MODELING TRAVEL DEMAND AND CRASHES AT MACROSCOPIC AND MICROSCOPIC LEVELS

by

Venkata Ramana Duddu

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

2012

Approved by:

____
Dr. Srinivas S. Pulugurtha

____
Dr. Martin R. Kane

____
Dr. Rajaram Janardhanam

____
Dr. Edd Hauser

____
Dr. Shen-En Chen

____
Dr. Harish Cherukuri

ABSTRACT

VENKATA RAMANA DUDDU. Modeling travel demand and crashes at macroscopic and microscopic levels. (Under the direction of DR. SRINIVAS S. PULUGURTHA)

Accurate travel demand / Annual Average Daily Traffic (AADT) and crash predictions helps planners to plan, propose and prioritize infrastructure projects for future improvements. Existing methods are based on demographic characteristics, socio-economic characteristics, and on-network (includes traffic volume) characteristics. A few methods have considered land use characteristics but along with other predictor variables. A strong correlation exists between land use characteristics and these other predictor variables. None of the past research has attempted to directly evaluate the effect and influence of land use characteristics on travel demand/AADT and crashes at both area and link level. These land use characteristics may be easy to capture and may have better predictive capabilities than other variables. The primary focus of this research is to develop macroscopic and microscopic models to estimate travel demand and crashes with an emphasis on land use characteristics.

The proposed methodology involves development of macroscopic (area level) and microscopic (link level) models by incorporating scientific principles, statistical and artificial intelligent techniques. The microscopic models help evaluate the link level performance, whereas the macroscopic models help evaluate the overall performance of an area. The method for developing macroscopic models differs from microscopic models. The areas of land use characteristics were considered in developing macroscopic models, whereas the principle of demographic gravitation is incorporated in developing microscopic models. Statistical and back-propagation neural network (BPNN) techniques are used in developing the models.

The results obtained indicate that statistical and neural network models ensured significantly lower errors. Overall, the BPNN models yielded better results in estimating travel demand and crashes than any other approach considered in this research. The neural network approach can be particularly suitable for their better predictive capability, whereas the statistical models could be used for mathematical formulation or understanding the role of explanatory variables in estimating AADT. Results obtained also indicate that land use characteristics have better predictive capabilities than other variables considered in this research. The outcomes can be used in safety conscious planning, land use decisions, long range transportation plans, prioritization of projects (short term and long term), and, to proactively apply safety treatments.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1: INTRODUCTION

Road transportation being a major means of transport in transportation sector plays a vital role on both economy and environment. Rapid growth in population over the past two decades has led to an increased travel demand resulting in congestion, safety and environmental issues. Economic losses due to traffic congestion and crashes are noteworthy. According to “The 2007 Urban Mobility Report”, the cost of congestion due to delays and resulting fuel wastage in the year 2005 is around $78 billion. Due to congestion, travelers have burnt 2.9 billion gallons of gas and wasted 4.2 billion hours on the roads. As traffic and congestion increases with growth in population, the conflicts that arise because of human interaction, off- and on-network characteristics also increase. According to the National Highway Traffic Safety Administration (NHTSA), more than 5.8 million reported crashes occurred in the United States during 2008 (NHTSA, 2008). Over 37,000 people were killed and 2.55 million people were injured in these crashes. Federal agencies have made reducing crashes and improving safety on roads as one of their top priorities (Federal Highway Administration-FHWA, 2006). Also, the present transportation infrastructure is inadequate to accommodate the current needs of the users, which are expected to worsen in the future.

Influencing land use characteristics by considering appropriate combinations and percentages of various land use categories and rezoning has significant potential to effect trip productions and attractions, and provide better mobility on roads. Understanding the causes of the congestion and crashes, identifying appropriate solutions, and, proactively adopting or using them at macroscopic (area) or microscopic (link) levels can affect travel patterns and traffic safety. Further, an accurate forecast of travel demand and crash prediction helps planners to plan, propose and prioritize infrastructure projects for future improvements.

1.1 Problem Statement

Data collected from the field and outputs from well calibrated traditional four-step planning process are potential sources of traffic volume at link level. While the data from field are collected based on the need or as a part of traffic count programs and are available for selected links / intersections, traffic volumes are available as outputs for all major roads from traditional four-step planning process. Resource constraints (include funds) often limit practitioners from collecting traffic volume data for all links / intersections in the transportation network but rely on outputs from calibrated traditional four-step planning process.

Traffic volumes from calibrated traditional four-step planning process depend on estimated trip productions and trip attractions of each traffic analysis zone (TAZ) in the study area. TAZ is defined as an area delineated by state transportation officials to evaluate and tabulate trip productions and attractions for their use in transportation planning models. The trips produced from or attracted to a TAZ are generally estimated as a function of demographic / socio-economic characteristics. These demographic and socio-economic characteristics to estimate trip productions and trip attractions at TAZ level depend on land use characteristics (example, population of a TAZ depends on the number of dwelling / household units and the average household size of the TAZ).

