most cases), in a Poisson model the probability of counts are determined by a Poisson distribution (a probability distribution of non negative integers) in which the coefficients are exponentiated. By doing so, the estimated outcome is always positive. The assumption of Homoscedasticity mentioned earlier in the linear regression model can be addressed in a Poisson model since it allows the variance of the residuals to increase as the mean increases. Also, the problems with the non-normal distribution and heteroscedasticity can be addressed simply by using the Poisson model. If a discrete variable such as AADT termed as ‘y’ is assumed to be Poisson distributed, with a mean annual daily traffic, $\mu > 0$, over all sites and unit time period, then the probability density function (PDF) of 'y' is given by (Cameron and Trivedi, 1998):

$$p\ (y \mid \mu) = \frac{\exp(-\mu)\mu^{y}}{y!}, \quad y = 0, 1, 2, \ldots. \qquad [2.5]$$

$$\mu = E\ (y \mid \mu) = \exp\ (X\beta) \qquad [2.6]$$

Where,

$(y \mid \mu) \sim \text{Poisson}\ (\mu)$

X is a matrix of covariates.

The constraint in using the Poisson model is that the model assumes that the mean equals the variance (equidispersion), which is violated in many practical situations (Long, 1997). The violation of this assumption called overdispersion leads to a fairly different model called the Negative Binomial model.

$$E\ (y \mid \mu) = \text{Var}\ (y \mid \mu) = \mu \qquad [2.7]$$

Negative Binomial Models:

According to Gourieroux et al. (1984), the Poisson regression estimates could be consistent with a correctly specified mean structure but inefficiency exists with overdispersion and also a downwardly biased stand errors results in large z-values that may over estimate the significance of the variables (Cameron and Trivedi, 1986) as cited in Long (1997). Various Poisson mixture models to account for overdispersion have been implemented and used but Negative Binomial models (traditionally characterized as a Poisson-Gamma mixture model) addresses overdispersion using a simple mathematical representation of means and variances by incorporating a random component (Hilbe, 2007).

The conditional mean of y: $\mu = \exp(X\beta)$ is replaced by a random variable $\tilde{\mu}$ (Long, 1997) as follows

$\tilde{\mu} = \exp(X\beta + \varepsilon)$

$$\tilde{\mu} = \exp(X\beta).\exp(\varepsilon) = \mu.\exp(\varepsilon)\ \tilde{\mu} \qquad [2.8]$$

The unobserved heterogeneity in μ is introduced by ε. For convenience, it is assumed that $\exp(\varepsilon)$ is independent and Gamma distributed with unit mean and variance ν. If the PDF of $\exp(\varepsilon)$ is assumed to be $g(\varepsilon)$ then

$$f\ (y) = \int_0^\infty p(y|\mu).g(\varepsilon)d(\varepsilon) \qquad [2.9]$$

Various Poisson mixture models can be derived depending on the parameters imposed on the function $g(\varepsilon)$. In the equation 2.9 the first term $p(y|\mu)$ represents the Poisson function and the second term $g(\varepsilon)$ represents the gamma function (ν, ν) thus making the Poisson – Gamma mixture model, Negative Binomial (μ, ν). The PDF of Negative Binomial (μ, ν) is given by (Hilbe, 2007):

$$p(y, v) = \frac{\Gamma(y+\frac{1}{\alpha})}{\Gamma(y+1)\Gamma\frac{1}{\alpha}}\left(\frac{1}{1+\alpha\mu}\right)^{1/\alpha}\left(\frac{\alpha\mu}{1+\alpha\mu}\right)^{y} \qquad [2.10]$$

The mean and variance of the Negative Binomial variable is given by

$$E(y) = \mu;\ V(y) = \mu + \alpha\mu^2 \qquad [2.11]$$

The term α is called the dispersion (over) parameter of the Poisson-Gamma (or Negative Binomial) distribution. Also, as α value approaches zero the distribution becomes Poisson distribution (mean equals variance). Both the Poisson and the Negative Binomial models address the common problems of count data, heteroscedasticity (allows increase in variance as the mean) and protects the non negativity and discreteness of count data without having to transform them to a different scale. While Poisson models do not account for overdispersion, the Negative Binomial (a mixture of Poisson – Gamma) models accounts for the overdispersion (variance greater than mean). Both the models are widely used for count data though there are some limitations to specific cases. A more detailed discussion on count models can be found at Long, 1997; Cameron and Trivedi, 1998; and Hilbe, 2007.

2.7.3 Generalized Linear Models (GLMs)

GLMs are a class of statistical models that are obtained by a natural generalization of standard linear models (McCullagh and Nelder, 1989). Often categorized as extensions to the standard linear regression, GLMs can incorporate various response outcomes such as count, binary, proportions and positive valued continuous variables (Hilbe, 1994).

Scaling problems have been greatly reduced with the introduction of GLMs. This is one of the reasons of increased momentum in the use of GLMs in statistics. Moreover, the normality and constant variance of errors are no longer required (McCullagh and Nelder, 1989). A typical GLM is characterized by a random component 'Y' with a distribution belonging to the exponential family, a systematic component with covariates and parameters that produces a linear predictor 'η' and a canonical link function that connects the two components. The link function is the key of a GLM that can take any monotonic differentiable function. It explains how the linear predictor η is related to the mean of expected response μ, given by

$$\eta = \sum X\beta \; ; \; \mu = g(\eta^{-1}) = E(Y) \qquad [2.12]$$

It is particularly useful when transformations are needed and the expected mean can take only few values on the real line without losing the originality of the data. For instance, a Poisson count model (where mean can only take positive integers) uses a link function $\eta = \log \mu$, where the inverse function is $\mu = e^{\eta}$. Here the additive effects of η are converted to multiplicative effects where μ can never be negative. Several goodness of fit statistics are used for model selection in the GLM method, such as, likelihood ratio, Akaike Information Criterion (AIC) and corrected AIC (AICC).

Like any other model in the GLM, it is assumed that the variables are independently and identically distributed and the models are based on maximum likelihood theory for independent observations (McCullagh and Nelder, 1989). GLM might give inaccurate parameter estimates if the observations are correlated. It is particularly not suited for repeated measurements where correlation among the observed values is predominant.

2.7.4 Generalized Estimating Equations (GEE)

Liang and Zeger (1986) developed Generalized Estimating Equations (GEEs) as an extension to the GLMs to analyze longitudinal data. The GEE method is based on quasilikelihood theory (Wedderburn, 1974) and do not make any assumptions regarding the distributions of response observations. Quasilikelihood methods allow calculation of parameter estimates exclusively by specifying the mean and variance of the observations rather than specifications originating from single-parameter exponential family distributions (Hilbe, 2007). As the repeated measures are expected to be correlated, the GEE method requires specification of one of the working correlations structures available in most of the standard statistical software namely: independent, exchangeable, autoregressive, stationary, non-stationary and unstructured correlations structures. For a more detailed discussion on GEE models readers may want to read Hardin and Hilbe (2003).

Since the GEE methods uses quasilikelihood method, AIC statistic derived from likelihood theory cannot be applied directly (Pan, 2001a). Pan (2001a, b) developed two statistics namely Quasilikelihood under independence model criterion (QIC) and corrected QIC (QICC) respectively. While QIC is used to select the best correlation structure, QICC is used to select the best subset of model variables for a particular correlation structure (Hilbe, 2007). GEEs are preferred to GLMs even for independent observations as the former models are based on quasilikelihood methods that are considered to be better than the likelihood based methods used in the GLMs.

2.8 Limitations of Past Research and Need for Current Research

The traditional FSM process is a serial step by step aggregate approach in which transfer of errors is inevitable. It cannot address all aspects of demand jointly and in the presence of congestion. Though feedback loops are employed to address induced demand, the iterative processes might not give single convergent solutions. Combined travel demand models address some of the problems associated in the FSM process. However, being an aggregate approach by itself, combined models carry the basic assumptions and thus the drawbacks associated with the FSM. Activity based models employ sound social, behavioral and theoretical techniques. However, they are relatively new modeling methods that did not gained expected momentum. This is because of the enormous amount of data needed, computational complexities involved, and lack of familiarity of its concepts to a normal modeler. Also, modeling each and every individual’s daily tours and travel choices is very hard and difficult.

On the other hand, several state and local DOTs and planning organizations estimate AADT on roads from available short term traffic counts by applying seasonal and daily factors with continuous traffic counts as a reference. However there is a large difference between the available traffic counts and the number of road links in a huge transportation network. It is practically not feasible to collect traffic volumes on all the roads. With very few number of traffic counts available, estimation of AADT results in significant errors and uncertainties.

Use of AADT estimation models in urban travel demand estimation is still rudimentary in some aspects. Several AADT estimation models were developed in the past, but they are limited in the scope by several ways. While a few conducted time series analysis using historic traffic counts, a few others developed models using various statistical models for rural, county, non-state roads and other functional classes. They are also limited by the type of spatial data used (limited demographic and employment data were used to develop models).

Overall, none of the past research attempted to directly estimate traffic on the road network links by spatially capturing the characteristics of travel demand such as demographic, socio-economic, and land use data. Those that attempted did not consider the effect of spatial proximity or integrating data using spatial weights that decrease with an increase in distance. Spatial dependence of road network links was also not tested in the previous research. Therefore, there is a need to develop, a simple and systematic methodology, to estimate link level travel bypassing the tedious FSM process. Such a methodology should use scientific principles, spatial analytical methods, and statistical techniques for accurate estimation of traffic. It should be easy to adopt at any scale, size and level. A detailed discussion on the methodology developed is presented next.

CHAPTER III: METHODOLOGY

The primary objective of this research is to develop and assess models to estimate traffic demand on road network links for selected road functional classes based on on-network characteristics such as number of lanes and speed limit and off-network characteristics such as area type, demographic, socio-economic, and land use data in the vicinity of the network links. Vicinity for a selected road link in the network is the area within the generated spatial network buffer (or service area) around the road link (say, 1 mile).

The proposed methodology uses various spatial principles and statistical techniques to estimate traffic on a road link. It includes the following steps.

1. Select study area and links pertaining to each road functional class
2. Generate buffers around each link
3. Spatial overlay, data processing and integration
4. Develop statistical models to estimate traffic demand
    - Spatial proximity
    - Spatial weights
    - Statistical analysis
5. Validation of the developed models

Each of the above listed steps in the methodology is discussed next in detail.

3.1 Select Study Area and Links Pertaining to each Road Functional Class

A study area should be selected such that it represents various levels of urbanization. Travel activity varies based on the area type (CBD, Urban and Suburban), demographic, socio-economic, and land use characteristics. Traffic levels vary based on the type of road functional class. Higher functional classes are designed to carry high vehicular volumes at higher speeds. Likewise, major and minor arterials serve various levels of traffic and speeds. The higher the speed limit is, the higher is the throughput expected on the roadway section. Hence, it is necessary to consider various roadway types to develop and assess models.

Study locations should be selected in such a way that they are geographically distributed throughout the study area to capture various on- and off-network characteristics equally throughout the study area. For obtaining statistically meaningful estimates while developing models, sample size of the study locations selected should be sufficiently large. A sample size of 30 or more is generally considered as large enough for conducting statistical analysis. Maintaining large sample sizes among various road functional classes as well as area types helps to obtain statistically unbiased estimates.

3.2 Generate Buffers around each Link

People live and work at different locations in an urban region and access various roadways at various accessible distances as part of their daily travel. Twenty percent of driving trips are less than “one” mile whereas 87% of walking trips and 59% of biking trips are less than “one” mile (NHTS, 2009). It is expected that trips less than “half” or “one” mile are always walk or bike trips and hence a minimum buffer width of “one” mile is needed to capture driving trips. In a typical urban setting, land developments takes place along the transportation corridors and are expected to limit their access to around "five" miles. Depending on the roadway accessibility, a majority of users may access various roadways anywhere from "one" mile to "five" miles.

Various off-network characteristics data such as demographic, socio-economic and land use data influence the estimation of traffic on various roads. To quantitatively assess the influence of such characteristics, they should be captured and analyzed to estimate traffic. While it is evident that off-network characteristics data needs to be captured, there is no documented evidence on the distance at which the data has to be captured. Also, various functional classes of roadways collect traffic at varying distances. Collectors and local streets serve minor arterials and minor arterials serve major arterials. While smaller functional classes serve short trips or serve long trips as a medium to connect to higher functional classes, higher functional class roadways typically serve longer trips. Freeways and expressways typically have a wider reach than other functional classes. Therefore, it is necessary to capture data at various buffers widths to develop and assess models to estimate traffic. Hence, buffer widths of 1 mile, 1.5 mile, 2 mile, 3 mile, 4 mile, and 5 miles are considered to capture data to assess models to estimate traffic on different roadways.

3.2.1 Network Buffers (Service Area):

Three types of buffers, circular, polygon based network buffer and line based network buffers may be used to capture spatial data (Oliver et al., 2007). Circular buffers have been employed to capture the spatial information in many GIS based applications. Circular buffers capture the spatial information around a location with a constant Euclidean distance (generally described as a path of flying crow) covered by a radius 'r' irrespective of the accessibility to the road network. Circular buffers are a straight forward and a more crude form of spatial analysis.

A network buffer approach captures information based on accessibility of street network. In this case, the end points of a specified, say ‘y’ mile accessible network are joined to form an irregular polygon. Most off-the-shelf GIS software has tools available to generate network buffers using basic path search algorithms. Figure 3.1 shows an example 2-mile network buffer. Figure 3.2 below compares a 2-mile circular buffer with the corresponding polygon based road network buffer. It can be clearly observed from the figure that a circular buffer is spread over a larger area compared to a polygon based network buffer.

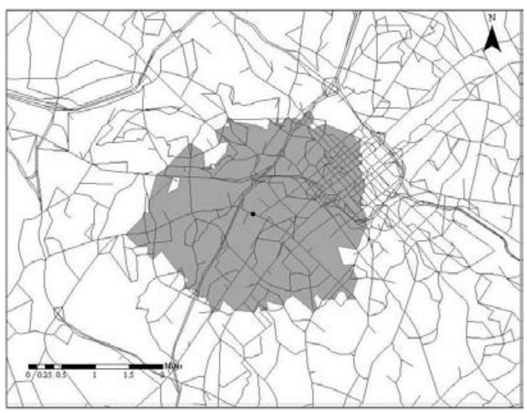


FIGURE 3.1: A 2-Mile Polygon Based Network Buffer

A line based road network buffer is similar to the polygon based network buffer but an additional constraint is laid on the network say, ‘x’ feet distance all along the road. The endpoints are not connected here but the buffer is drawn alongside the road with an ‘x’ feet distance up to ‘y’ miles.

It was felt that the polygon based network buffer approach is suitable for the present study because the accessible distances employed in the study are very large (up to 5 miles). It would be computationally very intensive and also difficult to determine the optimal line buffer distance ‘x’. Also, in the regional planning network, secondary road network information is typically not available. The secondary road information, if considered to generate line based network buffer, will cover most of the buffer area generated using the polygon based approach.

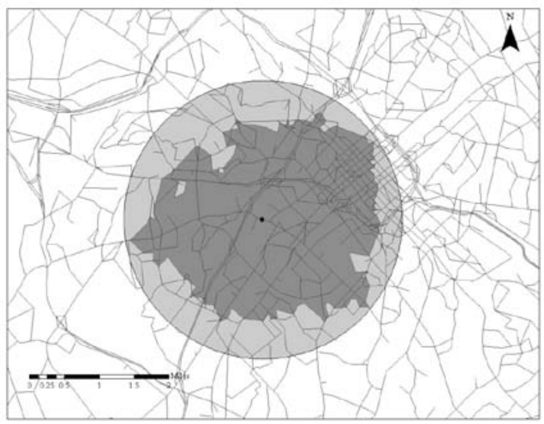


FIGURE 3.2: Comparison of Circular and a Polygon based Network Buffer

3.3 Spatial Overlay, Data Processing and Integration

In this step, the generated spatial buffers are overlaid on demographic, socio-economic and land use data to capture and extract data for processing and model development. Figure 3.3 shows an overlay of a network spatial buffer on TAZ layer with demographic and socio-economic information.

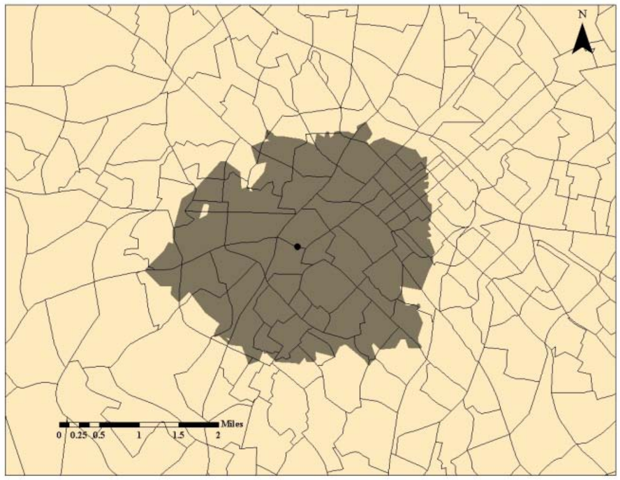


FIGURE 3.3: Spatial Network Buffer Overlaid on a TAZ Layer

Spatial buffers are typically spread over multiple TAZ’s or land use polygons. As an example, Figure 3.4 shows a spatial network buffer overlaid on the TAZ layer with TAZs numbered 1, 2, 3, .., and 8. The attributes of the spatial buffer are calculated based on the degree of overlap of the TAZs.

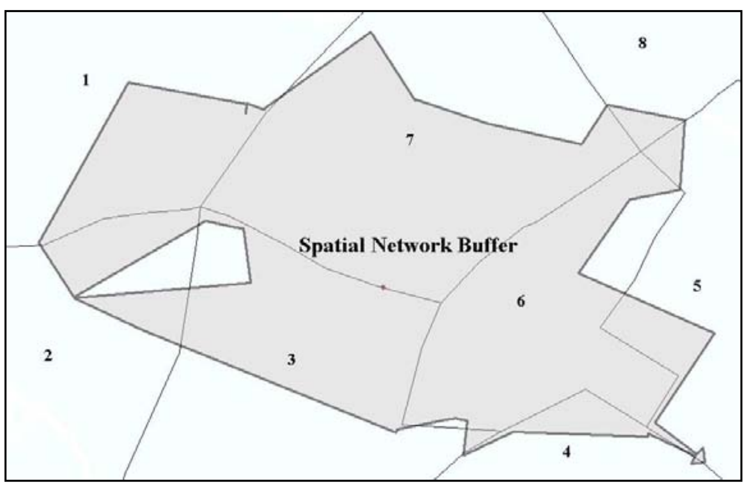


FIGURE 3.4: TAZs and Overlay of Spatial Network Buffer

3.3.1 Computation of Demographic and Socio-economic related Characteristics within a Spatial Buffer

The demographic and socio-economic characteristics (excluding mean income) are assumed to be distributed uniformly throughout a TAZ. A weighted aerial interpolation method was used to calculate the attributes. Based on this method, the ratio of the area of the TAZ in the buffer zone to the area of the TAZ gives the proportion of the attribute data (example, population) in the portion of the TAZ in the buffer zone. Summing data from portions of these TAZs gives the data within the service area or spatial buffer. As an example, the population in the generated spatial buffer is estimated using the following equation.

$$\mathrm{P_i} = \sum_j \frac{\mathrm{A_{j,i}}}{\mathrm{A_j}} * \mathrm{P_j} \qquad [3.1]$$

where,

P_i = Population of a buffer zone,

P_j = Population in a TAZ “j”,

$A_{j,i}$ = Area of a TAZ “j” in the buffer zone, and,

A_j = Area of a TAZ “j”.

Similar equations are developed and used to compute other attributes of the TAZ, such as, number of households, household population, number of employees of various types, and number of pupils enrolled in private and public kindergarten, elementary, middle and high schools and colleges and universities.