Further, infrastructure (on-network characteristics) related decisions often depend on traffic characteristics which as stated above depend on socio-economic / demographic / land use characteristics.

As trips produced from and attracted to a TAZ increases, the traffic volume on links in the TAZ increases. This is naturally expected to result in an increase in the number of crashes. Land use planning decisions can therefore influence demographic / socio-economic characteristics, and hence traffic volume and crashes both at area level and link level.

Most of the studies in the past are based on the data within the vicinity of a location. Crash estimation models developed based on data within the vicinity of a location could help improve safety at the specific location. Typical location- or site-specific improvements are related to engineering, education and enforcement, and are generally short-term in nature. Aggregating the transportation system at the TAZ level and developing TAZ level crash estimation models as a function of land use characteristics helps reduce some of the difficulties caused by ‘lumpiness’ of random events that one see across intersections or across road segments (Washington, et al., 2006). However, TAZ level models for estimating traffic volumes, that is, trip attractions and trip productions is not required as they are primarily used to evaluate link level traffic volumes.

While macroscopic or area level models assist in planning for better future (new developments or by rezoning), microscopic or link level models assist in design and operational analysis. However, traffic volume data is not available for all the links in the network. Developing link-level travel demand and crash estimation models will help evaluate the effect of design and operational related changes on mobility and safety. The outcomes can be proactively adopted to achieve better performance on the transportation network. While resource limitations restrict collecting traffic volume data for every link in the network, lack of link-level crash estimation models that rely on variables other traffic volume or travel demand is another limitation.

It is well known in the literature that the effect of an area (land use characteristic) on the road link decreases with the increase in distance (Principle of Demographic Gravitation: Stewart, 1948). Evaluating link level traffic and hence crashes as a function of land use characteristics based on distance gradient method would not only improve the accuracy of the models in estimating travel demand but also helps better understand the role of safety in long range transportation planning and land use planning decisions. It would minimize dependence on analyzing crash data, identifying causes / solutions, and implementing countermeasures as an afterthought. Therefore, spatial variations of land use characteristics based on gradient distance method that decrease with an increase in distance should be incorporated in the estimation process.

1.2 Research Goal and Objectives

The primary goal of this research is to develop methods and assess models to accurately estimate travel demand and crashes at macroscopic and microscopic levels for transportation planning, operational and safety analysis and decision-making. Several objectives were identified to achieve this goal. They are:

- Identify the characteristics that explain travel demand and crashes.
- Investigate spatial autocorrelation to evaluate whether there is any influence of crashes in a TAZ on its adjacent TAZ’s.

- Incorporate spatial variations of land use characteristics based on gradient distance method that decrease with an increase in distance.
- Develop the following models based on statistical techniques and artificial neural network techniques.
    - Macroscopic (TAZ) level crash estimation models with an emphasis on land use characteristics.
    - Microscopic (link) level travel demand and crash estimation models to estimate link level traffic volume and crashes incorporating the Principle of Demographic Gravitation.
- Compare the performance of the statistical models with neural network models.
- Validate outcomes and demonstrate the applicability of models to estimate travel demand and crashes.

1.3 Organization of the Dissertation

The dissertation is organized as follows. A review of past literature, underlying methods and their limitations are discussed in Chapter 2. A brief description of neural networks, how it effectively helps solve the problem and their applications in the area of transportation are discussed in Chapter 3. The proposed new research methodology and the databases required to incorporate it are discussed in Chapter 4. Chapter 5 and Chapter 6 discuss the development of macroscopic and microscopic models, and also present their predictive performance to verify/validate the models developed. Conclusions and directions for future research are presented in Chapter 7.

CHAPTER 2: LITERATURE REVIEW

A discussion of past related literature is presented in this chapter. In Section 2.1, a review of various methods used in estimating crashes is discussed. In Section 2.2 and 2.3, a review of travel demand modeling and traditional urban transportation planning models is presented. In Section 2.4, a review of the studies and practices in AADT estimation is described. The motivation of this research is discussed in Section 2.5.

2.1 Crash Estimation

Researchers have examined the role of demographic / socio-economic, land use and on-network characteristics on crashes, and have developed crash estimation models in the past. Poisson models, Negative Binomial models, linear regression models, and empirical analyses techniques were used to develop such models. A coherent discussion on the effect of these characteristics on crashes and techniques used is presented next.

Levine et al. (1995), based on data for Honolulu, Hawaii, found that most of the crashes occur closer to employment centers than residential areas. Crashes in rural and sub-urban areas are most likely to be fatal or serious injury crashes that can be related to night-time driving and alcohol.

Abdel-Aty et al. (2000) studied the relationship between crashes on principal arterials and on-network characteristics, user characteristics (age and gender) and roadway geometry. The Negative Binomial model developed indicated that average AADT, degree of horizontal curves, and lane, shoulder and median widths play a significant role on the frequency of crashes. The results obtained also indicated that female, younger and older drivers experience more crashes on roads with heavy traffic volumes and lower lane, shoulder and median widths than when compared to male drivers and mid-aged drivers. The tendency of younger drivers to be involved in crashes was observed to be higher on roadway curves and while speeding.

Ivan et al. (2000) explained highway crash rates by predicting single-vehicle and multi-vehicle crashes separately using Poisson regression models. The results obtained indicate that daytime, volume-capacity ratio, segments with no passing zones, the shoulder width, the number of intersections and driveways are significant variables that can explain single-vehicle crashes, whereas daylight conditions, the number of intersections and driveways are significant variables that can explain multi-vehicle crashes.

Kim et al. (2002) evaluated crash patterns based on land use characteristics using empirical analysis and geographical information systems (GIS). The results obtained indicate that residential neighborhoods, which have higher traffic volumes only during the peak hours tend to have higher crashes than commercial centers that have high traffic volumes throughout the day.

Wood (2002) described the underlying mechanism of generalized linear crash models and practical resolution to address the ‘low mean value’ problem. Noland et al. (2003) examined road casualties in England using Negative Binomial models based on land use and road characteristics. The results obtained indicate that traffic casualties are higher in areas with higher levels of social deprivation and in areas with higher employment density. However, urbanized areas with high density of population tend to have fewer traffic casualties.

Greibe (2003) researched on crash estimation models for urban junctions and links using generalized linear modeling technique. The results obtained indicate that road environment variables, minor side roads, parking facilities and speed limits are significant variables to explain crashes on road links. Only vehicle traffic flow was found to be a significant variable for road junctions.

Ladron de Guevara et al. (2004) forecasted crashes using Negative Binomial models and concluded that population density, intersection density, the number of employees and traffic volume play a significant role in predicting crashes. The study suggested that “planning-level safety models are feasible and may play a role in future planning activities”.

Noland et al. (2004) examined that an increase in the number of lanes and road widths lead to increased traffic-related crashes. This study concluded that an increase in shoulder width may help in decreasing traffic-related crashes. The changes in vertical and horizontal curvatures have no statistical association with traffic safety.

Wood (2005) explained errors that were mainly related to crash rates and variables using generalized linear models with logarithmic function. Aguero-Valverde et al. (2006) used injury and fatal crash data for Pennsylvania and compared full bayes hierarchical models and Negative Binomial models. The study revealed that variables that are highly significant in Negative Binomial models are equally significant in full bayes hierarchical models. However, marginally significant variables in Negative Binomial models are not significant in full bayes models. It was concluded that the counties with higher percentages of the population below the poverty line, with younger and older drivers, and with increased road mileage and density have significantly increased crash risks.

Kim et al. (2006) explored the relationship between land use, population, the employment by sector, economic output and motor vehicle crashes. The authors indicate that new development or increasing the intensity or changing the nature of existing economic activity will have implications on safety. The results from the Negative Binomial models developed indicate that land use characteristics such as parks, schools and commercial areas are highly associated with crashes. Similarly, Kim et al. (2010) examined relationship between demographic, land use, and roadway accessibility variables and types of crashes in Honolulu, Hawaii using binomial logistic regression. The authors concluded that demographic variables such as job count and number of people living below the poverty level are significantly associated with injury crashes and pedestrian and bike crashes. Accessibility measures such as the number of bus stops and the number of intersections are associated with increases in all types of crashes. Land use characteristics such as Business and commercial areas are strongly associated with increased total as well as injury and fatal crashes.

Caliendo et al. (2007) used Poisson, Negative Binomial and negative multinomial regression models separately to predict crashes on tangents and curves on multilane roads. Regression models were developed separately for total crashes, fatal crashes and injury crashes. Length, longitudinal slope, sight distance, curvature, side friction coefficient and, AADT were considered as the explanatory variables to predict crashes. This study concluded that length, curvature and AADT are significant variables that can explain crashes on curves, whereas length, AADT and junctions are significant variables that can explain crashes on tangents.

Mitra and Washington (2007) evaluated the nature of over-dispersion in crash estimation models. This study was motivated to corroborate the findings of Miaou and Lord (2003) regarding the variance structure in over-dispersed crash models. Four geometric factors and four traffic flow explanatory variables were considered in developing the crash estimation models. These models were compared using significance of coefficients, standard deviance, chi-square goodness of fit, and deviance information criteria statistics.