3.3.2 Computation of Mean Income within a Spatial Buffer

Since mean income is an average income of the household in the spatial buffer, it cannot be summed based on the ratio of area representing the TAZ. It is computed as a function of number of households in a TAZ in the spatial buffer and mean income of the TAZ. The computation of mean income of the spatial buffer is mathematically expressed as follows.

$$I_i = \frac{\sum_j H_{j,i} * I_j}{\sum_j H_{j,i}} \qquad [3.2]$$

where,

I_i = Mean income of the spatial buffer,

I_j = Mean income in a TAZ "j", and,

$H_{j,i}$ = Number of households in TAZ "j" in the spatial buffer "i".

3.3.3 Computation of Land Use Area within a Spatial Buffer

Generated spatial buffers are overlaid on the Land use polygon data to extract the area of each land use category in each buffer. Once the area of each land use in a spatial buffer are calculated, pivot tables are used to sum the areas of each land use type and are presented in separate columns using equation 3.3 below. Figure 3.5 below shows a network buffer overlaid on a land use zoning layer.

$$A_{lu} = \sum a_{lu} \qquad [3.3]$$

where,

A_{lu} = Total area of a land use type in a spatial buffer, and,

a_{lu} = Area of a type of land use zone in the spatial buffer.

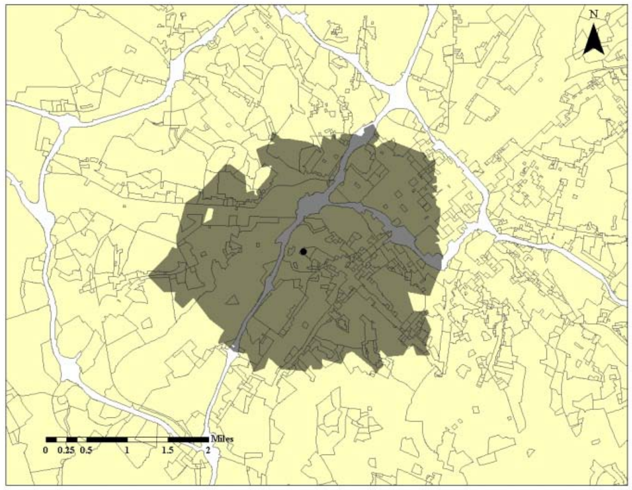


FIGURE 3.5: Network Buffer Overlaid on a Land use Zoning Layer

3.4 Develop Statistical Models to Estimate Traffic Demand

Two different methods are adopted to develop models in the present study. While the first one is based on spatial proximity, the second one is based on spatial weights that decrease with an increase in distance. They are discussed next in detail.

3.4.1 Spatial Proximity

The rationale behind developing models based on spatial proximity is to identify the ideal buffer width to capture and extract spatial data. In this case, models are developed for each spatially accessible network distance from the selected link or the traffic count location. In this research, spatial information is captured within 1.0 mile, 1.5 mile, 2 mile, 3 mile, 4 mile, and 5 mile accessible network distances from the study location. Models are then developed for each individual buffer distance. In other words, a 2 mile road network buffer would capture all the spatial information between the study location and the boundary line of the corresponding 2 mile network buffer. The best model from the above six models was selected based on goodness of fit and variable coefficients.

3.4.2 Spatial Weights

The rationale behind developing models based on spatial weights is to capture the spatial variations in the network level planning variables and land use data in the process of estimating travel demand. Geographically weighted regression models have been employed previously in many applications for capturing the variations in the local level attributes. It revolves around the observation points located at various distances in space. The proposed spatial weight method uses concepts from such models.

In a spatial weighting method, network buffers are generated in such a way that the spatial information is captured between the buffers. The spatial information is captured from 0 to 1 mile; 1 to 1.5 mile; 1.5 to 2 mile; 2 to 3 mile; 3 to 4 mile and 4 to 5 mile distances.

A distance decay function is used to apply weights for data captured from different spatial buffers. As discussed earlier in the GWR approach the sphere of influence or the traffic supply density is assumed to decrease as the distance from the point of interest increases. Figure 3.6 below (drawn for illustration) depict the decrease in traffic supply from an area around a point in a link with an increase in distance or travel time. Varying intensities of color is used to represent the change in traffic supply.

The strength of the method vests in selecting the critical accessible distance for a roadway. The sphere of influence or the area of attraction decreases from top to bottom in the hierarchical functional classes (freeways/expressways being on the top and minor thoroughfare in this case at bottom). A freeway/expressway attracts trips or traffic from a larger surrounding area from several lower level functional classes (major and minor thoroughfares and arterials). On the other hand, major and minor thoroughfares collect traffic from hierarchically lower level functional classes from a comparatively smaller surrounding area in the same order. So the weighting functions and bandwidths for such models based on functional class needs to be carefully derived based on the land use, demographic, socio-economic and employment characteristics in the vicinity of particular links.

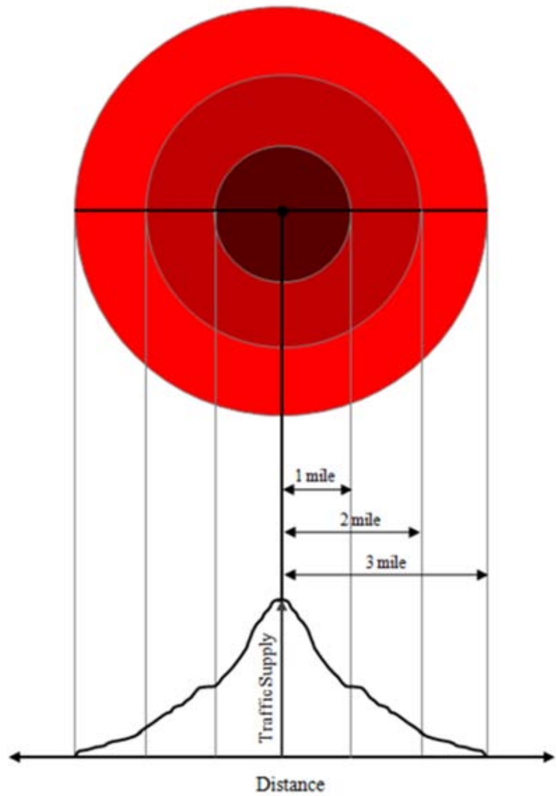


FIGURE 3.6: Distance Vs Traffic Supply

In this research, the weights for each bandwidth of network buffers were applied based on the inverse of the square of the network distance and corresponding proportion of the total. Figure 3.7 depicts the bandwidths used and the weighting represented by the darkness of the color. The darker the color, the higher the weight for a bandwidth.

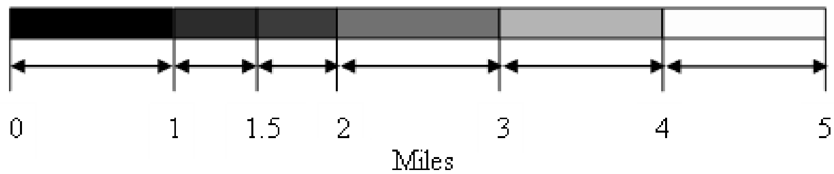


FIGURE 3.7: Bandwidths indicating the Intensity of Weights

If $b_{i\text{-}j}$ is the bandwidth between 'i' miles and 'j' miles, the weight applied to that bandwidth (i-j) is

$$\text{Weight of i-j, } W_{i\text{-}j} = \frac{\left(1/_{j^2}\right)}{\Sigma\left(1/_{j^2}\right)} \qquad [3.4]$$

The idea is to keep the total weight equal to "1" (100%). As an example, Table 3.1 below shows the weights (in proportions) for each bandwidth for a five mile accessible distance. As stated before, a 5 mile accessible distance was considered for all road functional classes combined and freeways/expressways. Using the above formula, $W_{0\text{-}1}$ is 0.52 (52%) for bandwidth of "0" to "1" mile, $W_{1\text{-}1.5}$ is 0.23 (23%) for bandwidth of "1" to "1.5" miles, $W_{1.5\text{-}2}$ is 0.13 (13%) for bandwidth of "1.5" to "2" miles, $W_{2\text{-}3}$ is 0.06 (6%) for bandwidth of "2" to "3" miles, $W_{3\text{-}4}$ is 0.03 (3%) for bandwidth of "3" to "4" miles, and $W_{4\text{-}5}$ is 0.02 (2%) for bandwidth of "4" to "5" miles. The sum of all weights is equal to 1.00 (100%).

TABLE 3.1: Weights in each Bandwidth for a Five Mile Accessible Network Distance Considered

Bandwidth (miles)		$1/j^2$	W_{i-j}
i	j		
0	1	1.00	0.52
1	1.5	0.44	0.23
1.5	2	0.25	0.13
2	3	0.11	0.06
3	4	0.06	0.03
4	5	0.04	0.02
Total		1.91	1.00

Likewise, weights were calculated for major thoroughfares and minor thoroughfares for an accessible network distance of “3” miles and “2” miles respectively without changing the bandwidth sizes.

Sphere of influence for a freeway/expressway, major thoroughfare and minor thoroughfare for corresponding accessible distances considered in this research are shown in Figure 3.8 (a), (b) and (c) below.

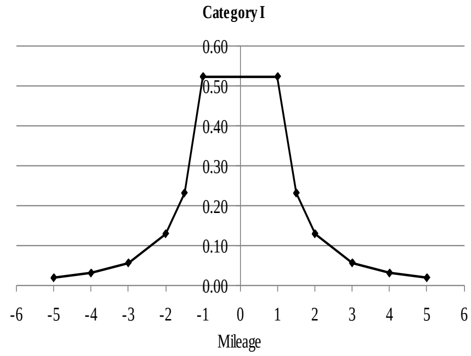
The weights are applied to the data captured in each corresponding bandwidth and summed to obtain the final data set. Mathematical representation of the spatial weighting applied on various off-network spatial variables data is given below.

$$X_a = \sum(X_{a\,(i-j)} \times W_{i-j}) \qquad [3.5]$$

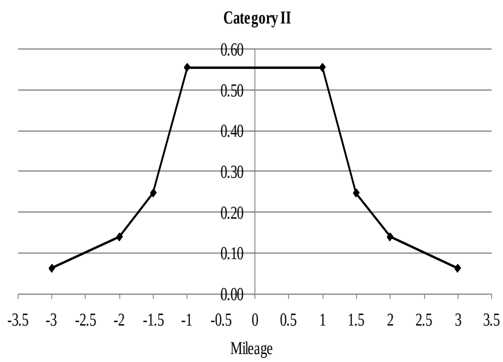
where,

X_a is the off-network characteristics variable with the combined weighted data, and,

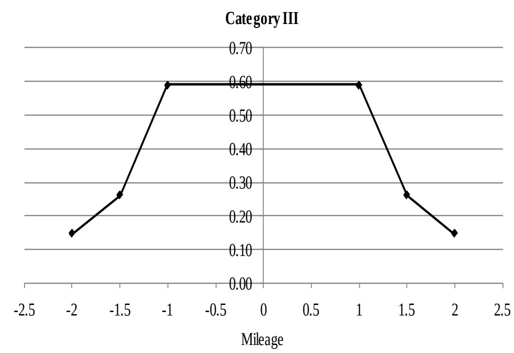
$X_{a\,(i-j)}$ is the off-network characteristics variable corresponding to the bandwidth i-j.



(a). Freeway/Expressway



(b). Major Thoroughfare



(c). Minor Thoroughfare

FIGURE 3.8: Sphere of Influence around Various Road Functional Classes

3.4.3 Statistical Analysis

The data gathered, processed and cleaned in the previous steps is then used to test correlation between independent variables, conduct statistical analysis and identify parameters significant to estimate AADT (dependent variable).

Correlation and Multicollinearity:

Prior to developing the models, correlations among the independent variables are studied using Pearson correlation coefficient. Pearson correlation coefficient is the ratio of covariance to the product of standard deviations. It represents the strength of association or linear relationship between two variables. The higher the absolute value the higher is the relationship. It ranges from -1 to +1, the two boundaries of perfect negative and positive correlations. The negative lowest (-1) is a strong negative correlation that indicates that an increase in a variable value decreases the other at the corresponding rate. The positive highest (+1) is a strong positive correlation that indicates an increase in a variable value with a decrease in the other variable at the corresponding rate. Table 3.2 below shows the Pearson correlation ranges and the corresponding association strength.

TABLE 3.2: Pearson Correlation Coefficient Range and the Associated Strength

Pearson Correlation Ranges		Strength of Correlation
-1.0	-0.7	Strong Negative
-0.7	-0.3	Weak Negative
-0.3	+0.3	Very little or No
+0.3	+0.7	Weak Positive
+0.7	+1.0	Strong Positive

Highly correlated predictor (independent) variables could produce significant overall p-values even when the variables do not produce an effect on the dependant variables or insignificant t-values because of high standard errors. This problem of unreliable results due to correlation among the independent variables is called as multicollinearity effect. It is recommended that the correlations among the (independent) variables considered for developing the models are in the range (-0.3, +0.3). A rational approach should be adopted in identifying and eliminating variables with inter-correlations. A more detailed explanation of variables retained or eliminated is given in the subsequent chapters.

Development of Models:

Determination of optimal buffer size to be used to capture spatial data for estimating travel demand typically is based on the road functional class type. All the data is classified based on the functional class (freeway/expressway, major and minor thoroughfares in a regional network model). Models are developed for each road functional class, as a function of demographic, socio-economic and land use characteristics. Generalized Estimating Equations (GEE) are used to develop and assess the models and identify significant predictors with comparatively better predicting capability.

"Spatial proximity" and "spatial weighting" models are developed separately. While the former models are developed based on the characteristics of data close to the roadway links considered, the later models applies various weighting factors based on the distance from the roadway links.

In each of the methods after conducting the correlation analysis and retaining the most relevant independent variables with minimal inter-correlations, GEE models are developed and tested for various probability distributions. Further, independent variables with a significance value less than 90 percent (P-Value greater than 0.1) were eliminated in each of the models. The models are recalibrated and the process is repeated until all the independent variables had the specified level of significance. The final model consists of variables with minimal inter-correlations (±0.3) and a significance value less than 90 percent.

3.5 Validation of the Developed Models

Model validation is designed in two steps. Firstly, percent differences between the actual traffic counts (AADT) and estimated AADT are calculated. The average of all the absolute percent differences is then calculated. Secondly, model validation is carried out using Pearson's Chi Square (X^2) Statistic. The mathematical formula to compute Pearson's Chi Square (X^2) Statistic (McCullagh and Nelder, 1989, Page No. 34) is shown next.

$$X^2 = \sum (y - \mu)^2 / V(\mu) \qquad [3.6]$$

$$V(\mu) = \mu + k\mu^2 \qquad [3.7]$$

where,

X^2 is the chi-square statistic,

y is the observed mean,

μ is the predicted mean,

V (μ) is the variance, and,

k is the dispersion parameter of the Negative Binomial model.

Also, known as the modified or concocted chi-square (X_m^2), the above equation is used to calculate the Chi-Square Statistic (CSS) of the observed and predicted values for model validation. The CSS thus calculated is compared with Critical-CSS at 99% confidence level. If CSS is less than Critical-CSS, the test is satisfied and the predicted values are statistically close to the observed values.

CHAPTER IV: DATA COMPILATION

This chapter discusses study area, selection of locations and various data elements to develop models. It also describes data elements and the data collected for this research.

4.1 Study Area and Locations/Sites

Mecklenburg County in the state of North Carolina (NC) is selected as study area. It is located in the southern part of the state of North Carolina. The total area (land) of the County is 526.28 square miles with a population of approximately nine hundred thousand distributed at a rate of 1322.2 persons per square mile according to US Census Bureau quick facts (US Census, 2008). It consists of city of Charlotte and towns such as Mathews, Pineville, Mint Hill, Davidson, Cornelius and Huntersville. However, major area of the Mecklenburg County is occupied by the city of Charlotte. Charlotte is the financial capital of North Carolina and a rapidly growing urbanized area with a population of close to 700,000 (US Census Estimates, 2008). It ranks top in terms of population, traffic congestion and urbanization in the state of North Carolina. According to the 2008 US Census population estimates, the city is one of the fastest growing cities and one among the twenty most populous cities in the United States.

The city's downtown attracts large number of traffic from within the Mecklenburg and several adjacent counties. The Mecklenburg County highway network is a radial design with a spoke like extension of roads from the Central Business District (CBD) to the outskirts. Several major and minor arterials serve the traffic generated from various origins and destinations. There are two routes that serve as inner and outer loops (ring roads) of the region. The inner loop consists of Interstate-277 and Interstate-77 surrounding downtown Charlotte, whereas the outer loop is served by Interstate-485 and Interstate-85 connecting major suburban centers. Besides these there is another loop Route-4. It is served mostly by four lane roads and partly with a limited access parkway and I-85. It surrounds uptown Charlotte at a radius of about 4 miles between inner and outer loops. This is essentially the basic structure of the highway network of the study region that also helps to distinguish various areas types of the same. The Charlotte – Mecklenburg highway network (approximately 1,100 TAZ's) is extracted from the regional network covering 13 counties (~3000 TAZ's) considering availability and consistency of data such as land use zoning. Figure 4.1 below gives a map with road network links (approximately, 7,375) in Mecklenburg County extracted from the 2005 regional travel model network.

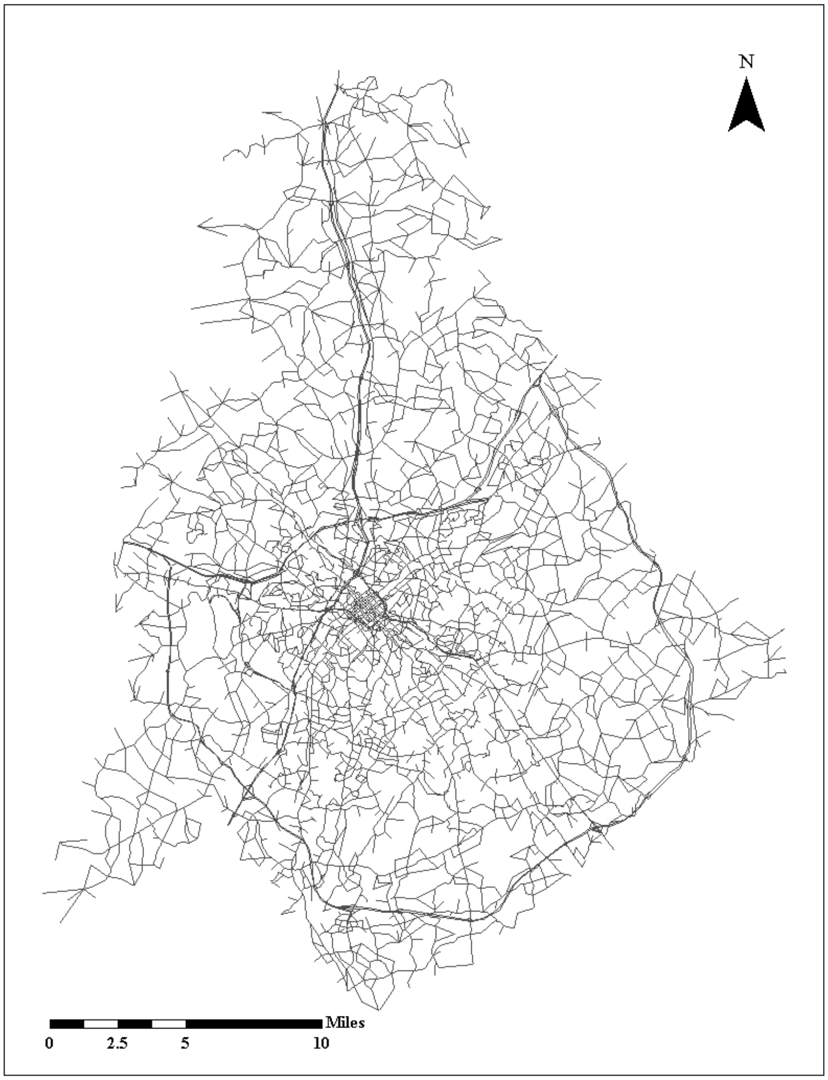


FIGURE 4.1: 2005 Mecklenburg County Highway Network

Geographically distributed study locations are selected based on road functional class and area type to develop models based on road functional class. Three area types (CBD, Urban and Suburban) were considered in the study.