Ma et al. (2008) developed multivariate Poisson lognormal regression (MVPLN) models using crash data for Washington State roadway segments. Crash data was used as a dependent variable, while roadway geometric characteristics were used as independent variables.

Quddus et al. (2008) developed relationships between traffic casualties and traffic characteristics, road characteristics and socio-demographic characteristics using both non-spatial Negative Binomial models and spatial Bayesian hierarchical models. Area or ward (census track) level data was used to evaluate the correlations. This study concluded that results from both non-spatial and spatial models are quite similar in many cases. The Bayesian hierarchical models developed indicate that casualties increase with traffic flow, and households with no cars and total employment are statistically significant variables in all the models.

Wier et al. (2009) developed simple bivariate models to predict the changes in vehicle-pedestrian injury collisions based on changes in traffic volume on highways and freeways in San Francisco, California. They used street characteristics (traffic volume, the number of intersections, the percentage of residential streets, the percentage of arterial streets without public transit, the percentage of arterial streets with public transit, and, the percentage of freeways and highways), land use characteristics (percentages of commercial, industrial, neighborhood commercial, residential, higher density residential, and residential neighborhood commercial areas in square miles), population characteristics (employee population, resident population, the percentage of population age 65 and older, the percentage of population age 17 and under, the percentage of population living below the poverty line, the percentage of population unemployed), commute behavior (the percentage of workers commuting to work by walking and the percentage of workers commuting to work by public transit) to develop multivariate area-level regression model for vehicle- pedestrian injury collisions from 2001-2005. Ordinary squares regression (OLS) was used to model the natural log of the number of vehicle-pedestrian injury collisions over a 5-year period. Traffic volume was observed to have the highest adjusted partial correlation with vehicle-pedestrian collisions, followed by the number of employees, the proportion of neighborhood commercial area, the proportion of arterial streets without transit and the residential population.

Naderan et al. (2010) have developed crash generation models for total, property damage, severe, injury, and fatal crashes separately using trip productions and attractions of each TAZ.

Ukkusuri et al. (2011) investigated the role of built environment on pedestrian crash frequency at census tract level. The results from the study indicate that census tracts with greater fraction of industrial, commercial, and open land use types have greater likelihood for pedestrian crashes while tracts with a greater fraction of residential land use have significantly lower likelihood of pedestrian crashes. The authors also concluded that as schools and transit stops are determinants of pedestrian activity, census tracts with greater number of schools and transit stops are more likely to have greater pedestrian crashes.

A TAZ level non-parametric safety analysis by Siddiqui et al. (2012) indicated that the total number of intersections per TAZ, airport trip productions, light truck productions, total roadway segment length with 35 mph posted speed limit, total roadway length with 15 mph posted speed limit, total roadway length with 65 mph posted speed limit, and non-home based work productions are significant variables in predicting total crashes. Whereas, total number of intersections per TAZ, light truck productions, total roadway length with 35 mph posted speed limit, and total roadway length with 65 mph posted speed limit are identified as significant variables for severe crashes.

2.2 Travel Demand Modeling

Travel demand modeling means the development of mathematical formulations which represent travel patterns on the road links of a transportation network. Travel demand modeling is an integral part of the planning process which helps predict the traffic volumes, flows on road links of transportation network and transit patronage in the future by considering socio-economic characteristics, demographic characteristics, land use characteristics and any new developments or transportation infrastructure projects in a region. The basic objective of this process is to provide a comprehensive and continuing guidance to the Metropolitan Planning Organizations (MPO’s) for the development, evaluation and implementation of future transportation planning proposals, policies and in prioritizing projects to allocate available funds for future investments. In practical terms, the purpose of travel demand modeling is to provide a tool which helps predict, or forecast travel patterns under various conditions which represent the state of transportation network at a future time.

Presently, a majority of the MPO's use a traditional sequential four-step model (FSM) to estimate and forecast traffic on the road network. Incorporated in 1950's when the first transportation study was conducted in Detroit, Michigan and subsequently by many metropolitan areas in the United States and developed European countries, the FSM has progressively evolved into an established methodology for predicting and forecasting traffic over the past fifty to sixty years. The following section discusses method and models that are used in practice in the traditional sequential FSM.

2.3 Traditional Urban Transportation Planning Models

The traditional FSM for urban transportation planning consists of the steps: trip generation, trip distribution, mode split and traffic assignment. A significant amount of data is required for FSM to define travel and transportation systems. The data needs include graphical representation of the transportation network with road links and nodes. The travel or activity data needed for the FSM is gathered by surveys and is typically aggregated to a zonal level of convenient size. These zones are typically termed as traffic analysis zones (TAZ).

As the level of aggregation has a significant effect on the results and ultimately the policy measures, defining TAZs play an important role in modeling travel demand. The selection of TAZ size and number depends on several factors such as socio-economic, demographic and land use characteristics and also on the project objectives and the type of study conducted. It is assumed that all the attributes of each TAZ in the study area are represented by its respective centroids. So, 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. The trips that are expected from the areas out of the study area are modeled as external zones or stations, which are connected to the network on its periphery. A brief overview of each step in the traditional FSM process is discussed below.

2.3.1 Trip Generation

The process of estimating the total number of trips produced or attracted by each TAZ is known as trip generation. There are two kinds of trip generation models: trip production model and trip attraction model. Trip production models estimate the number of home-based trips to and from each TAZ and where trip makers reside. Trip attraction models estimate the number of home-based trips to and from each TAZ at the non-home end of the trip. Different production and attraction models are used for each trip purpose. Special generation models are used to estimate non home-based, truck, taxi, and external trips. As the trips produced in an origin should have a destination, all the productions in origin TAZs should be balanced with trip attractions in the destination TAZs to ensure that total trip productions and attractions are equal. The number of trips produced in a TAZ depends on population size and density, household size, 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. 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. Trip generation is generally carried out using two methods: linear regression analysis and the cross classification/category analysis (Ortuzar and Willumsen, 2001).

2.3.2 Trip Distribution

Trip distribution is a process of distributing the trips generated among the destination TAZs. Among the various to model trip distribution, the gravity model adapted from Newton's gravitational law of physics is commonly used in trip distribution process. According to gravity method, the trip attractions between origin-destination TAZs diminish with an increase in the distance between the TAZs. According to Stopher and Meyburg (1975), growth rates observed from the historical data and experience were also generally applied for trip distribution. However, temporal variations of trips, special attractions, and future developmental attractions are not considered in the traditional gravity model. The gravity model need travel time matrices for inter and intra-zonal trips for both base year and forecast year. However, 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".

2.3.3 Mode Split

Mode split is the process of splitting or distributing the origin destination volumes by available alternate modes. The model split is typically performed after trip distribution, though in some cases it is performed after trip generation and before distribution. 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).

2.3.4 Trip Assignment

Trip assignment is the process of assigning the trips from a given origin to a given destination on a given mode obtained from the previous step to routes comprising of a set of links in the network. The trip assignment process is 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. Equilibrium principles developed by Wardrop (1952) are used for trip assignment or network assignment in congested networks. Wardrops's first and second principles, the user optimal and system optimal principles 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 Combined Four Step Methods

Combined models were developed to carry out all the four steps simultaneously in order to reduce errors and uncertainties transferred in the traditional FSM approach. 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).

2.4 Estimation of AADT

Sharma et al. (1993) developed an index of assignment effectiveness to evaluate the duration and frequency of control of seasonal counts. The authors concluded that a 1-week count repeated in 4 different months was much more accurate than a 1-week count repeated twice. However, repetition more than 4 times would contribute little additional improvement. Similarly, Sharma et al. (1994) investigated the problem of determining the duration and timing of a seasonal count given a specified precision. As an extension to these studies, Sharma et al. (1996) 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. AADT values were calculated using respective expansion factors of the ATR group by assigning SPTC sites to homogeneous ATR groups. Appropriateness of volume adjustment factors was expressed in terms of assignment effectiveness and was used to represent the degree of correctness in assigning the sample sites to a given ATR group. The need for effective assignment of count sites was discussed and found that estimates of a properly assigned 6 hour counts proved better than the improperly assigned 72 hour count sites. However, during the last decades, control counts became more and more unpopular in the United States and the newest edition of the Traffic Monitoring Guide (TMG, 2001) has left out the suggested use of control counts.

Smith et al. (1997) focused on developing traffic volume forecasting models for two sites on Northern Virginia Capital Beltway. Four models such as historical average, time-series, neural network and nonparametric regression models were developed to estimate freeway traffic flow that represents 15 minute future traffic volume. From Wilcoxon signed-rank test conducted, the authors reveal that the nonparametric models which are easy to implement and proved to be portable, experienced significantly lower errors that any other model tested. A similar study was done by Smith et al. (2002) to compare the performance of parametric and nonparametric regression models using seasonal autoregressive integrated moving average (ARIMA) for traffic flow forecasting. The results indicate that the traffic condition data is characteristically stochastic as opposed to chaotic. Their research concluded that larger databases would provide a better set of neighbors to use in producing forecasts and in addition different state definitions and/or distance metrics may lead to better results.

Stamatiadis et al. (1997) studied the relationships and developed seasonal adjustment factors for the state of Kentucky to understand the relationship between the data obtained in a short-term period to those for the entire year. The preliminary analysis indicated that seasonal adjustment factors are essential in developing accurate estimates of traffic volumes for each vehicle type, and their use can improve the estimation of daily volumes.