The following road functional classes are considered in the study:

- Freeway or Expressway
- Major thoroughfares
- Minor thoroughfares

As discussed earlier the three loops (ring roads) in the study region are used to select links from various area types. The links inside and around the inner loop were considered as area type “CBD”. Links away from inner loop and around Route 4 are considered as “Urban” and links away from Route 4 and around the present outer loop and proposed extension of Interstate-485 are considered as “Suburban”. The Charlotte Department of Transportation (CDOT) uses population and employment densities to classify area types. However, from a general observation of a map representing area types using the CDOT approach, several urban areas are represented as suburban and vice versa. To be consistent and to be representative of any metropolitan setting nationwide, the above method is used to classify area types in this dissertation. Figure 4.2 below shows the selected study locations by area type for major thoroughfares from 2005 base year Mecklenburg region highway network for Charlotte metropolitan area for illustration.

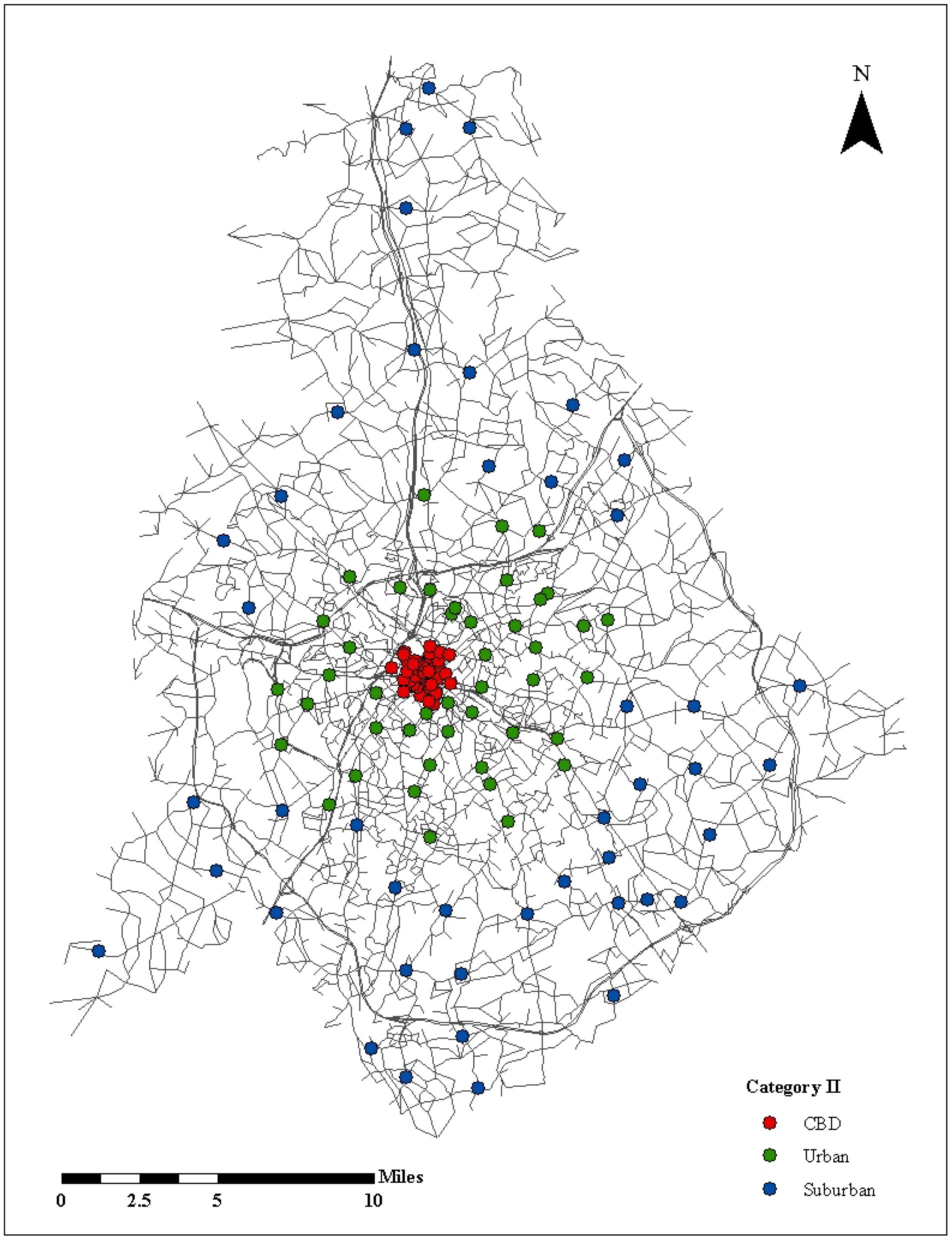


FIGURE 4.2: 2005 Study Locations for Major Thoroughfares by Area Type

The initial data consisted of nine datasets (a combination of three roadway types and three area types). The sample size of each data set consisted of not less than 30 links. More study links were considered where ever possible. For instance freeway/expressway type links in the CBD area were very few and hence only 33 links with traffic counts available were considered. Also the sample locations/links are considered in such a way that they are geographically distributed over the study area. Table 4.1 below shows the number of links considered for each road functional class and area types.

TABLE 4.1: Study Links in Each Road Functional Class and Area Type

Roadway / Area Types	CBD	Urban	Suburban	All Area Types
Freeways/Expressways	30	37	35	102
Major Thoroughfare	35	38	38	111
Minor Thoroughfare	29	33	36	98
All Roadway Types	94	108	109	311

4.2 Data Collection

The regional network and travel demand model data is obtained from the CDOT. Highway network and TAZ level planning variables data of Mecklenburg County is extracted from the regional travel demand model in a GIS format. Local or regional transportation planning objectives are considered while defining the TAZs whereas the geographical disintegration of census data typically at block group or zonal level is carried out with broader objectives. TAZ data is preferred in this research to the census block group data. Also the census data is available for the year 2000 at the time of this research while other data are for year 2005.

For consistency all the traffic, network, demographic, socio-economic, employment, and land use data for the year 2005 is used as a basis for data collection. Though year 2009 base year data is available 2005 base year data is considered to account for the effect of recession (started in late 2007 and intensified from early 2008) and LYNX blue line light rail transit that started in November 2007.

Network characteristics such as number of lanes, and speed limit of the study links and the corresponding spatially dependant links such as upstream and downstream links and upstream and downstream cross street links were collected and stored in each data set. Spatial network buffers of varying sizes were generated for each study link to capture the demographic, socio-economic, and land use characteristics in the vicinity of the study link. Finally three databases were generated by combining the data sets of each road functional class. The final dataset for each road functional class contains three different area types. The area type is included as an independent variable in the final database.

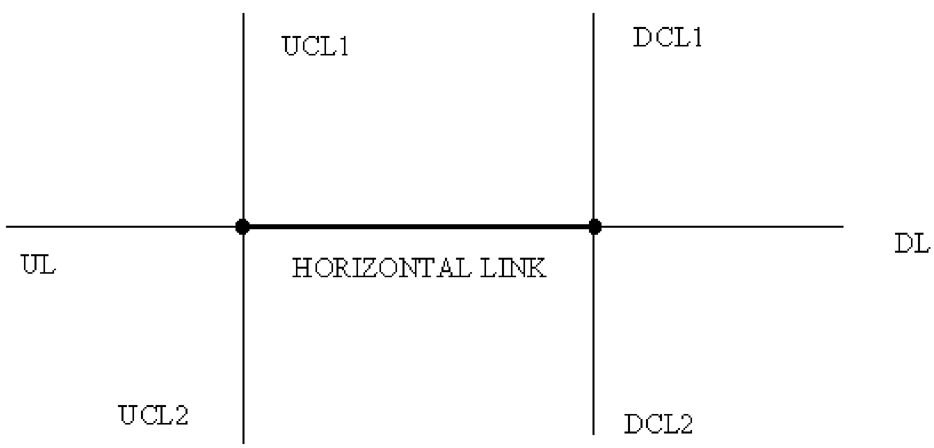
4.2.1 Traffic Volumes

AADT volumes are obtained from the North Carolina Department of Transportation (NCDOT) for the years 2002 to 2008 (that would be used for calibration and validation of the models in the later part). In addition average annual weekday traffic (AAWT) from short term traffic counts conducted by Charlotte Department of Transportation (CDOT) during the years 2004, 2005 and 2006 were obtained. Appropriate adjustment factors provided by the CDOT were used to calculate AADT. 2005 traffic counts were used where ever possible and appropriate adjustments were made for 2004 and 2006 data where 2005 data is not available.

4.2.2 Network Characteristics Data

Number of lanes, speed limit and various road functional classes discussed earlier were collected for all the study links. In addition, number of lanes and speed limits of upstream and downstream links and upstream and downstream cross street links were also collected. It was expected that traffic on study links are subjected to spatial dependency of network and hence the characteristics of the corresponding upstream and downstream network links were also included in the analysis as independent variables.

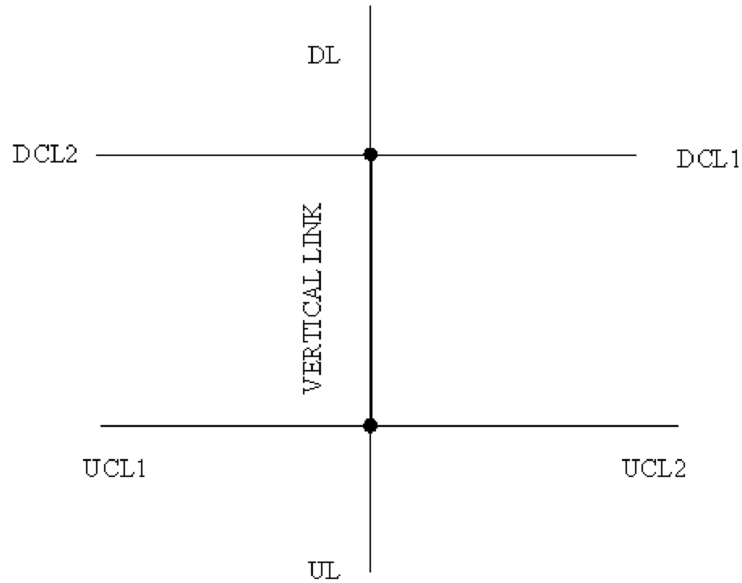
Roadway running towards or away from the downtown is considered as vertical and the perpendicular roads are considered as horizontal. For consistency, links on the left are considered as cross-street link 1 and links to the right are considered as cross-street link 2. It was felt that cross-street links cannot be considered as one due to the differences observed in the number of lanes and speed limit between the two cross street links (left and right of the study link) while collecting the cross-street information. When T-intersections are encountered, there will be only one cross-street. So, it was felt more appropriate to consider the two cross-street links both upstream and downstream as two separate entities. For T-intersections, the cross-street variable values are taken as 'zeros'. Figure 4.3 below shows typical links in a highway network and the corresponding spatially dependent links.



Horizontal Links

Left – Upstream (US); Right – Downstream (DS);

Above – UCL1, DCL1; Below – UCL2, DCL2



Vertical Links

Above – Upstream (US); Below – Downstream (DS)

Left – UCL1, DCL1; Right – UCL2, DCL2.

FIGURE 4.3: Illustration of Spatially Dependent Links

4.2.3 Socio – Demographic and Employment Data

Demographic characteristics such as number of households, population, household population; social characteristics such as mean income; and employment by type and total employment based on Standard Industry Classification (SIC), pupil enrolled in kindergarten, elementary, middle and high schools, colleges and universities were considered in the analysis.

4.2.4 Land use Data

Land use zoning characteristics such as single family and multi-family, urban residential, office, business, institutional, industrial, commercial/retail and other land uses expected to be significant attractors and generators of trips were included in the analysis.

Zoning regulations are adopted by the Charlotte-Mecklenburg planning department to promote and protect similar types of development in support of public policy objectives (City of Charlotte), each land use codified by the City of Charlotte Zoning Ordinance (2010). Such land use zoning allows development of land uses of certain category coded in the text of the ordinances. The zoning districts are classified as residential and non-residential land uses depending on the purpose served by each. Residential land uses include single family housing, multifamily housing, urban residential and primarily residential mixed use land uses. Other land uses such as industrial, business, commercial center, office, institutional, research district and other mixed use districts are also considered. Land uses such as rural, mobile residential, planned unit development and neighborhood service district are small in area and are under mostly residential category. Table 4.2 below gives a description of each land use district described by the City of Charlotte Zoning Ordinance in general. Table 4.3 below summarizes the characteristics of all the data used in the analysis. All the demographic and socio-economic variables are converted to thousands (for example, thousand population, and employees in thousands) and the land use variables are converted to thousand square ft. All the calculations are carried out using SQL Server 2005®.

TABLE 4.2: Description of various Land use Characteristics

Land Use Characteristics	Description
Single Family Residential Area	3 to 8 dwelling units/acre
Multi Family Residential Area	12 – 43 dwelling units/acre
Urban Residential Area	Encouraged for infill development and increase density near employment core
Office District	Establishments for office, institutional, and commercial operations not involving sale of merchandise
Industrial	Wholesale, manufacturing, processing and assembling units, transportation terminals and a broad variety of industrial operations
Research District	Higher end research, development and high technology manufacturing
Urban Residential - Commercial Area	Residential, retail outlets and offices conducive for higher density patterns
Institutional	Major educational, medical, government, cultural and religious and other institutions
Business	Neighborhood business to support residential areas to general and distributive businesses for retail merchandise and business parks
Mixed use District	Full range of housing types, and compatible non residential uses to provide goods, services and employment primarily for the planned community residents
Mobile Residential	Proper location and planning of manufactured homes and mobile home parks and subdivisions
Rural District	zones that are rural in nature
Manufactured House	Permits farms, manufactured homes, service buildings (laundry and recreational facilities), parks, greenways and detached dwellings
Commercial Center	Retail establishment or shopping center of larger than 70,000 square feet (approximately)
Neighborhood Service District	Retail and service activity needs of neighborhoods and accommodates mixed use existence
Innovative	Planned for development for specific public policy objectives
Planned Unit Development	Planned for development for specific public policy objectives
Mixed use Development	Zones that are planned and encouraged for mixed use developments
Right of Way	State owned right of ways of interstates, major and minor thoroughfares and any right of way for roads

TABLE 4.3: Characteristics of the Data

Network Characteristics	**Area Type**
Roadway Type / Category	CBD
# Lanes	Urban
Speed Limit	Suburban
Demographic	**Land Use Characteristics**
# Households	Single Family Residential Area
Total Population	Multi Family Residential Area
Household Population	Urban Residential Area
Group Quarters Population	Office District
Socio Economic	Industrial
Mean Income	Research District
# Manufacturing, Industrial, Warehouse, Transportation, Communication, and Utilities Employees	Urban Residential - Commercial Area
	Institutional
# Retail Employees	Business
# Highway Retail Employees	Mixed District
# Low Visitor Service Employees	Mobile Residential
# High Visitor Service Employees	Rural District
# Office and Government Employees	Manufactured House
# Bank Employees	Commercial Center
# School, College and University Employees	Neighborhood Service District
# Pupils Enrolled in Public or Private Kindergarten, Elementary and Middle Schools	Innovative
	Planned Unit Development
# Pupils Enrolled in Public or Private High Schools	Mixed use Development
# Pupils in Public or Private Colleges and Universities	Right of Way
Total Number of Employees (Sum of all Employees)	

CHAPTER V: MODELS BASED ON SPATIAL PROXIMITY

Buffers of width 1 mile, 1.5 mile, 2 mile, 3 mile, 4 mile and 5 miles were generated around each study link belonging to freeway/expressway, major thoroughfare and minor thoroughfare road functional classes. Demographic, socio-economic and land use data were then overlaid on the generated buffers. Network characteristics were also collected as discussed earlier. Data are then processed and integrated to generate databases for model development. An examination of correlation between independent variables and development of models for different buffer widths was then conducted. The models are developed for all road functional classes, freeways/expressways, major thoroughfares and minor thoroughfares, with and without considering network characteristics to estimate AADT for transportation planning and analysis purposes. The examination of correlation between independent variables followed by models developed using spatial proximity (individual spatial buffers) method is discussed next.

5.1 Correlation Matrices and Selection of Independent Variables

To avoid or minimize multicollinearity effect, a two tailed, bivariate Pearson correlation analysis is conducted among all the independent variables. A correlation matrix is generated in the SPSS® software. Independent variables are screened, such that, no final selected variables have correlations out of (-0.3, +0.3) range. Variable selection is carried out in a logical way. Among the four types of independent variables (network, demographic, socio-economic, and land use variables), the variables that are expected to be directly related to the AADT are included whereas the variables that are comparatively unexplainable are excluded in the case of correlation between the two. For example in the network characteristics data, more number of lanes allows more throughputs and hence is included. The posted speed limit facilitates higher thru traffic but highly fluctuates depending on the actual speeds and hence is considered only when not correlated with the number of lanes. Functional classification depends on the number of lanes and speed limit and hence is eliminated when found correlated with lanes or speed limit. Since, separate models were developed for each road functional class, it was felt that functional class is not a significant variable. Also, if the upstream or downstream number of lanes or posted speed limit is correlated with the selected link number of lanes or speed limit the upstream or downstream link variables are eliminated.

Likewise, population living in a service or buffer area can be easily quantified whereas, households and household population are difficult to track (there might be empty households, houses on rent and other). Population could account for the people living in the households and could be accounted for twice if both are considered. It was found that for almost all the spatial buffers (for all accessible distances considered) for all the road functional classes, number of households, household population and group quarter population are correlated with population. Likewise, if the population is related with any of the employee types they are eliminated. Also the “total number of employees” variable is calculated as a sum of employees in all the employment categories. Hence, if the total employment is correlated with employment types they are eliminated.

In the land use characteristics data, elimination is carried out on a case by case basis. The objective is to eliminate the effect of multicollinearity to the extent possible and to keep the number of variables in the final model precise. Population in an area depends on the type of land use such as single family and multifamily residential, urban residential, office, institutional and other land uses. So if population is correlated with any of these land uses they are eliminated. If they have no correlation with population but are correlated among the land use types care was taken in eliminating the land use that is insignificant in terms of traffic generation. Neighborhood service district, innovative, planned unit development, right of way land uses are less in area. If they are correlated with other land uses they are eliminated. They are included only when they are not correlated with any other socio-economic and land use variables. Common land uses that are representative of a typical urbanized area are retained in case of a correlation, such that the models developed could be used nationally. Multifamily land uses bear more population compared to single family that can be accounted for in population. So, if single family and multifamily land uses are correlated, single family land use is retained. Research district is retained if research district and employees and pupils enrolled in public and private colleges or universities are correlated. Oftentimes, it was observed that land use variables and “population” are correlated. Since all the final independent variables retained in each set of processed data have inter correlations in the +/- 0.3 range, any significant multicollinearity is ruled out.

The above process was repeated for all selected buffer widths for each type of road functional class considered in this research. The final set of independent variables (that are not correlated to each other) selected to develop the models for each buffer width for various road functional classes considered are listed tables 5.1, 5.2, 5.3, and 5.4.

TABLE 5.1: Variables Selected for each Buffer Width for All Road Functional Classes

Variable / Spatial Buffer	**1 Mile**	**1.5 Mile**	**2 Mile**	**3 Mile**	**4 Mile**	**5 Mile**
CBD	✓	✓	✓	✓	✓	✓
Urban	✓	✓	✓	✓	✓	✓
Suburban	✓	✓	✓	✓	✓	✓
Freeway/Expressway	✓	✓	✓	✓	✓	✓
Major Thoroughfare	✓	✓	✓	✓	✓	✓
Minor Thoroughfare	✓	✓	✓	✓	✓	✓
# Lanes	✓	✓	✓	✓	✓	✓
Upstream Cross Street Link 1 # Lanes	✓	✓	✓	✓	✓	✓
Downstream Cross Street Link 1 Speed Limit	✓	✓	✓	✓	✓	✓
Population	✓	✓	✓	✓	✓	✓
Mean Income			✓			
# Pupils Enrolled in Public or Private High Schools	✓	✓	✓	✓		
SingleFamily	✓	✓				
Industrial	✓					
Research District	✓	✓				
Institutional	✓	✓	✓	✓	✓	✓
Mobile Residential	✓	✓	✓			
Rural District			✓	✓	✓	✓
Manufactured House		✓	✓	✓	✓	✓
Commercial Center	✓	✓				
Neighborhood Service District	✓					
Innovative					✓	✓
Planned Unit Development	✓	✓		✓		
RightofWay	✓	✓				

TABLE 5.2: Variables Selected for each Buffer Width for Freeways/Expressways

Variable / Spatial Buffer	1 Mile	1.5 Mile	2 Mile	3 Mile	4 Mile	5 Mile
CBD	✓	✓	✓	✓	✓	✓
Urban	✓	✓	✓	✓	✓	✓
Suburban	✓	✓	✓	✓	✓	✓
# Lanes	✓	✓	✓	✓	✓	✓
Speed Limit	✓	✓	✓	✓	✓	✓
Upstream Cross Street Link 1 # Lanes	✓	✓	✓	✓	✓	✓
Upstream Cross Street Link 1 Speed Limit	✓	✓	✓	✓	✓	✓
Upstream Cross Street Link 2 # Lanes	✓	✓	✓	✓	✓	✓
Downstream Link Speed Limit	✓	✓	✓	✓	✓	✓
Downstream Cross Street Link 1 # Lanes	✓	✓	✓	✓	✓	✓
Downstream Cross Street Link 1 Speed Limit	✓	✓	✓	✓	✓	✓
Population	✓	✓	✓	✓	✓	✓
# Pupils Enrolled in Public or Private High Schools			✓			
Institutional	✓	✓	✓	✓	✓	✓
Mobile Residential	✓	✓	✓			
Rural District					✓	✓
Manufactured House		✓	✓	✓	✓	✓
Commercial Center	✓	✓				
Innovative	✓	✓				
Planned Unit Development	✓	✓	✓	✓		
RightofWay	✓					

TABLE 5.3: Variables Selected for each Buffer Width for Major Thoroughfares

Variable / Spatial Buffer	1 Mile	1.5 Mile	2 Mile	3 Mile	4 Mile	5 Mile
CBD	✓	✓	✓	✓	✓	✓
Urban	✓	✓	✓	✓	✓	✓
Suburban	✓	✓	✓	✓	✓	✓
# Lanes	✓	✓	✓	✓	✓	✓
Speed Limit	✓	✓	✓	✓	✓	✓
Upstream Link Speed Limit	✓	✓	✓	✓	✓	✓
Upstream Cross Street Link 1 # Lanes	✓	✓	✓	✓	✓	✓
Upstream Cross Street Link 2 Speed Limit	✓	✓	✓	✓	✓	✓
Downstream Cross Street Link 1 # Lanes	✓	✓	✓	✓	✓	✓
Downstream Cross Street Link 2 Speed Limit	✓	✓	✓	✓	✓	✓
Population	✓	✓	✓	✓	✓	✓
# Pupils Enrolled in Public or Private High Schools	✓	✓	✓	✓		
SingleFamily	✓	✓	✓			
OfficeDistrict	✓					
Industrial	✓					
Research District	✓	✓	✓			
Institutional	✓			✓	✓	✓
Mixed use District					✓	✓
Mobile Residential		✓	✓			
Rural District				✓	✓	✓
Manufactured House				✓	✓	✓
Commercial Center	✓					
Neighborhood Service District	✓	✓				✓
Innovative				✓	✓	✓
Planned Unit Development	✓	✓				
RightofWay	✓	✓		✓	✓	✓

TABLE 5.4: Variables Selected for each Buffer Width for Minor Thoroughfares

Variable / Spatial Buffer	1 Mile	1.5 Mile	2 Mile	3 Mile	4 Mile	5 Mile
CBD	✓	✓	✓	✓	✓	✓
Urban	✓	✓	✓	✓	✓	✓
Suburban	✓	✓	✓	✓	✓	✓
# Lanes	✓	✓	✓	✓	✓	✓
Upstream Link Speed Limit	✓	✓	✓	✓	✓	✓
Upstream Cross Street Link 1 # Lanes	✓	✓	✓	✓	✓	✓
Upstream Cross Street Link 2 Speed Limit	✓	✓	✓	✓	✓	✓
Downstream Link Speed Limit	✓	✓	✓	✓	✓	✓
Downstream Cross Street Link 1 # Lanes	✓	✓	✓	✓	✓	✓
Downstream Cross Street Link 2 Speed Limit	✓	✓	✓	✓	✓	✓
Population	✓	✓	✓	✓	✓	✓
Mean Income	✓					
# Pupils Enrolled in Public or Private High Schools	✓	✓	✓	✓		
SingleFamily	✓	✓	✓			
Industrial		✓				
Research District	✓	✓	✓		✓	✓
Institutional	✓	✓	✓	✓	✓	✓
Business						
Mobile Residential	✓	✓				
Rural District			✓	✓	✓	✓
Manufactured House				✓	✓	✓
Commercial Center		✓				
Neighborhood Service District	✓					
Innovative					✓	✓
Planned Unit Development				✓		
RightofWay		✓	✓	✓	✓	

The datasets thus retained were used to run the generalized estimating equations in the SPSS®. The independent variable is the thousand AADT, and all the other independent variables that are not correlated to each other such as network characteristics, demographic, socio-economic and land use characteristics (in thousands) are considered as covariates (that includes the categorical data in binary format). Statistical analyses and model development is discussed next.

5.2 Statistical Analyses and Assessment of Models to Estimate AADT

Initially general and stepwise Multiple Linear Regression (MLR) models were developed along with a log transformed MLR model. Results with a log transformed model sounded to be comparatively better than the simple MLR model based on goodness of fit statistics like R-Square, adjusted R-Square and Predicted Error Sum of

Squares (PRESS). This indicated that a non-linear relationship is predominant between the dependent and independent variables. Hence, Generalized Estimating Equation (GEE) models were developed using simple normal distribution, and, normal, Poisson and Negative Binomial distributions with log-links.

The Quasi Information Criteria (QIC), and the corrected QIC (QICC) are used as goodness of fit statistics. In general, QIC is used for selecting the correlation structure, whereas, QICC is used to select a model among a group of models. Normal and Normal with log-link have comparatively very high QIC and QICC values (fifty times larger than Poisson with log-link). Hence, models were developed using Poisson and Negative Binomial distribution with a log-link (typical count models). SPSS® is used to run the GEE analysis for developing the models.

Independent variables are further screened based on predicting power (significance levels) in the models. A 90% level of significance is used to develop the models. Thus independent variables with a P – Value greater than or equal to 0.1 were eliminated and the analysis is repeated. Wald Chi Square should be typically larger (greater than 1.0). The elimination process is continued till all the variables in the final model have a P – Value less than 0.1. Since, most of the variables were eliminated in the correlation analysis the final model is obtained typically by second or third run.

Since the data collected is not longitudinal in nature (AADT for the year 2005 only), correlation structures were not evaluated. Evaluation and selection of models is based on low QICC and difference between QIC and QICC (the lesser the difference, the better the model). The variables included and their coefficients were also examined to select the final model.

The models for all road functional classes, freeways/expressways, major thoroughfares, and minor thoroughfares, for all buffer widths considered are presented and discussed next. To facilitate planners to develop models for analysis as well as planning purposes, models with and without network characteristics were developed.

5.3 Models with Network Characteristics

Models were developed to estimate AADT for analysis purposes using the final variables obtained from the correlation analysis conducted among all the independent variables such as network characteristics, socio-economic and land use characteristics data. Models developed based on all and each individual road functional classes are presented in this section. Only the variables that contributed to the final model and their corresponding parameters were presented in the results tables. Variables that do not contribute to the final model were left blank in the tables. All land uses in the tables are in thousand square feet.

5.3.1 Models based on All Road Functional Classes

A summary of parameters and the corresponding goodness of fit statistics for each selected buffer width for Poisson and Negative Binomial distributions with log link were presented in Table 5.5 and Table 5.6, respectively.

It is evident from the goodness of fit statistics that the Negative Binomial with log link fits the model better than the Poisson with log link. While the Poisson based models have QIC and QICC around 1,900, the Negative Binomial models have QIC and QICC around 60 and 70. The QIC and QICC are relatively close to each other for all buffer widths examined in this research.

The Negative Binomial model based on one and a half mile buffer width has the lowest QIC and QICC. This model indicates that area type (urban), freeways/expressways, major thoroughfares, number of lanes, single family housing and the presence of manufactured house land use are significant variables in estimating AADT. It is expressed mathematically as follows.

$(\text{AADT in thousands})_{1.5\ \text{Mile}} = \text{Exp}\ (1.578 + 0.149 \times \text{Urban} + 2.117 \times \text{Freeway/Expressway} + 0.660 \times \text{Major Thoroughfare} + 0.106 \times \text{Number of Lanes} + 0.000007 \times \text{Single Family housing} - 0.013 \times \text{Manufactured House})$

TABLE 5.5: Generalized Estimating Equations Models based on Poisson - Log Function with Network Characteristics - All Road Functional Classes

Variables	1.0 Mile		1.5 Mile		2 Mile		3 Mile		4 Mile		5 Mile	
	Coefficient	P Value	Coefficient	P Value	Coefficient	P Value	Coefficient	P Value	Coefficient	P Value	Coefficient	P Value
(Intercept)	1.530	.000	1.604	.000	1.813	.000	1.803	.000	1.812	.000	1.919	.000
CBD											.510	.057
Urban					.116	.047	.123	.037	.122	.037	.412	.008
Freeways/Expressways	2.129	.000	2.085	.000	1.930	.000	1.933	.000	1.926	.000	1.942	.000
Major Thoroughfares	.713	.000	.674	.000	.674	.000	.679	.000	.674	.000	.706	.000
# Lanes	.120	.000	.116	.000	.120	.000	.120	.000	.119	.000	.113	.000
Population											-.004	.073
Single Family	0.000014	.000	0.000007	.000								
Industrial	0.000014	.011										
Research District	0.000035	.080										
Mobile Residential												
Commercial Center												
Neighborhood Service District												
Rural District							0.001640	.085				
Manufactured House			-0.014088	.000	-0.013476	.000					-0.002863	.012
Innovative											0.000020	.046
QIC	1882.234		1869.950		1952.710		1976.275		1977.016		1962.042	
QICC	1813.390		1814.764		1897.426		1918.809		1919.433		1854.635	

TABLE 5.6: Generalized Estimating Equations Models based on Negative Binomial - Log Function with Network Characteristics - All Road Functional Classes

Variables	1.0 Mile		1.5 Mile		2 Mile		3 Mile		4 Mile		5 Mile	
	Coefficient	P Value	Coefficient	P Value	Coefficient	P Value	Coefficient	P Value	Coefficient	P Value	Coefficient	P Value
(Intercept)	1.665	.000	1.578	.000	1.585	.000	1.875	.000	1.858	.000	1.909	.000
Urban	.176	.003	.149	.006	.239	.000	.236	.000	.248	.000	.251	.000
Freeways/Expressways	2.061	.000	2.117	.000	1.975	.000	1.927	.000	1.925	.000	1.942	.000
Major Thoroughfares	.670	.000	.660	.000	.674	.000	.672	.000	.667	.000	.668	.000
# Lanes	.112	.000	.106	.000	.123	.000	.120	.000	.121	.000	.116	.000
Population							-.003	.035	-.001	.080	-.001	.031
Mean Income					.003	.060						
Single Family	0.000010	.001	0.000007	.000								
Mobile Residential												
Commercial Center												
Neighborhood Service District												
Manufactured House			-0.012998	.000	-0.012757	.000	-0.005117	.098			-0.003701	.004
Innovative												
QIC	61.052		59.200		61.969		62.105		62.283		61.597	
QICC	71.047		71.292		73.891		73.615		72.202		73.306	

5.3.2 Models based on Freeways/Expressways

A summary of parameters and the corresponding goodness of fit statistics for each selected buffer width for Poisson and Negative Binomial distributions with log link were presented in Table 5.7 and Table 5.8, respectively.

It is evident from the goodness of fit statistics that the Negative Binomial with log link fits the model better than the Poisson with log link. It was also observed that the goodness of fit statistics were better than the models for all road functional classes. While the Poisson based models have QIC and QICC values around 1,000, the Negative Binomial models have QIC and QICC around10 and 20. The QIC and QICC are relatively close to each other for all the buffer widths considered in the research.

The Negative Binomial model based on two mile buffer width has the lowest QIC and QICC. The difference between these two statistics is also comparatively lower. This model indicates that area type (CBD), number of lanes, downstream link speed limit, downstream cross-street link 1 – number of lanes, mobile residential and manufactured house land use are significant variables in estimating AADT. It is expressed mathematically as follows.

$(\text{AADT in thousands})_{2\ \text{Mile}} = \text{Exp}\ (2.425 + 0.203 \times \text{CBD} + 0.138 \times \text{Number of Lanes} + 0.017 \times \text{Downstream Link Speed Limit} + 0.078 \times \text{Downstream Cross Street Link 1 Number of lanes} - 0.0000056 \times \text{Mobile Residential} - 0.011473 \times \text{Manufactured House})$

TABLE 5.7: Generalized Estimating Equations Models based on Poisson - Log Function with Network Characteristics – Freeways/Expressways

Variables	1.0 Mile		1.5 Mile		2 Mile		3 Mile		4 Mile		5 Mile	
	Coefficient	P Value	Coefficient	P Value	Coefficient	P Value	Coefficient	P Value	Coefficient	P Value	Coefficient	P Value
(Intercept)	2.377	.000	2.371	.000	2.620	.000	2.320	.000	2.693	.000	2.724	.000
CBD	.208	.002	.213	.002			.205	.003				
Urban												
# Lanes	.132	.000	.136	.000	.108	.000	.134	.000	.128	.000	.122	.000
Downstream Link Speed Limit	.019	.000	.019	.000	.019	.000	.020	.000	.020	.000	.020	.000
Downstream Cross Street Link 1 # Lanes	.082	.003	.074	.006	.053	.099	.080	.003				
# Students enrolled in High Schools					0.080317	.086						
Mobile Residential	-0.000746	.003			-0.000084	.000						
Rural District									-0.000937	.007		
Manufactured House			-0.011399	.000	-0.013618	.000					-0.003669	.028
Planned Unit Development	-0.000114	.000					-0.000018	.001				
Right of Way												
QIC	939.845		943.504		979.174		936.508		1044.391		1026.497	
QICC	879.788		885.731		923.215		875.047		1012.259		985.475	

TABLE 5.8: Generalized Estimating Equations Models based on Negative Binomial - Log Function with Network Characteristics – Freeways/Expressways

Variables	1.0 Mile		1.5 Mile		2 Mile		3 Mile		4 Mile		5 Mile	
	Coefficient	P Value	Coefficient	P Value	Coefficient	P Value	Coefficient	P Value	Coefficient	P Value	Coefficient	P Value
(Intercept)	1.791	.000	2.374	.000	2.425	.000	1.829	.000	2.499	.000	2.688	.000
CBD	.363	.000	.217	.002	.203	.004	.316	.001				
Urban	.195	.036					.163	.089				
# Lanes	.126	.000	.144	.000	.138	.000	.128	.000	.139	.000	.134	.000
Speed Limit	.013	.061					.012	.088				
Downstream Link Speed Limit	.014	.000	.018	.000	.017	.000	.016	.000	.019	.000	.019	.000
Downstream Cross Street Link 1 # Lanes	.081	.002	.076	.004	.078	.004	.079	.003	.052	.100		
Mobile Residential	-0.000466	.000			-0.000056	.000						
Rural District									-0.001115	.001		
Manufactured House			-0.011216	.000	-0.011473	.000					-0.003562	.002
Planned Unit Development	-0.000058	.032					-0.000011	.042				
Right of Way	0.000145	.087										
QIC	10.606		10.754		10.582		10.764		11.874		11.678	
QICC	29.406		21.987		23.795		25.660		21.205		19.146	

5.3.3 Models based on Major Thoroughfares

A summary of parameters and the corresponding goodness of fit statistics for each selected buffer width for Poisson and Negative Binomial distributions with log link were presented in Table 5.9 and Table 5.10, respectively.

It is evident from the goodness of fit statistics that the Negative Binomial with log link fits the model better than the Poisson with log link. It was also observed that the goodness of fit statistics were better than the models for all road functional classes combined. While the Poisson based models have QIC and QICC values around 350, the Negative Binomial models have QIC and QICC around 20 and 30. The QIC and QICC are relatively close to each other for all the buffer widths considered in the research.

The Negative Binomial model based on one and a half mile buffer width has the lowest QIC and QICC. The difference between these two statistics is also comparatively lower. This model indicates that area type (urban), speed limit, planned unit development and right of way land uses are significant variables in estimating AADT. It is expressed mathematically as follows.

$(\text{AADT in thousands})_{1.5\ \text{Mile}} = \text{Exp}\ (0.753 + 0.509 \times \text{Urban} + 0.048 \times \text{Speed Limit} + 0.000045 \times \text{Planned Unit Development} + 0.000384 \times \text{Right of Way})$

TABLE 5.9: Generalized Estimating Equations Models based on Poisson - Log Function with Network Characteristics – Major Thoroughfares

Variables	1.0 Mile		1.5 Mile		2 Mile		3 Mile		4 Mile		5 Mile	
	Coefficient	P Value	Coefficient	P Value	Coefficient	P Value	Coefficient	P Value	Coefficient	P Value	Coefficient	P Value
(Intercept)	1.679	.000	1.308	.001	1.380	.002	1.477	.000	1.589	.000	.614	.067
CBD	-.423	.003	-.268	.044								
Urban			.360	.000	.467	.000	.469	.000	.528	.000	.412	.000
Speed Limit	.024	.014	.033	.000	.035	.000	.032	.000	.029	.002	.048	.000
Upstream Link Speed Limit	.007	.016	.005	.080	.006	.039	.006	.040	.007	.030	.005	.092
Downstream Cross Street Link 1 # Lanes									.049	.043		
Population					-.014	.020	-.006	.007	-.005	.001		
# Education related Employees												
Single Family	0.000008	.061										
Research District												
Rural District									-0.002672	.068		
Manufactured House							-0.003486	.002	-0.004298	.000		
Planned Unit Development			0.000029	.020								
QIC	394.175		362.481		368.204		364.917		355.015		378.744	
QICC	375.003		344.011		349.836		348.610		335.867		362.360	

TABLE 5.10: Generalized Estimating Equations Models based on Negative Binomial - Log Function with Network Characteristics – Major Thoroughfares

Variables	1.0 Mile		1.5 Mile		2 Mile		3 Mile		4 Mile		5 Mile	
	Coefficient	P Value	Coefficient	P Value	Coefficient	P Value	Coefficient	P Value	Coefficient	P Value	Coefficient	P Value
(Intercept)	1.456	.000	.753	.027	1.539	.000	1.594	.000	1.589	.000	.594	.075
CBD	-.378	.003										
Urban	.323	.000	.509	.000	.514	.000	.511	.000	.534	.000	.471	.000
Speed Limit	.028	.003	.048	.000	.029	.003	.028	.003	.029	.001	.053	.000
Upstream Link Speed Limit	.005	.079			.006	.045	.006	.059	.005	.077		
Downstream Cross Street Link 1 # Lanes	.050	.067			.054	.047	.057	.039	.053	.049		
Population					-.018	.002	-.008	.001	-.005	.001		
Research District												
Manufactured House							-0.003913	.001	-0.003921	.000		
Neighborhood Service District												
Planned Unit Development			0.000045	.000								
Right of Way			0.000384	.000								
QIC	21.562		21.366		21.441		21.404		21.314		22.322	
QICC	31.926		30.246		31.826		33.753		33.689		27.363	

5.3.4 Models based on Minor Thoroughfares

A summary of parameters and the corresponding goodness of fit statistics for each selected buffer width for Poisson and Negative Binomial distributions with log link were presented in Table 5.11 and Table 5.12, respectively.