Granato (1998) presented an analysis by utilizing data from an (automatic traffic recording) ATR station maintained by the Iowa DOT in Cedar Rapids, Iowa. The analysis was to determine how much day of week/month of year factors can reduce the error of prediction of AADT from a short-term traffic count. Results indicate that continuous consecutive day count improved the estimation only by 5%. Using a single ATR count station data, a 25% error reduction in the AADT estimates was found with the application of day of week and month of the year factors when compared to using continuous 24 hour. The author suggested using multi day traffic counts scattered across two or three weeks over consecutive-day counts.

Mohamad et al. (1998) developed AADT prediction model for county roads. Traffic data was collected using automatic traffic counters for 40 counties out of 92 in Indiana State which were selected based on population, state highway mileage, per capita income, and the presence of interstate highways. AADT was calculated using factors and multiple regression analysis was conducted to develop models. County population and county arterial mileage, location and accessibility were found to be the significant factors affecting the daily traffic on paved county roads. Similarly, Xia et al. (1999) attempted to estimate AADT for non-state roads that do not have traffic counts in Broward County, Florida with more predictor variables in regression analysis. The authors used predictor variables such as roadway characteristics such as the number of lanes, area type and functional classification, socio-economic data variables such as 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 were found to be significant predictors to estimate AADT.

Horowitz and Farmer (1999) have reviewed travel forecasting practices that were being undertaken by many states in the United States. The review consisted of interviews and documents from 45 states and articles from the past literature which included passenger models, freight models and time-series models. The statewide models were compared to intercity models and found that in spite of the inherent differences in scales, planning needs, and data availability, states with complete models tended to follow an urban modeling framework and used software originally designed for urban travel forecasting. Recommendations were provided for improved statewide travel forecasting.

Sharma et al. (2000) proposed a neural network approach for estimating AADT from 48-hour coverage counts. The authors carried out a detailed comparison between the neural network approach and the traditional method by using data from ATR-equipped segments in Minnesota. The results showed that the traditional method produced better AADT estimates than the neural network approach for a single 48-hour coverage count when it was correctly assigned to a factor group. The error for two 48-hour counts using the neural network approach was comparable to that for only a single 48-hour count using the traditional method. The two 48-hour counts from different months were used by the neural network approach. The authors concluded that the error could be much higher for coverage count locations assigned incorrectly to factor groups in practice when using the traditional method. The neural network approach was extended to estimate AADT on low-volume roads by Sharma et al. (2001).

Seaver et al. (2000) estimated traffic volumes on the rural roads using statistical techniques. The average daily traffic (ADT) on the rural roads was modeled based on the road type using the data related to 80 counties in Georgia State. Forty five variables related to 8 categories such as population demographics (population density, population percentage changes, persons per household), education, transportation (travel time to work, means of transportation to work, leaving time for work), income (per capita income and median household income), employment (types of employment, unemployment, and place of employment (in or out of the county or state)), farming, urbanization and housing are considered as the predictors of ADT. The authors have developed several regression equations and suggested to classify the county and then choose an appropriate model to predict ADT.

Lingras et al. (2000) applied time series analysis for predicting daily traffic volumes. The analysis is applied based on different types of road groups according to trip purpose and trip length distribution. The study involved comparison of statistical and neural network techniques for time series analysis in predicting daily traffic volumes. The authors concluded that neural network models perform better than auto regression models and the prediction errors for predominantly recreational roads were higher than those for predominantly commuter and long-distance roads supporting the fact that commuter and long-distance traffic patterns are relatively more stable than recreational traffic patterns.

Zhao and Chung (2001) extended the study by Xie et al. (1999) using a larger data set that included all state roads in Broward County. The predictor variables were updated by analyzing land-use and accessibility variables more extensively. They presented four models with different combinations of explanatory variables considered. The authors examined and compared the predictive power of the models and suggested that the method of estimating AADT by not using traffic counts might not be adequate to meet the need of engineering design and planning. However, it could be used for tasks that do not need a high level of accuracy.

Davis and Yang (2001) developed an algorithm for computing probability of a match between a short count site, and each of a set of permanent counting stations showing distinct trends using Bayesian decision theory. The authors 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. The median or the 50th percentile of the predictive distribution using the probable ranges and their associate probability values were used to obtain total traffic volumes. The authors concluded that that for estimating mean daily traffic (MDT) by vehicle class, the samples should be taken between May and October, and between Tuesday and Thursday so that the error of Bayes estimates of classified MDT are about 10-12% on average and within 26%, 95% of time.

McCord et al. (2003) estimated AADTs from several air photos and satellite images for several highway segments in Ohio. A sequential approach of five steps was proposed to produce the AADT estimate from a single image: 1) obtain the vehicle density from the image; 2) covert the density to a volume (“t minutes”); 3) expand the t-minute volume to an hourly volume; 4) expand the hourly volume to a daily volume; 5) de-seasonalize the daily volume to produce an average yearly volume. These AADTs were compared with the AADTs from traditional ground-based estimates. As the empirical errors were small enough, the study concluded that AADT estimation errors and ground-based sampling efforts could both be reduced by combining satellite-based data with traditional ground-based data.

Tang et al. (2003) adapted time-series, neural network, non-parametric regression, Gaussian maximum likelihood to develop models for predicting traffic volumes by day of the week, by month and AADT for the entire year. Analysis was conducted based on Hong Kong’s historical traffic data from 1994-1998. The daily flows estimated by the four models were used to calculate the AADT for the year of 1999. The results from the four models were compared and the authors indicated that the Gaussian maximum likelihood model appears to be the most promising and robust of these four models for extensive applications to provide the short-term traffic forecasting database for the whole territory of Hong Kong.

Zhao and Park (2004) estimated AADT's using geographic weighted regression (GWR) technique which 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. When compared with OLS models, the GWR models were found to be more accurate and were useful for studying the effects of the regressors at different locations.

A study by Zhong et al. (2004) developed genetically designed neural network and regression models, factor models, and autoregressive integrated moving average (ARIMA) models to evaluate missing traffic counts from permanent traffic counts. The authors found that genetically designed regression models based on data from before and after the failure had the most accurate results. Average errors for refined models were found to be lower than 1% and the 95th percentile errors were below 2% for counts with stable patterns.

Li et al (2004) identified several significant factors through regression analysis that contribute to seasonal traffic patterns considering land use, demographic and socio-economic data which also include seasonal movement of seasonal residents and tourists, retired people between ages 65 and 75 with high income, and retail employment. The results indicate a possible way to directly estimate the seasonal factors for short-period counts based on land use variables and the possibility of assigning established seasonal groups to short-period counts based on similarity in land use, demographic and socio-economic characteristics. However, 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. Based on known seasonal factors groupings of permanent count stations and their four land use categories, the authors developed a fuzzy tree construct to determine the seasonal factors 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) developed a method that exploits the correlations between 24 hour segment volumes. The method can be used with only two daily traffic volumes on the coverage count segment and is based on generalized least squares estimation of AADT, rather than on ordinary least squares estimation, which is traditionally used. Monte Carlo simulation on a small network representing intercity flows in Ohio was used to compare the performance of the correlation-based method with that of the traditional method. Results showed that the correlation-based method resulted in less error in AADT estimation than the traditional method when the segment volumes were highly correlated. However, when these segment volumes had low correlation, both the methods showed similar performance, providing estimates of approximately equal magnitude with approximately equal frequency.

Jiang (2005) and Jiang et al. (2006) developed a method to estimate AADT by combining the ground-based traffic data information with in-image traffic data information. In this method, the weighted combination of earlier year coverage counts and current year image containing traffic data were is proposed for the AADT estimation.

An empirical study was conducted using data for 122 highway segments and single contemporary image for 10 year period from 1994 and 2003. The results indicate an increased accuracy in AADT estimation with reduction in the average error and increased likelihood of producing an AADT with percent error less than 10%.

Lam et al. (2006) developed models for short-term traffic forecasting based on historical traffic data collected for annual traffic census in Hong Kong. Non-parametric regression model and Gaussian maximum likelihood model were used for traffic forecasting. AADT’s were calculated from the predicted daily vehicular flows. The results for the prediction of models and comparison show that the non-parametric regression model produces better forecasts than the Gaussian maximum likelihood model.

Gadda et al. (2007) quantified the level of uncertainty in AADT estimates from extrapolating short-term local counts over time and space by quantifying different errors such as factoring errors, spatial errors and temporal errors. The research also explored how these errors vary by day of week, month of year, area type, functional class, number of lanes, duration and distance to nearest SPTC station. The classification of the site by fine clustering, on the basis of functional class, lane count, and multiple area types may prove very useful for the analysis. Results obtained appeared consistent across states which supports the notion of their transferability to other contexts. Results from the quantification of spatial errors indicate an increase in the errors dramatically beyond 0.5 miles from the count site in urban areas and 1 mile in rural areas.

Wang and Kockelman (2009) used Kriging-based method for mining network and count data over time and space. The study concluded that Kriging performed far better than other options for spatial extrapolation, such as assigning AADT on the basis of a point’s nearest sampling site. Similarly, Selby and Kockelman (2011) explored the application of Euclidean distance and network distance based Kriging methods for prediction of ADT counts across the Texas network. Universal Kriging was found to reduce over non-spatial regression techniques. However, errors remained quite high at the sites with low counts and/or in less measurement-dense areas. Results based on comparison indicate that the estimation of Kriging parameters by network distances showed no enhanced performance over Euclidean distances, which require less data and are much more easily computed.