It is evident from the goodness of fit statistics that the Negative Binomial with log link fits the model better than the Poisson with log link. It was also observed that the goodness of fit statistics were better than the models for all road functional classes combined. While the Poisson based models have QIC and QICC values around 130, the Negative Binomial models have QIC and QICC around 15 and 25. The QIC and QICC are relatively close to each other for all the buffer widths considered in the research.

The Negative Binomial model based on one and a half mile buffer has the lowest QIC and QICC, The difference between these two statistics is also comparatively lower. This model indicates that area type (urban), upstream link speed limit, downstream link speed limit and institutional land use are significant variables in estimating AADT. It is expressed mathematically as follows.

$(\text{AADT in thousands})_{1.5\ \text{Mile}} = \text{Exp}\,(1.756 + 0.204 \times \text{Urban} + 0.006 \times \text{Upstream Link Speed Limit} + 0.004 \times \text{Downstream Link Speed limit} - 0.000027 \times \text{Institutional})$

TABLE 5.11: Generalized Estimating Equations Models based on Poisson - Log Function with Network Characteristics – Minor Thoroughfares

Variables	1.0 Mile		1.5 Mile		2 Mile		3 Mile		4 Mile		5 Mile	
	Coefficient	P Value	Coefficient	P Value	Coefficient	P Value	Coefficient	P Value	Coefficient	P Value	Coefficient	P Value
(Intercept)	1.742	.000	1.755	.000	1.742	.000	1.894	.000	1.742	.000	1.742	.000
Urban	.190	.012	.197	.009	.190	.012	.183	.017	.190	.012	.190	.012
Upstream Link Speed Limit	.006	.002	.006	.003	.006	.002	.006	.002	.006	.002	.006	.002
Downstream Link Speed Limit	.004	.093	.005	.069	.004	.093			.004	.093	.004	.093
Downstream Cross Street Link 1 # Lanes												
Institutional			-0.000025	.030								
Research District												
Mobile Residential												
Manufactured House							-0.011807	.000				
QIC	133.198		132.146		133.198		129.872		133.198		133.198	
QICC	131.702		132.002		131.702		130.810		131.702		131.702	

TABLE 5.12: Generalized Estimating Equations Models based on Negative Binomial - Log Function with Network Characteristics – Minor Thoroughfares

Variables	1.0 Mile		1.5 Mile		2 Mile		3 Mile		4 Mile		5 Mile	
	Coefficient	P Value	Coefficient	P Value	Coefficient	P Value	Coefficient	P Value	Coefficient	P Value	Coefficient	P Value
(Intercept)	1.860	.000	1.756	.000	1.860	.000	1.886	.000	1.860	.000	1.860	.000
Urban	.190	.012	.204	.006	.190	.012	.182	.015	.190	.012	.190	.012
Upstream Link Speed Limit	.007	.001	.006	.003	.007	.001	.007	.002	.007	.001	.007	.001
Downstream Link Speed Limit			.004	.091								
Downstream Cross Street Link 1 # Lanes												
Institutional			-0.000027	.009								
Research District												
Mobile Residential												
Manufactured House							-0.011727	.000				
QIC	15.610		15.397		15.610		14.794		15.610		15.610	
QICC	20.829		24.268		20.829		22.044		20.829		20.829	

5.4 Models without Network Characteristics

Models were developed to estimate AADT for planning purposes using the final variables obtained from the correlation analysis conducted among all the independent variables such as socio-economic and land use characteristics data (network characteristics are excluded here). Models developed based on all and each individual road functional classes are presented in this section. Only the variables that contributed to the final model and their corresponding parameters were presented in the results tables. Variables that do not contribute to the final model were left blank in the tables. All land uses in the tables are in thousand square feet.

5.4.1 Models Based on All Road Functional Classes

A summary of parameters and the corresponding goodness of fit statistics for each selected buffer width for Poisson and Negative Binomial distributions with log link were presented in Table 5.13 and Table 5.14, respectively.

It is evident from the goodness of fit statistics that the Negative Binomial with log link fits the model better than the Poisson with log link. While the Poisson based models have QIC and QICC around 7,000 and 10,000, the Negative Binomial models have QIC and QICC around 230 and 310. The QIC and QICC are relatively close to each other for all buffer widths examined in this research.

The Negative Binomial model based on one mile buffer width has the lowest QIC and QICC. This model indicates that area types (CBD, urban), population, single family housing, industrial, institutional, mobile residential, and right of way land uses are significant variables in estimating AADT It is expressed mathematically as follows.

$(\text{AADT in thousands})_{1.0\ \text{Mile}} = \text{Exp}\,(4.063 + 0.969 \times \text{CBD} + 0.898 \times \text{Urban} - 0.240 \times \text{Population} - 0.00003 \times \text{Single Family housing} - 0.00002 \times \text{Industrial} - 0.00009 \times \text{Research District} - 0.00005 \times \text{Institutional} - 0.00029 \times \text{Mobile Residential} + 0.0007 \times \text{Right of Way})$

TABLE 5.13: Generalized Estimating Equations Models based on Poisson - Log Function without Network Characteristics - All Road Functional Classes

Variables	One Mile		One Half Mile		Two Mile		Three Mile		Four Mile		Five Mile	
	Coefficient	P Value	Coefficient	P Value	Coefficient	P Value	Coefficient	P Value	Coefficient	P Value	Coefficient	P Value
(Intercept)	4.050	.000	4.100	.000	5.914	.000	4.194	.000	4.406	.000	4.453	.000
CBD	1.236	.000	1.159	.000	1.198	.000	2.130	.000	2.455	.000	2.289	.000
Urban	.905	.000	.882	.000	.611	.000	1.073	.000	1.483	.000	1.515	.000
Population	-.259	.000	-.097	.000	-.084	.000	-.045	.000	-.029	.000	-.019	.000
Mean Income					-.026	.000						
# Students Enrolled in High Schools							.144	.008				
Single Family	-0.00003	.000	-0.00001	.000								
Industrial	-0.00002	.028										
Institutional			-0.00004	.091	-0.00005	.002						
Mobile Residential	-0.00031	.008	-0.00021	.010	-0.00018	.000						
Rural District					-0.00959	.000	-0.00715	.030	-0.00487	.046		
Manufactured House					-0.01491	.000						
Planned Unit Development							-0.00003	.066				
Right of Way	0.00053	.003	0.00021	.003								
QIC	7661.710		9436.570		8144.379		10123.752		10747.583		10970.454	
QICC	7271.903		9038.138		7828.160		9695.196		10412.094		10663.981	

TABLE 5.14: Generalized Estimating Equations Models based on Negative Binomial - Log Function without Network Characteristics - All Road Functional Classes

Variables	One Mile		One Half Mile		Two Mile		Three Mile		Four Mile		Five Mile	
	Coefficient	P Value	Coefficient	P Value	Coefficient	P Value	Coefficient	P Value	Coefficient	P Value	Coefficient	P Value
(Intercept)	4.063	.000	4.054	.000	5.408	.000	4.284	.000	4.383	.000	4.520	.000
CBD	.969	.000	.931	.003	1.404	.000	2.352	.000	2.358	.000	2.298	.000
Urban	.898	.000	.817	.000	.808	.000	1.431	.000	1.564	.000	1.684	.000
Population	-.240	.000	-.084	.000	-.088	.000	-.048	.000	-.029	.000	-.020	.000
Mean Income					-.018	.000						
Single Family	-0.00003	.000	-0.00001	.001								
Industrial	-0.00002	.022										
Research District	-0.00009	.005										
Institutional	-0.00005	.069	-0.00005	.018	-0.00005	.000						
Mobile Residential	-0.00029	.003	-0.00014	.043	-0.00015	.000						
Rural District					-0.00938	.000	-0.00726	.004	-0.00509	.007	-0.00382	.029
Manufactured House					-0.01272	.000						
Right of Way	0.00070	.001	0.00031	.003								
QIC	238.018		285.765		250.753		296.476		309.808		312.192	
QICC	245.981		289.546		259.964		298.512		310.462		313.156	

5.4.2 Models based on Freeways/Expressways

A summary of parameters and the corresponding goodness of fit statistics for each selected buffer width for Poisson and Negative Binomial distributions with log link were presented in Table 5.15 and Table 5.16, respectively.

It is evident from the goodness of fit statistics that the Negative Binomial with log link fits the model better than the Poisson with log link. It was also observed that the goodness of fit statistics were better than the models for all road functional classes. While the Poisson based models have QIC and QICC values around 1,300 and 1,400, the

Negative Binomial models have QIC and QICC around 15 and 25. The QIC and QICC are relatively close to each other for all the buffer widths considered in the research.

The Negative Binomial model based on one and a half mile buffer width has the lowest QIC and QICC. The difference between these two statistics is also comparatively lower. This model indicates that area type (urban), population, and manufactured house land use are significant variables in estimating AADT. It is expressed mathematically as follows.

$$(\text{AADT in thousands})_{1.5\ \text{Mile}} = \text{Exp}\ (4.282 + 0.206 \times \text{urban} + 0.026 \times \text{Population} - 0.010164 \times \text{Manufactured House})$$

TABLE 5.15: Generalized Estimating Equations Models based on Poisson - Log Function without Network Characteristics – Freeways/Expressways

Variables	1.0 Mile		1.5 Mile		2 Mile		3 Mile		4 Mile		5 Mile	
	Coefficient	P Value	Coefficient	P Value	Coefficient	P Value	Coefficient	P Value	Coefficient	P Value	Coefficient	P Value
(Intercept)	4.411	.000	4.284	.000	4.353	.000	4.306	.000	4.370	.000	4.306	.000
CBD							.296	.002	.290	.003	.296	.002
Urban	.142	.068	.199	.009			.311	.001	.352	.000	.311	.001
Population	.046	.039	.026	.000	.008	.064						
# Students enrolled in High Schools					.141	.028						
Institutional									-0.000008	.099		
Mobile Residential	-0.000839	.002			-0.000112	.000						
Rural District									-0.001900	.001		
Commercial Center	-0.000178	.052										
Manufactured House			-0.010218	.000	-0.012600	.000						
Innovative	-0.000718	.000										
Planned Unit Development	-0.000151	.000										
QIC	1479.961		1479.328		1471.010		1521.122		1505.162		1521.122	
QICC	1389.640		1411.386		1401.984		1445.380		1408.240		1445.380	

TABLE 5.16: Generalized Estimating Equations Models based on Negative Binomial - Log Function with Network Characteristics – Freeways/Expressways

Variables	1.0 Mile		1.5 Mile		2 Mile		3 Mile		4 Mile		5 Mile	
	Coefficient	P Value	Coefficient	P Value	Coefficient	P Value	Coefficient	P Value	Coefficient	P Value	Coefficient	P Value
(Intercept)	4.415	.000	4.282	.000	4.407	.000	4.306	.000	4.372	.000	4.333	.000
CBD							.296	.002	.286	.003	.268	.006
Urban	.141	.070	.206	.008			.311	.001	.343	.000	.284	.002
Population	.045	.041	.026	.000	.010	.022						
Institutional									-0.000008	.080		
Mobile Residential	-0.000860	.000			-0.000106	.000						
Rural District									-0.001860	.001		
Commercial Center	-0.000199	.014										
Manufactured House			-0.010164	.000	-0.012546	.000					-0.003116	.041
Innovative	-0.000721	.000										
Planned Unit Development	-0.000148	.000										
QIC	17.464		17.708		17.833		18.217		18.074		18.084	
QICC	30.345		24.881		25.226		23.328		26.895		25.053	

5.4.3 Models based on Major Thoroughfares

A summary of parameters and the corresponding goodness of fit statistics for each selected buffer width for Poisson and Negative Binomial distributions with log link were presented in Table 5.17 and Table 5.18, respectively.

It is evident from the goodness of fit statistics that the Negative Binomial with log link fits the model better than the Poisson with log link. It was also observed that the goodness of fit statistics were better than the models for all road functional classes. While the Poisson based models have QIC and QICC values around 400, the Negative Binomial models have QIC and QICC around 20 and 30. The QIC and QICC are relatively close to each other for all the buffer widths considered in the research.

The Negative Binomial model based on one mile buffer width has the lowest QIC and QICC. The difference between these two statistics is also comparatively lower. This model indicates that area type (CBD), population, and neighborhood service district land use are significant variables in estimating AADT. It is expressed mathematically as follows.

(AADT in thousands)$_{1.0\ \text{Mile}}$ = Exp (2.964 – 0.976 × CBD + 0.062 × Population – 0.000379 × Neighborhood Service District)

TABLE 5.17: Generalized Estimating Equations Models based on Poisson - Log Function without Network Characteristics – Major Thoroughfares

Variables	1.0 Mile		1.5 Mile		2 Mile		3 Mile		4 Mile		5 Mile	
	Coefficient	P Value	Coefficient	P Value	Coefficient	P Value	Coefficient	P Value	Coefficient	P Value	Coefficient	P Value
(Intercept)	2.964	.000	2.883	.000	3.048	.000	3.074	.000	3.032	.000	3.033	.000
CBD	-.976	.000	-.997	.000	-.668	.000	-.693	.000	-.652	.000	-.653	.000
Urban					.232	.012	.203	.033	.227	.016	.228	.015
Population	.062	.014	.037	.002								
# Education related Employees												
Research District					-0.000026	.061						
Neighborhood Service District	-0.000379	.005	-0.000253	.029								
Manufactured House							-0.005796	.000	-0.005375	.000	-0.005995	.000
Right of Way							-0.000062	.010				
QIC	401.162		396.218		402.576		395.771		404.322		403.395	
QICC	387.225		381.429		384.379		382.429		391.286		390.245	

TABLE 5.18: Generalized Estimating Equations Models based on Negative Binomial - Log Function with Network Characteristics – Major Thoroughfares

Variables	1.0 Mile		1.5 Mile		2 Mile		3 Mile		4 Mile		5 Mile	
	Coefficient	P Value	Coefficient	P Value	Coefficient	P Value	Coefficient	P Value	Coefficient	P Value	Coefficient	P Value
(Intercept)	2.988	.000	2.847	.000	3.046	.000	3.073	.000	3.032	.000	3.032	.000
CBD	-.942	.000	-.991	.000	-.667	.000	-.692	.000	-.652	.000	-.653	.000
Urban					.231	.012	.199	.037	.227	.016	.227	.015
Population	0.054308	.035	0.038948	.001								
Research District					-0.000024	.087						
Manufactured House							-0.005785	.000	-0.005375	.000	-0.005695	.000
Neighborhood Service District	-0.000395	.001	-0.000268	.010								
Right of Way			0.000332	.060			-0.000055	.001				
QIC	23.185		22.573		23.214		22.728		23.148		23.109	
QICC	29.993		31.255		29.765		31.547		30.021		29.982	

5.4.4 Models based on Minor Thoroughfares

A summary of parameters and the corresponding goodness of fit statistics for each selected buffer width for Poisson and Negative Binomial distributions with log link were presented in Table 5.19 and Table 5.20, respectively.

It is evident from the goodness of fit statistics that the Negative Binomial with log link fits the model better than the Poisson with log link. It was also observed that the goodness of fit statistics were better than the models for all road functional classes. While the Poisson based models have QIC and QICC values around 140, the Negative Binomial models have QIC and QICC around 15 and 20. The QIC and QICC are relatively close to each other for all the buffer widths considered in the research.

The Negative Binomial model based on one mile buffer width has the lowest QIC and QICC. The difference between these two statistics is also comparatively lower. These models indicate that area type (urban), and mean income are significant variables in estimating AADT. It is expressed mathematically as follows.

$$(\text{AADT in thousands})_{1.0\ \text{Mile}} = \text{Exp}\,(2.231 + 0.219 \times \text{urban} - 0.003 \times \text{mean income})$$

TABLE 5.19: Generalized Estimating Equations Models based on Poisson - Log Function without Network Characteristics – Minor Thoroughfares

Variables	1.0 Mile		1.5 Mile		2 Mile		3 Mile		4 Mile		5 Mile	
	Coefficient	P Value	Coefficient	P Value	Coefficient	P Value	Coefficient	P Value	Coefficient	P Value	Coefficient	P Value
(Intercept)	2.060	.000	2.081	.000	2.060	.000	2.072	.000	2.060	.000	2.060	.000
Urban	.227	.005	.234	.004	.227	.005	.216	.007	.227	.005	.227	.005
Institutional			-0.000026	.082								
Research District												
Mobile Residential												
Manufactured House							-.014	.000				
QIC	145.235		144.526		145.235		138.818		145.235		145.235	
QICC	143.892		144.160		143.892		139.599		143.892		143.892	

TABLE 5.20: Generalized Estimating Equations Models based on Negative Binomial - Log Function with Network Characteristics – Minor Thoroughfares

Variables	1.0 Mile		1.5 Mile		2 Mile		3 Mile		4 Mile		5 Mile	
	Coefficient	P Value	Coefficient	P Value	Coefficient	P Value	Coefficient	P Value	Coefficient	P Value	Coefficient	P Value
(Intercept)	2.231	.000	2.080	.000	2.060	.000	2.072	.000	2.060	.000	2.060	.000
Urban	.219	.008	.236	.003	.227	.005	.216	.007	.227	.005	.227	.005
Mean Income	-.003	.051	-0.000026	.048								
Research District												
Mobile Residential												
Manufactured House							-.014	.000				
QIC	16.692		16.739		16.853		15.807		16.853		16.853	
QICC	21.911		22.084		20.297		21.269		20.297		20.297	

5.5 Summary: Models based on Spatial Proximity Method

The above results indicate that Negative Binomial is better than Poisson models in all the cases. The goodness of fit statistics indicates that the models performed better, when each individual road functional classes are modeled than the models for all the road functional classes combined. This suggests that the models should be developed considering each road functional class.

The models with network characteristics have comparatively better goodness of fit statistics than that of the models without including network characteristics. This satisfies the intuition that better models can be obtained when network characteristics are also considered. Also, upstream and downstream network link variables are found to be significant in the models based on freeways/expressways and minor thoroughfares. This suggests that spatial dependency could affect link level travel.

A closer look at the best models among each road functional class reveals some important findings. All road functional classes combined have a better model at “1.5” mile and “1.0” mile buffers for models with and without network characteristics respectively. The freeway/expressway road functional class have a better model at “2” mile and “1.5” mile buffers, while major thoroughfare and minor thoroughfare type of road functional classes have a better model at “1.5” and “1.0” mile buffers with and without network characteristics respectively. This suggests that the sphere of influence decreases as the hierarchy of road functional class decreases. In other words, freeways and expressways have a sphere of influence or accessibility levels higher than compared to the major and minor thoroughfares. However, the results indicate that accessible distances for various roads are varying from 1 mile to 2 mile and are not consistent. Accessibility and intensity of travel differs as the distance increases and the variations in the way people access roadways are expected. It is expected that better estimates can be obtained by applying gradually decreasing weights with respect to accessible distance.

Hence, there is a need to use multiple buffers and apply varying weights depending on the proximity to the study link.

CHAPTER VI: MODELS BASED ON SPATIAL WEIGHTS

An examination of results obtained for different spatial buffers showed that the sphere of influence decreases as the accessible distance increases from the roadway. It was also felt that the lower functional classes have comparatively lower sphere of influence in attracting traffic than the higher functional class roads. Hence, it was felt that developing models by combining data for different spatial buffers based on spatial weights would give better results. The spatial weights for different spatial buffers are applied based on spatial gravity principles derived from the “Gravity” method of the trip distribution. Spatial weights are calculated based on the assumption that the trip density decreases proportionally with the square of the distance. The calculation of spatial weights, statistical analysis and models developed based on combined data are discussed next.

6.1 Selection of Weights

Various weights are applied for the spatial data captured from “0” to “5” mile network distances based on the type of road functional class. The data are captured at various network bandwidths (0-1, 1-1.5, 1.5-2, 2-3, 3-4 and 4-5 mile) in the shape of a multiple “donuts” of various sizes.

It is observed in the spatial proximity method that the best models for freeways/expressways are based on “2” mile and “1.5” mile buffer width data, major thoroughfares are based on “1.5” mile buffer width data, and minor thoroughfares are based on “1.5” mile and “1.0” mile buffer width data for models with and without network characteristics respectively. Considering the best models selected previously, a larger network distance, “5” mile (all the six network buffer bandwidths considered) was used for all road functional classes combined and for freeways/expressways to capture the data. Comparatively less buffer widths were considered for major and minor thoroughfares as the best models are indicating a comparatively decreasing sphere of influence. For major thoroughfares, data captured over a “3” mile network distance was considered while data captured over a “2” mile network distance was considered for minor thoroughfares.

The computed weights for all road functional classes, freeways/expressways, major thoroughfares and minor thoroughfares are shown in Table 6.1 (a, b & c).

TABLE 6.1: Computed Spatial Weights for Different Road Functional Classes
(a). Weights for All Road Functional Classes Combined and for Freeways/Expressways

Bandwidth (miles)		$W_{i\text{-}j}$
i	j	
0	1	0.52
1	1.5	0.23
1.5	2	0.13
2	3	0.06
3	4	0.03
4	5	0.02
Total		1.00

(b). Weights for Major Thoroughfares

Bandwidth (miles)		W_{i-j}
i	j	
0	1	0.55
1	1.5	0.25
1.5	2	0.14
2	3	0.06
Total		1.00

(c). Weights for Minor Thoroughfares

Bandwidth (miles)		W_{i-j}
i	j	
0	1	0.59
1	1.5	0.26
1.5	2	0.15
Total		1.00

Data are processed to develop an integrated database using the above weights for each road functional class. Note that only demographic, socio-economic and land use data are multiplied by spatial weights. On-network characteristics were not multiplied by spatial weights.

Correlation analysis was conducted (as in the case of spatial proximity method) to eliminate variables with Pearson coefficient beyond the +/- 0.3 range and retain the variables within the +/- 0.3 range to avoid multicollinearity. The final set of independent variables (that are not correlated to each other) selected to develop the models for all and each road functional class considered, are listed in Table 6.2 below. GEE Models were developed with and without network characteristics using Poisson and Negative Binomial with log-link. The model parameters and goodness of fit statistics are presented and discussed next.

TABLE 6.2: Variables Selected for All and each Individual Road Functional Classes Considered

Variable / Spatial Buffer	All Roadways	Freeway/Expressway	Major Thoroughfare	Minor Thoroughfare
CBD	✓	✓	✓	✓
Urban	✓	✓	✓	✓
Suburban	✓	✓	✓	✓
Freeway/Expressway	✓			
Major Thoroughfare	✓			
Minor Thoroughfare	✓			
# Lanes	✓	✓	✓	✓
Speed Limit		✓	✓	
Upstream Link Speed Limit			✓	✓
Upstream Cross Street Link 1 # Lanes	✓	✓	✓	✓
Upstream Cross Street Link 1 Speed Limit		✓		
Upstream Cross Street Link 2 # Lanes		✓		
Upstream Cross Street Link 2 Speed Limit			✓	✓
Downstream Link Speed Limit		✓		✓
Downstream Cross Street Link 1 # Lanes		✓	✓	✓
Downstream Cross Street Link 1 Speed Limit	✓	✓		
Downstream Cross Street Link 2 Speed Limit			✓	✓
Population	✓	✓	✓	✓
# Pupils Enrolled in Public or Private High Schools			✓	✓
SingleFamily			✓	✓
Industrial				✓
Research District	✓	✓	✓	✓
Institutional	✓			✓
Mobile Residential			✓	
Rural District	✓	✓	✓	✓
Manufactured House	✓	✓	✓	
Commercial Center				✓
Neighborhood Service District			✓	
Innovative	✓	✓	✓	✓
RightofWay				✓

6.2 Models with Network Characteristics

Models based on spatial weighted data were developed to estimate AADT for analysis purposes using the final variables obtained from the correlation analysis. Models developed based on all and each individual road functional classes are presented in this section. Only the variables that were found to be significant in the final model and their corresponding parameters are presented in the results tables. Variables that do not contribute to the final model were left blank in the tables. All land uses in the tables are in thousand square feet.

6.2.1 Models based on All Road Functional Classes

Results obtained for all road functional classes are summarized in Table 6.3. As expected the goodness of fit statistics indicate that the Negative Binomial model is better than the Poisson model. The model parameters for the Negative Binomial model indicate that area type (urban), freeway/expressway, major thoroughfare, number of lanes, population, and manufactured house, and innovative land uses are significant variables in estimating AADT. It is expressed mathematically as follows.

(AADT in thousands) = Exp (1.837 + 0.244 $\times$ urban + 1.961 $\times$ Freeway/Expressway + 0.651 $\times$ Major Thoroughfare + 0.115 $\times$ number of lanes - 0.013 $\times$ population – 0.067 $\times$ manufactured house + 0.0002 $\times$ innovative)

TABLE 6.3: Generalized Estimating Equations Models for All Road Functional Classes with Network Characteristics

Variables	Negative Binomial		Poisson	
	Coefficient	P Value	Coefficient	P Value
(Intercept)	1.837	.000	1.777	.000
Urban	.244	.000	.120	.040
Freeways/Expressways	1.961	.000	1.962	.000
Major Thoroughfares	.651	.000	.651	.000
# Lanes	.115	.000	.117	.000
Population	-.013	.092		
Manufactured House	-.067	.001	-.050	.005
Innovative	0.0002	.051	0.0002	.049
QIC	61.336		1945.937	
QICC	74.756		1875.639	

6.2.2 Models based on Freeways/Expressways

Results obtained for freeways/expressways are summarized in Table 6.4. As expected, the goodness of fit statistics indicates that the models performed better than all the road functional classes combined. Even in this case, the Negative Binomial models are better than Poisson models. The model parameters indicate that number of lanes, downstream link speed limit, downstream cross street link 1 number of lanes, population and manufactured house land use are significant variables in estimating AADT. It is expressed mathematically as follows.

(AADT in thousands) = Exp (2.36 + 0.133 $\times$ number of lanes + 0.018 $\times$ downstream link speed limit + 0.07 $\times$ downstream cross street link 1 number of lanes + 0.029 $\times$ population – 0.036 $\times$ manufactured house)

TABLE 6.4: Generalized Estimating Equations Models for Freeways/Expressways with Network Characteristics

Variables	Negative Binomial		Poisson	
	Coefficient	P Value	Coefficient	P Value
(Intercept)	2.360	.000	2.325	.000
# Lanes	.133	.000	.122	.000
Downstream Link Speed Limit	.018	.000	.019	.000
Downstream Cross Street Link 1 # Lanes	.070	.010	.072	.009
Population	.029	.002	.032	.001
Manufacture House	-.036	.029		
QIC	10.850		953.215	
QICC	21.998		891.417	

6.2.3 Models based on Major Thoroughfares

Results obtained for major thoroughfares are summarized in Table 6.5. As expected, the goodness of fit statistics indicates that the models performed better than all the road functional classes combined. As observed previously, the Negative Binomial models are better than Poisson models. The model parameters indicate that area type (Urban), speed limit, upstream link speed limit, downstream cross street link 1 number of lanes, population and manufactured house land use are significant variables in estimating AADT. It is expressed mathematically as follows.

(AADT in thousands) = Exp (1.538 + 0.505 $\times$ urban + 0.03 $\times$ speed limit + 0.06 $\times$ upstream link speed limit + 0.053 $\times$ downstream cross street link 1 number of lanes - 0.051 $\times$ population – 0.059 $\times$ manufactured house)

TABLE 6.5: Generalized Estimating Equations Models for Major Thoroughfares with Network Characteristics

Variables	Negative Binomial		Poisson	
	Coefficient	P Value	Coefficient	P Value
(Intercept)	1.538	.001	1.366	.006
Urban	.505	.000	.458	.000
Speed Limit	.030	.006	.035	.001
Upstream Link Speed Limit	.006	.053	.006	.042
Downstream Cross Street Link 1 # Lanes	.053	.055		
Population	-.051	.007	-.037	.045
Manufactured House	-.059	.004	-.049	.015
QIC	21.685		372.446	
QICC	33.991		355.256	

6.2.4 Models based on Minor Thoroughfares

Results obtained for minor thoroughfares are summarized in Table 6.6. As expected, the goodness of fit statistics indicates that the models performed better than all the road functional classes combined. Also, the Negative Binomial models are better than Poisson models. The model parameters indicate that area type (Urban), speed limit, upstream link speed limit, and rural district land use are significant variables in estimating AADT. It is expressed mathematically as follows.

(AADT in thousands) = Exp (1.86 + 0.188 $\times$ urban + 0.007 $\times$ upstream link speed limit - 0.005 $\times$ rural district)

TABLE 6.6: Generalized Estimating Equations Models for Minor Thoroughfares with Network Characteristics

Parameter	Negative Binomial		Poisson	
	Coefficient	P Value	Coefficient	P Value
(Intercept)	1.860	.000	1.740	.000
Urban	.188	.014	.188	.014
Upstream Link Speed Limit	.007	.001	.007	.002
Downstream Link Speed Limit			.004	.089
Rural District	-.005	.095	-.007	.028
QIC	15.606		133.093	
QICC	22.819		133.524	

6.3 Models without Network Characteristics

Models based on spatial weighted data were developed to estimate AADT for planning purposes using the final variables obtained from the correlation analysis. Models developed based on all and each individual road functional classes are presented in this section. Only the variables that were found to be significant in the final model and their corresponding parameters are presented in the results tables. Variables that do not contribute to the final model were left blank in the tables. All land uses in the tables are in thousand square feet.

6.3.1 Models based on All Road Functional Classes

Results obtained for all road functional classes are summarized in Table 6.7. As expected, the goodness of fit statistics indicates that the Negative Binomial model is better than the Poisson model. The model parameters for the Negative Binomial model indicate that area type (CBD, urban, population, institutional, rural district and manufactured house land uses are significant variables in estimating AADT. It is expressed mathematically as follows.

(AADT in thousands) = Exp (4.729 + 2.936 $\times$ CBD + 1.778 $\times$ urban - 0.371 $\times$ population – 0.0001 $\times$ institutional – 0.094 $\times$ rural district – 0.09 $\times$ manufactured house)

TABLE 6.7: Generalized Estimating Equations Models for All Road Functional Classes without Network Characteristics

Variables	Negative Binomial		Poisson	
	Coefficient	P Value	Coefficient	P Value
(Intercept)	4.729	.000	4.729	.000
CBD	2.936	.000	3.227	.000
Urban	1.778	.000	1.764	.000
Population	-.371	.000	-.382	.000
Institutional	-0.00010	.067	-0.00012	.081
Rural District	-.094	.000	-.100	.003
Manufactured House	-.090	.085		
QIC	256.336		8580.255	
QICC	261.435		8297.823	

6.3.2 Models based on Freeways/Expressways

Results obtained for freeways/expressways are summarized in Table 6.8. As expected, the goodness of fit statistics indicates that the models performed better than all the road functional classes combined. Also, the Negative Binomial models are better than Poisson models. The model parameters indicate that area type (urban), population and manufactured house land use are significant variables in estimating AADT. It is expressed mathematically as follows.

(AADT in thousands) = Exp (4.276 + 0.168 $\times$ urban + 0.031 $\times$ population – 0.04 $\times$ manufactured house)

TABLE 6.8: Generalized Estimating Equations Models for Freeways/Expressways without Network Characteristics

Variables	Negative Binomial		Poisson	
	Coefficient	P Value	Coefficient	P Value
(Intercept)	4.276	.000	4.251	.000
Urban	.168	.029	.175	.023
Population	.031	.006	.034	.003
Manufacture House	-.040	.036		
QIC	18.211		1530.247	
QICC	25.239		1455.051	

6.3.3 Models based on Major Thoroughfares

Results obtained for major thoroughfares are summarized in Table 6.9. As expected, the goodness of fit statistics indicates that the models performed better than all the road functional classes combined. As observed previously, the Negative Binomial models are better than Poisson models. The model parameters indicate that area type (urban), population, single family, rural district, manufactured house and neighborhood service district land uses are significant variables in estimating AADT. It is expressed mathematically as follows.

(AADT in thousands) = Exp (3.078 + 0.398 $\times$ urban - 0.079 $\times$ population + 0.00001 $\times$ single family + 0.038 $\times$ rural district – 0.126 $\times$ manufactured house - 0.001 $\times$ neighborhood service district)

TABLE 6.9: Generalized Estimating Equations Models for Major Thoroughfares without Network Characteristics

Variables	Negative Binomial		Poisson	
	Coefficient	P Value	Coefficient	P Value
(Intercept)	3.078	.000	3.097	.000
Urban	.398	.001	.398	.001
Population	-.079	.000	-.078	.000
Single Family	0.00001	.047	0.00001	.080
Rural District	.038	.000	.034	.001
Manufactured House	-.126	.000	-.122	.000
Neighborhood Service District	-.001	.007	.000	.041
QIC	23.238		411.568	
QICC	35.680		395.220	

6.3.4 Models based on Minor Thoroughfares

Results obtained for minor thoroughfares are summarized in Table 6.10. As expected, the goodness of fit statistics indicates that the models performed better than all the road functional classes combined. As was seen previously, the Negative Binomial models are better than Poisson models. The model parameters indicate that area type (CBD) and population are significant variables in estimating AADT. It is expressed mathematically as follows.

(AADT in thousands) = Exp (1.903 - 0.422 $\times$ CBD + 0.081 $\times$ population)

TABLE 6.10: Generalized Estimating Equations Models for Minor Thoroughfares without Network Characteristics

Variables	Negative Binomial		Poisson	
	Coefficient	P Value	Coefficient	P Value
(Intercept)	1.903	.000	1.911	.000
CBD	-.422	.009	-.406	.011
Population	.081	.005	.079	.007
QIC	17.175		148.632	
QICC	22.345		146.633	

6.4 Summary: Models based on Spatial Weighting Method

In general, the above results indicate that Negative Binomial models are better than Poisson models in all the cases. The goodness of fit statistics indicates that the models performed better, when each individual road functional classes are modeled than the models for all the road functional classes combined. This suggests that the models should be developed considering each road functional class.

The models with network characteristics have comparatively better goodness of fit statistics than the models without including network characteristics. This satisfies the intuition that better models can be obtained when network characteristics are also considered. Also, upstream and downstream network link variables are found to be significant in the models developed for each individual road functional classes. This suggests that spatial dependency of network links can affect link level travel.

Models based on “spatial proximity” and “spatial weighting” methods were similar when goodness of fit statistics was compared. Though, in theory, it was felt that “spatial weighting” method is comparatively sound and would yield better results, the model results turned out to be similar. The best model selected in the spatial weighting method for all road functional classes has comparatively higher goodness of fit statistics than for all road functional classes in the spatial proximity method. This indicates that a common weighting scheme for all road functional classes is not suitable and that each road functional class should be applied a weighting scheme that is suitable for that particular road functional class. It can be observed from models developed that after applying different weights for each road functional class, depending on the accessible distances for each model, performance improved well.

CHAPTER VII: VALIDATION

Model validation using CSS and percent differences calculated between the actual and predicted AADT from the best model selected in each method for each road functional class are discussed in this Chapter. Validation of models developed for all road functional classes was not performed as the models for each road functional class is expected to be performing better. Links were selected separately for validation purpose. In other words, links selected for developing the models are not repeated while selecting links for validation. Links are selected in all three road functional classes and area types considered. The number of links selected for validation for each combination of road functional class and area types is around 10% of the links considered for developing the models. Table 7.1 below shows the number of links considered in each road functional class and area type. The links were selected in such a way that the AADT is varied from low to high. Table 7.2 below shows the minimum and maximum AADT of the links considered in the validation analysis.

TABLE 7.1: Links Considered in each Road Functional Class and Area Type

Road Functional Class / Area Type	CBD	Urban	Suburban	All Area Types
Freeways/Expressways	2	5	4	11
Major Thoroughfare	4	5	4	13
Minor Thoroughfare	2	4	3	9
All Roadway Types	8	14	11	33

TABLE 7.2: Minimum and Maximum AADT of Links Considered in each Road Functional Class

Road Functional Class	Minimum	Maximum
Freeways/Expressways	48	160
Major Thoroughfare	5	42
Minor Thoroughfare	6	18

Percent differences are calculated between the estimated AADT and actual AADT (observed). These are summarized in Table 7.3. It can be seen from the table that the models with network characteristics give better estimates than models without network characteristics. The spatial weighting models performed comparatively better than spatial proximity models.

TABLE 7.3: Average Percent Difference of Observed and Estimated AADT by Road Functional Class using Various Models

Road Functional Class / Spatial Method	Freeway/Expressway		Major Thoroughfare		Minor Thoroughfare	
	Spatial Proximity	Spatial Weighting	Spatial Proximity	Spatial Weighting	Spatial Proximity	Spatial Weighting
With Network Characteristics	27.12%	27.81%	39.24%	35.37%	26.02%	25.83%
Without Network Characteristics	37.97%	36.29%	36.38%	35.84%	30.24%	27.08%

As stated in the "Methodology" chapter, a modified Chi Square Statistic (X_m^2) was used to perform a Chi-Square test. The CSS calculated was compared with Critical CSS (at 99% confidence level) of the corresponding degrees of freedom (df = n – p) for each model, here, n is the validation sample size and p is the number of predictors of the corresponding model used.

Table 7.4 shows the CSS and Critical-CSS for each model. CSS is less than Critical-CSS in all the cases considered in the table. This indicates that both spatial proximity and spatial weighting methods, with and without network characteristics can be appropriately applied and used when and where they are deemed necessary.

TABLE 7.4: Chi-Square Test for Observed and Estimated AADT by Road Functional Class using Various Models

Road Functional Class / Spatial Method	Network Characteristics	Freeway/Expressway		Major Thoroughfare		Minor Thorougfare	
		CSS	Critical CSS	CSS	Critical CSS	CSS	Critical CSS
Proximity	Yes	9.06	13.28	13.56	20.09	8.37	13.28
	No	16.28	18.48	9.66	21.67	11.29	16.81
Weighting	Yes	10.28	15.09	10.85	16.81	8.00	15.09
	No	15.51	18.48	8.64	16.81	11.33	16.81

In general, model validation conducted indicates that models developed using both the spatial methods, spatial proximity and spatial weighting methods yielded comparatively equal estimates. Models with more details (with network characteristics) give comparatively more accurate results than models with less detail (without network characteristics).

CHAPTER VIII: CONCLUSIONS

Annual Average Daily Traffic (AADT), an output of Urban Transportation Planning Process (UTPP), is used in several planning, roadway design, operational and safety analysis by transportation planners and engineers. Accurately estimating or predicting travel demand helps planners and engineers to better utilize limited transportation funds to improve transportation system performance and cater to the future needs.

Existing methods are very complex and do not adequately address the modeling needs. The traditional four-step method is an aggregate sequential top down model. Errors and inaccuracies get carried to later steps often resulting in incorrect estimates of travel demand. Due to the aggregate nature of the modeling methods, spatial variations in the characteristics of data that influence travel demand are overlooked. Combined four-step and disaggregate tour based activity methods address some of the limitations of the aggregate model. However, they are computationally intensive and require enormous amounts of data limiting them currently to small scale applications. Estimation of travel demand using short term traffic counts using various seasonal and weekly factors is another easy method. However, estimating travel demand for all the links in the network from a very few number of traffic counts available leads to uncertainties and inaccuracies.

A methodology is proposed in this research involving scientific principles and statistical techniques but bypassing the tedious four-step method. As travel demand on roadways depends on road functional classification and their characteristics, three different road functional classes - freeways and expressways, major thoroughfares, and minor thoroughfares identified in the regional network model were selected to develop models to estimate travel demand by road functional class. Two methods are proposed to accomplish the task. While the first one is based on spatial proximity, the second is based on data integrated using spatial weights.

Various on-network and off-network characteristics data were considered as independent variables in the development of models. On-network characteristics of the study links, as well as upstream, downstream, and cross-streets network links, were considered to account for spatial dependency. Off-network characteristics considered include demographic, socio-economic, and land use characteristics.

In the first method, different buffer widths were considered to extract spatial data and develop models to examine the role of spatial proximity in estimating travel demand. The buffer widths considered include 1 mile, 1.5 mile, 2 mile, 3 mile, 4 mile and 5 mile distances. In the second method, spatially decreasing weights (based on gravity principles) are applied on data captured at various bandwidths and combined to develop models for various road functional classes. A summary of the research findings is presented next.

8.1 Summary of findings

Results obtained indicate that models based on Negative Binomial distribution yield better travel demand estimates than models based on Poisson distribution for data used in this research. This could be attributed to over-dispersion observed in the data. The goodness of fit statistics indicate that better travel demand estimates can be achieved when models are developed for each road functional class than when data was combined for all road functional classes.

Analysis based on various buffer widths to capture data showed that spatial proximity plays a vital role in accurately estimating travel demand. A comparison of results obtained from models generated using different buffer widths showed that using “2” mile or “1.5” mile buffer width (for higher road functional classes) followed by “1.0” mile buffer width (for lower road functional classes) to extract demographic, socio-economic, and land use data would yield the best estimates. In general, freeways and expressways, major thoroughfares and minor thoroughfares have decreasing sphere of influence or accessibility levels in the same order.

Based on the above research findings it was felt that the effect is expected to decrease away from the study location. Hence, spatial weights are applied using distance decay principles with decreasing weights applied on spatial off-network characteristics data away from study locations. Likewise, decreasing maximum accessible distances are considered for various road functional classes in a hierarchical order. It was proved that, based on the goodness of fit statistics, models developed for each road functional class performed better than for all road functional classes combined. The goodness of fit statistics indicates that models developed based on “spatial proximity” and “spatial weighting” methods have similar or not so different goodness of fit statistics.

Validation of the models developed was carried out using Chi-Square test. Chi-Square Statistic (CSS) was computed between estimates obtained from the proposed methodology and observed traffic counts and compared with critical value of CSS at 99 percent level of significance (P = 0.01). It was observed that the estimates are statistically close to the observed traffic counts. The average absolute percent difference is less than 30%, in general.

Models with more details (with network characteristics) were observed to be comparatively more accurate results than models with less detail (without network characteristics). Spatial dependency of network links was found to play a role in models developed based on both “spatial proximity” and “spatial weighting” methods. Significant differences were not observed between estimates obtained from “spatial proximity” and “spatial weighting” methods. Though the results are similar in terms of goodness of fit statistics and model validation, it was felt that spatial weighting method is theoretically sound and hence could better explain the travel compared to spatial proximity method. Also, the significant variables selected in the final models of the spatial weighting method are quantifiable and are widely used. They are easy to be forecasted in the future and hence easy to apply models for estimating future traffic.

The methods proposed in this research are easy to adopt and can be applied universally to urban settings of any size and level.

8.2 Limitations and Scope for future work

Data such as auto-ownership, number of drivers per household, and, population by gender and age group were not available in the planning variables data used in this research. Considering data along with those used in this research could increase predictive capability of the models.

Variation in travel by the time of the day was not considered in this research. Developing models by time of the day could help estimate demand and congestion on road links during peak and non-peak hours (by time of day). It can also help better design roads based on the duration of congestion.

The effect of mode choice was not considered in this research. The presence or access to public transportation systems could play a role on vehicular volume. The role of public transportation system on vehicular volume needs an examination.

Discrete spatial decay weights were applied for each road functional class considered to capture the effect of distance on travel demand. The application of continuous decay weights and varying weighting functions by road functional class may provide better and more accurate results. This merits an investigation.

Further, research needs to be carried out to determine the maximum accessible distances for each road functional class and the percentage of people accessing each road functional class within a specified buffer bandwidth.

REFERENCES

1. AASHTO. *Guidelines for Traffic Data Programs*. Washington, D.C., 1992.

2. Abbrahamsson, T. *Estimation of Origin–Destination Matrices using Traffic Counts – A Literature Survey*. Interim Report IR-98-021/May, International Institute for Applied Systems Analysis, Laxenburg, Austria, 1998.

3. Baek, S., Kim, H. and Y. Lim. Multiple–Vehicle Origin–Destination Matrix Estimation from Traffic Counts using Genetic Algorithm. *Journal of Transportation Engineering*, Vol. 130, No. 3, 2004, pp. 339–347.

4. Balinski, M. and K. Spielberg. *Methods for Integer Programming: Algebraic, Combinatorial and Enumeratial*. Aranofsky, M.J. (ed.), Program in Operations Research. John Wiley, New York, 1969.

5. Bar-Gera, H. and D. Boyce. Origin based Algorithms for Combined Travel Forecasting Models. *Transportation Research Part B*, Vol. 37, 2003, pp. 405–422.

6. Beckmann, M., McGuire, C. B. and C.B. Winsten. *Studies in the Economics of Transportation*, Yale University Press, New Haven, Connecticut, 1956.

7. Berry, W.D. *Understanding Regression Assumptions*. Sage Publications, Newbury park, CA, 1993.

8. Black, J. and M. Conroy. Accessibility Measures and the Social Evaluation of Urban Structure. *Environment and Planning A*, Vol. 9, 1977, pp. 1013–1031.

9. Blainey, S. Trip End Models of Local Rail Demand in England and Wales. *Journal of Transport Geography*, Vol. 18, No.1, 2010, pp. 153–165.

10. Boyce, D. and H. Bar-Gera. Multiclass Combined Models for Urban Travel Forecasting. *Networks and Spatial Economics*, Vol. 4, 2004, pp. 115–124.

11. Boyce, D. and H. Bar-Gera. Validation of Multiclass Urban Travel Forecasting Models Combining Origin-Destination, Mode, and Route Choices. *Journal of Regional Science, Vol. 43*, No. 3, 2003, pp. 517–40.

12. Boyce, D. Forecasting Travel on Congested Urban Transportation Networks: Review and Prospects for Network Equilibrium Models. *Network Spatial Economics*, Vol. 7, 2007, pp. 99–128.

13. Boyce, D. Is the Sequential Travel Forecasting Paradigm Counter–Productive? *Journal of Urban Planning and Development*, Vol. 128, No. 4, 2002, pp. 169–183.

14. Boyce, D., Lupa, M. and Y. Zhang. Introducing Feedback into Four-step Travel Forecasting Procedure Versus Equilibrium Solution of Combined Model. In *Transportation Research Record: Journal of the Transportation Research Board*, No. 1443, Transportation Research Board of the National Academies, Washington, D.C., 1994, pp. 65–74.

15. Brunsdon, C., Fotheringham, A.S. and M.E. Charlton. Spatial Nonstationarity and Autoregressive Models. *Environment and Planning A*, Vol. 30, No. 6, 1998b, pp. 957–973.

16. Brunsdon, C., Fotheringham, A.S. and M.E. Charlton. Geographically Weighted Regression: A Method for Exploring Spatial Nonstationarity. *Geographical Analysis*, Vol. 28, No. 4, 1996, pp. 281–298.

17. Brunsdon, C., Fotheringham, A.S. and M.E. Charlton. Geographically Weighted Regression – Modeling Spatial Non-stationarity. *The Statistician*, Vol. 47, No. 3, 1998a, pp. 431–443.

18. Brunsdon, C., Fotheringham, A.S. and M.E. Charlton. Geographically Weighted Discriminant Analysis. *Geographical Analysis*, Vol. 39, 2007, pp. 376–396.

19. Brunsdon, C., Fotheringham, A.S. and M.E. Charlton. Some Notes on Parametric Significance Tests for Geographically Weighted Regression. *Journal of Regional Science*, Vol. 39, No. 3, 1999, pp. 497–524.

20. Brunsdon, C., McClatchey, J. and D.J. Unwin. Spatial Variations in the Average Rainfall-Altitude Relationship in Great Britain – An Approach using Geographically Weighted Regression. *International Journal of Climatology*, Vol. 21, 2001, pp. 455–466.

21. Burns, L.D. *Transportation, Temporal, and Spatial Components of Accessibility*. Lexington Books, Lexington, MA., 1979.

22. Cameron, A.C. and P.K. Trivedi. *Regression Analysis of Count Data*. Cambridge University Press, Cambridge, UK, 1998.

23. Cascetta, E. Estimation of Trip Matrices from Traffic Counts and Survey Data: A Generalized Least Squares Estimator, *Transportation Research Part B*, Vol. 18B, No. 4/5, 1984, pp. 289–299.

24. Census 2008 Estimates. www.census.gov. Accessed, October 29, 2010.

25. Chow, L.F. Zhao, F., Liu, X., Li, M.T. and I. Ubaka. Transit Ridership Model Based on Geographically Weighted Regression. In *Transportation Research Record: Journal of the Transportation Research Board*, No. 1972, Transportation Research Board of the National Academies, Washington, D.C., 2006, pp. 105–114.

26. Church, R. and C. ReVelle. The Maximum Covering Location Problem. *Papers in Regional Science Association*, Vol. 32, 1974, pp. 101–118.

27. Church, R. and C. ReVelle. Theoretical and Computational Links between the p-Median, Location Set-Covering, and the Maximal Covering Location Problem. *Geographical Analysis*, Vol. 8, 1976, pp. 406–415.

28. City of Charlotte Zoning Ordinance (1992). *Charlotte Mecklenburg Planning Commission.* Codified through June 21, 2010. Accessed, October 19, 2010. http://www.charmeck.org/city/charlotte/planning/Rezoning/Pages/ZoningOrdinance.aspx

29. Clark, S.D. Estimating Local Car Ownership Models. *Journal of Transport Geography*, Vol. 15, 2007, pp. 184–197.

30. Current, J. and H. Min. Multi-objective Design of Transportation Networks: Taxonomy and Annotation. *European Journal of Operational Research*, Vol. 26, 1986, pp. 187–201.

31. Current, J.R., Cohon, J.L. and C.S. ReVelle. The Shortest Covering Path Problem: An Application of Locational Constraints to Network Design. J*ournal of Regional Science*, Vol. 24, 1984, pp. 161-183.

32. Current, J.R., ReVelle, C.S. and J.L. Cohon. The Application of Location Models to the Multi-objective Design of Transportation Networks. *Regional Science Review*, Vol. 14, 1985a, pp. 1–23.

33. Current, J.R., ReVelle, C.S. and J.L. Cohon. The Maximum Covering/Shortest Path Problem: A Multi-objective Network Design and Routing Formulation. *European Journal of Operational Research*, Vol. 21, 1985b, pp. 189–199.

34. Current, J.R., ReVelle, C.S. and J.L. Cohon. The Median Shortest Path Problem: A Multi-objective Approach to Analyze Cost Versus Accessibility in the Design of Transportation Networks. *Transportation Science*, Vol. 21, No. 3, 1987, pp. 188 – 197.

35. Dalvi, M.Q. and K.M. Martin. The Measurement of Accessibility: Some Preliminary Results. *Transportation*, Vol. 5, 1976, pp. 17–42.

36. Davis, G. Accuracy of Estimates of Mean Daily Traffic: A Review. In *Transportation Research Record: Journal of the Transportation Research Board*, No. 1593, Transportation Research Board of the National Academies, Washington, D.C., 1997, pp. 12–16.

37. Davis, G. and S. Yang. Accounting for Uncertainty in Estimates of Total Traffic Volume: An Empirical Bayes Approach. *Journal of Transportation Statistics*, Vol. 4, No. 1, 2001, pp. 27–38.

38. Davis, G. and Y. Guan. Bayesian Assignment of Coverage Count Locations to Factor Groups and Estimation of Mean Daily Traffic. In *Transportation Research Record: Journal of the Transportation Research Board*, No. 1542, Transportation Research Board of the National Academies, Washington, D.C., 1996, pp. 30–37.

39. Downes, J. D., Johnsen, S. and D. Morrell. *Household and Person Trip Generation Models*. Transport and Road Research Laboratory Report SR 401, Crowthome, England, 1978.

40. Downes, J.D. and D. Morrell. Variation of Travel Time Budgets and Trip Rates in Reading. *Transportation Research – A*, Vol. 15, 1981, pp. 47 – 53.

41. Elhai, J.D., Calhoun, P.S. and J.D. Ford. Statistical Procedures for Analyzing Mental Health Services Data. *Psychiatry Research*, Vol. 160, 2008, pp. 129 – 136.

42. Eom, J.K., Park, M.S., Heo, T-Y and L.F. Huntsinger. Improving the Prediction of Annual Average Daily Traffic for Non-Freeway Facilities by Applying a Spatial Statistical Method. In *Transportation Research Record: Journal of the Transportation Research Board*, No. 1968, Transportation Research Board of the National Academics, Washington, D.C., 2006, pp. 20–29.

43. Evans, S. P. Derivation and Analysis of Some Models for Combining Trip Distribution and Assignment. *Transportation Research*, Vol. 10, 1976, pp. 37–57.

44. Everitt, B.S. *The Cambridge dictionary of statistics*. Cambridge University Press, Cambridge, U.K., 1995.

45. Fawcett F. and Downes J. D. *The Spatial Uniformity of Trip Rates in the Reading Area in 1971*. Transport and Road Research Laboratory – Report LR 797, Crowthome, England, 1977.

46. FDOT. *Traffic Monitoring Handbook*. Transportation Statistics Office, Florida Department of Transportation. http://www.dot.state.fl.us/planning/statistics/tmh/ Accessed, December 10, 2009.

47. Fisch, O. Analytical Derivation and Behavioral Interpretation of a Model Combining Trip Generation and Distribution. *Socio-Economic Planning Sciences*, Vol. 19, No. 3, 1985, pp.159–165.

48. Fisk, C.S. and D.E. Boyce. A Note on Trip Matrix Estimation from Link Traffic Count Data, *Transportation Research Part B*, Vol. 17B, No. 3, 1983, pp. 245–250.

49. Florian, M. and S. Nguyen. A Combined Trip Distribution, Modal Split and Trip Assignment Model. *Transportation Research*, Vol. 12, 1978, pp. 241–246.

50. Florian, M., Nguyen, S. and J. Ferland. On the Combined Distribution-Assignment of Traffic. *Transportation Science*, Vol. 9, 1975, pp. 43–53.

51. Foody, G.M. Geographical Weighting as a Further Refinement to Regression Modelling - An Example Focused on the NDVI – Rainfall Relationship. *Remote Sensing of Environment*, Vol. 88, 2003, pp. 283–293.

52. Fotheringham, A.S., Brunsdon, C. and N. Charlton. *Geographically Weighted Regression: The Analysis of Spatially Varying Relationships*, John Wiley & Sons Ltd., West Sussex, United Kingdom, 2002.

53. Fox, J. *Regression Diagnostics*. Sage Publications, Newbury Park, CA, 1991.

54. Freedman, A.D. *Statistical Models – Theory and Practice*. Cambridge University Press, New York, NY, USA, 2005.

55. Friesz, T. L. An Equivalent Optimization Problem for Combined Multi-Class Trip Distribution, Assignment and Modal Split Which Obviates Symmetry Restrictions. *Transportation Research B*, Vol. 15, No. 5, 1981, pp. 361–369.

56. Gadda, S.C., Kockelman, K.M. and A. Magoon. Estimates of AADT: Quantifying the Uncertainty. *Eleventh World Conference on Transportation Research*, Berkeley, California, 2007. Accessed, October 1, 2009. http://www.ce.utexas.edu/prof/kockelman/public_html/TRB09AADT.pdf

57. Gardner, W., Mulvey, E.P. and E.C. Shaw. Regression Analyses of Counts and Rates: Poisson, Overdispersed Poisson, and Negative Binomial Models. *Psychological Bulletin*, 118, 1995, pp. 392–404.

58. Getis, A., and J. K. Ord. The Analysis of Spatial Association by Use of Distance Statistics. *Geographical Analysis*, Vol. 24, 1992, pp. 189–206.

59. Goel, P.K., McCord, M.R. and C. Park. Exploiting Correlations between Link Flows to Improve Estimation of Average Annual Daily Traffic on Coverage Count Segments: Methodology and Numerical Study. In *Transportation Research Record: Journal of the Transportation Research Board*, No. 1917, Transportation Research Board of the National Academics, Washington, D.C., 2005, pp.100–107.

60. Gourieroux, C., Monfort, A. and A. Trogon. Psuedo Maximum Likelihood Methods: Application to Poisson models. *Econometrica*, 52, 1984, pp. 701–720.

61. Granato, S. The Impact of Factoring Traffic Counts for Daily and Monthly Variation in Reducing Sample Counting Error. Proceedings of the Crossroads 2000

Conference (Ames, Iowa), 1998, pp.122-125. Accessed on October 1, 2009. http://www.ctre.iastate.edu/pubs/crossroads/122impact.pdf

62. Guy, C.M. The Assessment of Access to Local Shopping Opportunities: A Comparison of Accessibility Measures. *Environment and Planning B*, Vol. 10, 1983, pp. 219–238.

63. Hägerstrand, T. What about People in Regional Science? *People of the Regional Science Association*, Vol. 24, 1970, pp. 7–21.

64. Hall, R.W. Travel Outcome and Performance – The Effect of Uncertainty on Accessibility. *Transportation Research Part B*, Vol. 17B, No. 4, 1983, pp. 275–290.

65. Handy, S. Regional versus Local Accessibility: Implications for Non-Work Travel. In *Transportation Research Record: Journal of the Transportation Research Board*, No. 1400, Transportation Research Board of the National Academics, Washington, D.C., 1994, pp. 58–66.

66. Hansen, W.G. How Accessibility Shapes Land Use? *Journal of American Institute of Planners*, Vol. 25, No. 1, 1959, pp. 73–76.

67. Hardin, J.W. and J.M. Hilbe. *Generalized Estimating Equations*. Chapman & Hall, Boca Raton, FL, USA., 2003.

68. Hasan, M.K. and H.M. Dashti. A Multiclass Simultaneous Transportation Equilibrium Model. *Network Spatial Economics*, Vol. 7, 2007, pp. 197–211.

69. Hasan, M.K. and K.N.A. Safwat. Comparison of Two Transportation Network Equilibrium Modeling Approaches. *Journal of Transportation Engineering*, Vol. 126, No. 1, 2000, pp. 35–40.

70. Hasan, M.K. and S.A. Al-Gadhi. Application of Simultaneous and Sequential Transportation Network Equilibrium Models to Riyadh, Saudi Arabia. In *Transportation Research Record: Journal of the Transportation Research Board*, No. 1645, Transportation Research Board of the National Academics, Washington, D.C., 1998, pp. 127–132.

71. Hilbe, J.M. *Negative Binomial Regression*. Cambridge University Press, Cambridge, U.K., 2007.

72. Hilbe, J.M. Review: Generalized Linear Models. *The American Statistician*, Vol. 48, No.3, 1994, pp. 255–265.

73. Ho H.W., Wong S.C. and B.P.Y. Loo. Combined Distribution and Assignment Model for A Continuum Traffic Equilibrium Problem with Multiple User Classes. *Transportation Research Part B*, Vol. 40, 2006, pp. 633–650.

74. Hocking, R.R. *Methods and Applications of Linear Models – Regression and the Analysis of Variance*. Second Ed., John Wiley & Sons, Inc., Hoboken, N.J., 2003.

75. Ingram, D.R. The Concept of Accessibility: A Search for an Operational Form. *Regional Studies*, Vol. 5, 1971, pp. 101–107.

76. Jiang, Z. *Incorporating Image-Based Data in AADT Estimation - Methodology and Numerical Investigation of Increased Accuracy*. Ph.D. Dissertation, The Ohio State University, Columbus, OH, 2005.

77. Jiang, Z., McCord, M. R. and P. K. Goel. Improved AADT Estimation by Combining Information in Image and Ground-Based Traffic Data. *Journal of Transportation Engineering*, Vol. 132, No. 7, 2006, pp. 523–600.

78. Kockelman, K. Travel Behavior as a Function of Accessibility, Land Use Mixing, and Land Use Balance Evidence from San Francisco Bay Area. In *Transportation Research Record: Journal of the Transportation Research Board*, No. 1607, Transportation Research Board of the National Academics, Washington, D.C., 1997, pp. 116–125.

79. Koenig, G. *Les Indicateurs D'accessibilites Dans Les Studes Urbaines: De La Theorie A La Practique*. Revue Generale des Routes. July 1977.

80. Koenig, J.G. Indicators of Urban Accessibility: Theory and Application. *Transportation*, Vol. 9, 1980, pp. 145–172.

81. Kwan, M-P. Gender and Individual Access to Urban Opportunities: A Study Using Space-Time Measures. *Professional Geographer*, Vol. 51, 1999, pp. 210–227.

82. Kwan, M-P. Space–Time and Integral Measures of Individual Accessibility: A Comparative Analysis Using a Point-Based Framework. *Geographical Analysis*, Vol. 30, No. 3, 1998, pp. 191–216.

83. Lam W. H.K., Tang Y.F. and M.L. Tam. Comparison of Two Non-Parametric Models for Daily Traffic Forecasting in Hong Kong. *Journal of Forecasting*, 25, 2006, pp. 173–192.

84. Lam, W. and J. Xu. Estimation of AADT from Short Period Counts In Hong Kong – A Comparison between Neural Network Method and Regression Analysis. *Journal of Advanced Transportation*, Vol. 34, No. 2, 2000, pp. 249–268.

85. Leake, G. R. and A.S. Huzayyin. Accessibility Measures and their Suitability for use in Trip Generation Models. *Traffic Engineering Control*, Vol. 20, 1979, pp. 566–572.

86. Lenntorp, B. *Paths in Space-Time Environments: A Time Geographic Study of Movement Possibilities of Individuals*. Lund Studies in Geography, No. 44, Royal University of Lund, Sweden, 1976.

87. Li, M.T., Zhao, F. and L.F. Chow. Assignment of Seasonal Factor Categories to Urban Coverage Count Stations using a Fuzzy Decision Tree. *Journal of Transportation Engineering*, Vol. 132, No. 8, 2006, pp. 654–662.

88. Li, M.T., Zhao, F. and Y. Wu. Application of Regression Analysis for Identifying Factors that Affect Seasonal Traffic Fluctuations in Southeast Florida. In *Transportation Research Record: Journal of the Transportation Research Board*, No. 1870, Transportation Research Board of the National Academics, Washington, D.C., 2004, pp. 153–161.

89. Liang, K-Y. and S. L. Zeger. Longitudinal Data Analysis Using Generalized Linear Models. *Biometrika*, Vol. 73, No. 1, 1986, pp. 13–22.

90. Lim, H., and J.C. Thill. Intermodal Freight Transportation and Regional Accessibility in the United States. *Environment and Planning A*, Vol. 40, 2008, pp. 2006–2025.

91. Liu, S., and X. Zhu. Accessibility Analyst: An Integrated GIS Tool for Accessibility Analysis in Urban Transportation Planning. *Environment and Planning B*: *Planning and Design*, Vol. 31, 2004, pp. 105–124.

92. Lloyd, C. and I. Shuttleworth. Analyzing Commuting Using Local Regression Techniques - Scale and Sensitivity and Geographical Patterning. *Environment and Planning A*, Vol. 37, 2005, pp. 87–103.

93. Long, J.S. *Regression Models for Categorical and Limited Dependent Variables*. Sage Publications, Thousand Oaks, CA, 1997.

94. McCullagh, P. and J.A. Nelder. *Generalized Linear Models, Second Ed*. Chapman & Hall, London, U.K., 1989.

95. Miller H.J. and S. Shaw. *Geographic Information Systems for Transportation: Principles and Applications*. Oxford University Press Inc., New York, 2001.

96. Miller, H. J. Modelling Accessibility Using Space-Time Prism Concepts within Geographical Information Systems. *International Journal of Geographical Systems*, Vol. 5, 1991, pp. 287 – 301.

97. Miller, H.J. Measuring Space–Time Accessibility Benefits within Transportation Networks: Basic Theory and Computational Procedures. *Geographical Analysis*, Vol. 31, No.2, 1999, pp. 187–212.

98. Mohamad, D., Sinha, K. C., Kuczek, T. and C. F. Scholer. Annual Average Daily Traffic Prediction Model for County Roads. In *Transportation Research Record: Journal of the Transportation Research Board,* No. 1617, Transportation Research Board of the National Academics, Washington, D.C., 1998, pp. 69–77.

99. Morris J.M., Dumble P.M. and M.R. Wigan. Accessibility Indicators for Transport Planning. *Transport Research Part A*, Vol. 13, 1979, pp. 91–109.

100. Murray, A.T. A Coverage Model for Improving Public Transit System Accessibility and Expanding Access. *Annals of Operations Research*, Vol. 123, 2003, pp. 143–156.

101. Murray, A.T., and X. Wu. Accessibility Tradeoffs in Public Transit Planning. *Journal of Geographical Systems*, Vol. 5, 2003, pp. 93–107.

102. National Household Travel Survey (NHTS). Federal Highway Administration, Office of Policy, 2009.

103. NCDOT. *Traffic Survey Group*. Division of Highways, Transportation Planning Branch, North Carolina Department of Transportation. http://www.ncdot.org/doh/preconstruct/tpb/traffic_survey/ Accessed, December 10, 2009.

104. Neto, M.C., Jeong, Y., Jeong, M.K. and L.D. Han. AADT Prediction Using Support Vector Regression with Data-Dependent Parameters. *Expert Systems with Applications*, Vol. 36, 2009, pp. 2979–2986.

105. Neuburger, H. User Benefits in the Evaluation of Transport and Land-Use Plans. *Journal of Transport Economics and Policy*, Vol. 5, No.1, 1971, pp. 52–75.

106. Neveu, A. J. Quick-Response Procedures to Forecast Rural Traffic. In *Transportation Research Record: Journal of the Transportation Research Board*, No. 944, Transportation Research Board of the National Academics, Washington, D.C., 1983, pp. 47–53.

107. NYSDOT, *Traffic Data*, Engineering Division, New York State Department of Transportation.https://www.nysdot.gov/divisions/engineering/technical-services/highway-data-services/traffic-data. Accessed, December 10, 2009.

108. Oliver, L. N., Schuurman, N. and A.W. Hall. Comparing Circular and Network Buffers to Examine the Influence of Land Use on Walking for Leisure and Errands.

International Journal of Health Geographics, Vol. 6, No. 41, 2007, open access web document.

109. Oppenheim, N. *Urban Travel Demand Modeling: From Individual Choice to General Equilibrium*. John Wiley and Sons Inc., N.Y., 1995.

110. Ord, J. K. and A. Getis. Local Spatial Autocorrelation Statistics: Distributional Issues and an Application. *Geographic Analysis*, Vol. 27, 1995, pp 286–306.

111. Ord, J. K. and A. Getis. Testing for Local Spatial Autocorrelation in the Presence of Global Autocorrelation. *Journal of Regional Science*, Vol. 41, 2001, pp. 411–432.

112. Ortuzar, J.D. and L.G. Willumsen. *Modelling Transport*. Wiley, New York, 2001.

113. O'Sullivan, D., Morrison, A. and J. Shearer. Using Desktop GIS for the Investigation of Accessibility by Public Transport: An Isochrone Approach. *International Journal of Geographical Information Systems*, Vol. 14, 2000, pp. 85–104.

114. Paez, A. Exploring Contextual Variations in Land Use And Transport Analysis Using A Probit Model With Geographical Weights. *Journal of Transport Geography*, Vol. 14, 2006, pp. 167–176.

115. Pan, W. Akaike's Information Criterion in Generalized Estimating Equations. *Biometrics*, Vol. 57, No.1, 2001a, pp. 120–125.

116. Pan, W. On The Robust Variance Estimator in Generalized Estimating Equations. *Biometrika*, Vol. 88, No. 3, 2001b, pp. 901–906 .

117. Park, N. *Estimation of Average Annual Daily Traffic Using Geographic Weighted Regression and GIS*. Ph.D. Dissertation, Florida International University, Miami, FL, 2004.

118. Polzin, S.E. Transportation/Land-Use Relationship: Public Transit's Impact of Land Use. *Journal of Urban Planning and Development*, Vol. 125, 1999, pp. 135–151.

119. Pooler, J. A. The Use of Spatial Separation in the Measurement of Transportation Accessibility. *Transportation Research – A*, Vol. 29A, No. 6, 1995, pp. 421–427.

120. Pooler, J. and S. Saskatchewan. Measuring Geographical Accessibility: A Review of Current Approaches and Problems in the use of Population Potentials, *Geo Forum*, Vol. 18, No. 3, 1987, pp. 269–289.

121. Pulugurtha, S. S., Nambisan, S. S. and Srinivasan, N. Evaluating Accessibility and Optimizing Locations of Transit Service Facilities Using GIS. *Journal of Public Transportation*, Vol. 2, No. 4, 1999, pp. 75–94.

122. Recker, W.W., Chen, C. and M.G. McNally. Measuring the Impact of Efficient Household Travel Decisions on Potential Travel Time Savings and Accessibility Gains. *Transportation Research A*, Vol. 35, No. 4, 2001, pp. 339–369.

123. Reddy, H.K., and P. Chakraborthy. A Fuzzy Inference Based Assignment Algorithm to Estimate O-D Matrix from Link Volume Counts. Computer, Environment and Urban Systems, Vol. 22, No.5, 1998, pp. 409–423.

124. ReVille, C. and R. Swain. Central Facilities Location. *Geographical Analysis*, Vol. 2, 1970, pp. 30 – 42.

125. Ritchie, S.G. and M.E. Hallenbeck. *Statewide Highway Data Rationalization Study*. Report No. WA – RD – 83.1, Washington State Department of Transportation, Olympia, WA, 1986.

126. Robinson, R.V.F. and R.W. Vickerman. The Demand for Shopping Travel: A Theoretical and Empirical Study. *Applied Economics*, Vol. 8, 1976, pp. 267–281.

127. Roess, R.P., Prassas, E.S. and W.R. McShane. *Traffic Engineering – Third Ed.* Pearson Education Inc., New Jersey, 2004.

128. Safwat K.N.A. and T.L. Magnanti. A Combined Trip Generation, Trip Distribution, Modal Split, and Trip Assignment Model. *Transportation Science*, Vol. 22, 1988, pp. 14–30.

129. Safwat, K.N.A. and T.L. Magnanti. A Combined Trip Generation, Trip Distribution, Modal Split, and Trip Assignment Model. Working paper, Operations research center, Massachusetts Institute of Technology, Cambridge, MA, 1982. http://dspace.mit.edu/bitstream/handle/1721.1/5174/OR-112-82.pdf?sequence=1 Accessed, Oct. 29, 2010.

130. Safwat, K.N.A., and M.K. Hasan. Predicting International Freight Flows for Trade - Simultaneous Multimodal, Multi Commodity, Network Equilibrium Model. In *Transportation Research Record: Journal of the Transportation Research Board*, No. 1882, Transportation Research Board of the National Academics, Washington, D.C., 2004, pp. 129–139.

131. Sathisan, S.K., and N. Srinivasan. Evaluation of Accessibility of Urban Transportation Networks. In *Transportation Research Record: Journal of the Transportation Research Board,* No. 1617, Transportation Research Board of the National Academics, Washington, D.C., 1998, pp. 78–83.

132. Savigear, F. A Quantitative Measure of Accessibility. *The Town Planning Review*, Vol. 38, No.1, 1967, pp. 64–72.

133. Schrank, D. and T. Lomax. *The 2007 Urban Mobility Report*. Texas Transportation Institute and Texas A&M University System, College Station, TX, Sep. 2007.

134. Sharma, S.C. and Y. Leng. Seasonal Traffic Counts for a Precise Estimation Of AADT. *ITE Journal*, Vol. 64, No. 7, 1994, pp. 21–28.

135. Sharma, S.C., Gulati, B.M. and S.N. Rizak. Statewide Traffic Volume Studies and Precision of AADT Estimates. *Journal of Transportation Engineering*, Vol. 122, No. 6, 1996b, pp. 430–439.

136. Sharma, S.C., Kilburn, P. and Y. Wu. The Precision of AADT Volume Estimates from Seasonal Traffic Counts: Alberta Example. *Canadian Journal of Civil Engineering*, Vol. 23, No. 1, 1996a, pp. 302–304.

137. Sharma, S.C., Lingras, P., Liu, G.X. and F. Xu. Estimation of Annual Average Daily Traffic on Low-Volume Roads Factor Approach versus Neural Networks. In *Transportation Research Record: Journal of the Transportation Research Board*, No. 1719, Transportation Research Board of the National Academics, Washington, D.C., 2000, pp. 103–111.

138. Sharma, S.C., Lingras, P., Xu, F. and G.X. 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 Academics, Washington, D.C., 1999, pp. 24–31.

139. Shen, L.D., Zhao, F., and D. Ospina. *Estimation of Annual Average Daily Traffic for Off-System Roads In Florida*. Research Report, Florida Department of Transportation, Tallahassee, FL, 1999.

140. Siegel, J.D., De Cea, J., Fernandez, J.E., Rodriguez, R.E., and D. Boyce. Comparisons of Urban Travel Forecasts Prepared with the Sequential Procedure and a Combined Model. *Network Spatial Economics*, Vol. 6, 2006, pp. 135–148.

141. Smith, B.L., and Demetsky, M.J. Traffic Flow Forecasting: Comparison of Modeling Approaches. *Journal of Transportation Engineering*, Vol. 123, No. 1, 1997, pp. 261–266.

142. Song, S. Some Tests of Alternative Accessibility Measures: A Population Density Approach. *Land Economics*, Vol. 72, No. 4, 1996, 474–482.

143. SPSS®. SPSS Release 16.0 for Windows. Chicago, IL: SPSS, Inc, 2008.

144. SQL Server 2005®. © 2005 Microsoft Corp., Seattle, WA.

145. Stamatiadis, N. and D.L. Allen. Seasonal Factors Using Vehicle Classification Data. In *Transportation Research Record: Journal of the Transportation Research Board*, No. 1593, Transportation Research Board of the National Academics, Washington, D.C., 1997, pp. 23–28.

146. Stewart, J. Q. Demographic Gravitation: Evidence and Applications. *Sociometry*, Vol. 11, No. 1-2, 1948, pp. 31–58.

147. Stewart, J.Q., and W. Warntz. Physics of Population Distribution. *Journal of Regional Science*, Vol. 1, 1958, pp. 99–123.

148. Stopher, P.R. and A.H. Meyburg. *Urban transportation modeling and planning*. Lexington Books, D.C. Health and Company, Lexington, Massachusetts, 1975.

149. Tabachnick, B.G. and L.S. Fidell. *Using Multivariate Statistics*. HarperCollins College Publishers, New York, N.Y., 1996.

150. Tang, Y. F., Lam, W. H. K. and L. P. Pan. Comparison of Four Modeling Techniques for Short-Term AADT Forecasting in Hong Kong. *Journal of Transportation Engineering*, Vol. 129, No. 3, 2003, pp. 223–329.

151. Thill, J-C., and Kim, M. Trip Making, Induced Travel Demand and Accessibility. *Journal of Geographical Systems*, Vol. 7, 2005, pp. 229–248.

152. Toregas, C., Swain, R, ReVelle, C. and L. Bergman. The Location of Service Facilities. *Operations Research*, Vol. 19, 1971, pp. 1363–1373.

153. Toregas,C. and ReVelle, C. Optimal Location under Time or Distance Constraints. *Papers of the Regional Science Association*, Vol. 28, 1972, pp. 133–142.

154. *Traffic Monitoring Guide, Third Edition*. U.S. Department of Transportation (USDOT), Federal Highway Administration (FHWA), Washington, D.C., 1995.

155. *Traffic Monitoring Guide*. Office of Highway Policy Information, United States Department of Transportation, Federal Highway Administration, 2001. http://www.fhwa.dot.gov/ohim/tmguide/index.htm Accessed December 10, 2009.

156. Van Zuylen, H. J., and L.G. Willumsen. The Most Likely Trip Matrix Estimated from Traffic Counts. *Transportation Research Part B*, Vol. 14B, 1980, pp. 281–293.

157. VDOT, *Traffic Data*, Traffic Engineering Division, Virginia Department of Transportation. http://www.virginiadot.org/info/ct-TrafficCounts.asp. Accessed, December 10, 2009.

158. Wachs, M. and T.G. Kumagai. Physical Accessibility as a Social Indicator. *Socio-Economic Planning Science,* Vol. 6, 1973, pp. 357–379.

159. Wang, X. and M.K. Kockelman. Forecasting Network Data: Spatial Interpolation of Traffic Counts Using Texas Data. In *Transportation Research Record: Journal of the Transportation Research Board,* No. 2105, Transportation Research Board of the National Academics, Washington, D.C., 2009, pp. 100–108.

160. Wardrop, J. Some Theoretical Aspects of Road Traffic Research. Proceedings of the *Institute of Civil Engineers*, Vol. 1, No. 2, 1952, pp. 325–378.

161. Wedderburn, R.W.M. Quasi-likelihood Functions, Generalized Linear Models, and the Gauss-Newton Method. *Biometrika*, Vol. 61, No. 3, 1974, pp. 439–447.

162. Weibull, J.W. An Axiomatic Approach to the Measurement of Accessibility. *Regional Science and Urban Economics*, Vol. 6, 1976, pp. 357–379.

163. Williams, P.A. The Influence of Residential Accessibility on Household Trip-Making. *Socio-Economic Planning Science,* Vol. 23, No. 6, 1989, pp. 373–385.

164. Xia, Q., Zhao F., Chen Z., Shen L. D., and D. Ospina. Development of a Regression Model for Estimating AADT in a Florida County. In *Transportation Research Record: Journal of the Transportation Research Board*, No. 1660, Transportation Research Board of the National Academies, Washington, D.C., 1999, pp. 32–40.

165. Yang, C. and A. Chen. Sensitivity Analysis of the Combined Travel Demand Model with Applications. *European Journal of Operational Research*, Vol. 198, 2009, pp. 909–921.

166. Zahavi, Y. *The UMOT project*. Report No. DoT-RSPA-DPB-20-79-3, US Department of Transportation, Washington D.C. and Ministry of Transport, Federal Republic of Germany, Bonn, 1979.

167. Zhao, F. and S. Chung. Contributing Factors of Annual Average Daily Traffic in a Florida County. In *Transportation Research Record: Journal of the Transportation Research Board*, No. 1769, Transportation Research Board of the National Academies, Washington, D.C., 2001, pp. 113–122.

168. Zhao, F., and N. Park. Using Geographically Weighted Regression Models to Estimate Annual Average Daily Traffic. In *Transportation Research Record: Journal of the Transportation Research Board*, No. 1879, Transportation Research Board of the National Academies, Washington, D.C., 2004, pp. 99-107.

169. Zhao, F., Chow, L-F. and M-T. Li. *A Transit Ridership Model Based on Geographically Weighted Regression and Service Quality Variables.* Final Report

DO97591. Prepared for the Office of Public Transportation, Florida Department of Transportation, Tallahassee, Florida, 2005.

170. Zhou, Z., Chen, A. and S.C. Wong. Alternative Formulations of a Combined Trip Generation, Trip Distribution, Modal Split and Trip Assignment Model. *European Journal of Operational Research*, Vol. 198, 2009, pp. 129 –138.