2.5 Limitations of Past Research

Most researchers have focused on travel demand and crash estimation models based on demographic characteristics, socio-economic characteristics, and on-network (includes traffic volume) characteristics or on the effect of these variables on travel demand and crashes. A few researchers have considered land use characteristics but along with other predictor variables. A strong correlation may exist between land use characteristics and these other predictor variables. Further, land use characteristics may play a relatively stronger role on travel demand and crashes when compared to other predictor variables.

Not many authors have looked at area or TAZ level crash estimation models that would lower difficulties caused by ‘lumpiness’ of random events across intersections or across road segments (Washington, et al., 2006). Those that considered area level, spatial proximity or trip generation data (Naderan and Shahi, 2010; Abdel-Aty et al., 2010; An et al., 2011) for safety conscious planning did not examine the direct relationship between land use characteristics and crashes.

Similarly, literature documents various models in estimating crashes at link level. Most of those considered volume as one of the primary variable to estimate crashes at link level. However, these volumes are obtained from 1) traditional FSM in which uncertainties are transferred from the previous steps where error reduction is inevitable; 2) combined that FSM carry all the drawbacks associated with FSM, since these models are developed based on the basic assumptions associated with traditional FSM; and 3) AADT estimation models that were developed using various methods which are limited in scope. These methods include, statistical models based on area type (such as urban and rural), non-state roads and other functional classes, time series analysis based on historic traffic counts using neural networks, and density based methods.

Overall, none of the past research attempted to directly evaluate link level traffic and hence crashes as a function of land use characteristics. None of them have considered artificial intelligent (AI) techniques in directly estimating travel demand and crashes at both area and link level. These AI techniques and their applications in the area of transportation are discussed in-detail in the Chapter 3. Moreover, none of them have considered distance gradient method in estimating link level traffic and crashes. There is a need to develop a methodology which includes spatial analytical methods, statistical methods, basic scientific principles and artificial intelligent techniques. The new method should incorporate spatial variations of land use characteristics based on gradient distance method that decrease with an increase in distance in the estimation process to increase the predictability of the models.

CHAPTER 3: NEURAL NETWORKS

Artificial intelligence (AI) is an ability of a system that can independently perform tasks normally requiring human or animal intelligence (Nilsson, 1971). It is a system with an ability to learn, adapt and improve. The first known AI system is “Turing Machine” invented by Allen Turing in 1950. In the subsequent years, the research in the field of AI has grown rapidly and is sub-divided into many different areas based on their applicability to various fields in science and technology. Some of the applications that are commonly used in the field of civil engineering include Expert Systems, Genetic Algorithms, Intelligent Agents, Neural Networks, Logic Programming, and Fuzzy Logic. Each of the above mentioned applications are used based on the type of problem to be addressed.

Neural networks were chosen to develop the methods and models in this dissertation. A brief description of neural networks and how it effectively helps solve the problem are discussed in the following sections.

3.1 Neural Networks

Artificial neural networks, also called as neural networks, is a computational model that mimic at least partially the structure and functions of brains and nervous systems of living beings (Cichocki & Unbehauen, 1993). In general, a neural network is a computational model composed of simple processing elements called neurons or nodes, which are interconnected by links with weights that help perform parallel distributed processing in order to solve a desired problem. Neural networks have the ability to learn from the environment and to adapt to it in an interactive manner similar to their biological counterparts.

The interest in the neural networks has grown dramatically in the fields of science and engineering in the last few years. A basic neural network model consists of set of nodes connected by the links that has numeric weights associated with them. Each node has a set of input links from other nodes and a set of output links to other some nodes. The nodes from the input links are connected to an activation function to compute the activation level at the next time step. Figure 3.1 shows a model of an artificial neuron. The inputs to the artificial neuron are given in the form individual vector components given as x_i, for i = 1, 2, 3…, n. That is, the entire input is given as a vector signal $x \in \Re^{n\times1}$, where, $x = [x_1, x_2, x_3, …., x_n]^T$. Each input neuron '$x_i$' is connected to the neuron 'q' through a link called synapse, which is associated with a synaptic weight 'W_{qi}'. The neuron 'q' receives an input from 'x_i' as the product of the individual input vector component 'x_i' and the weight 'W_{qi}' associated with it. Since, there are multiple inputs to the neuron 'q', all these inputs are multiplied with their respective synaptic weights and then summed as $u_q = \sum_{i=1}^{n} W_{qi}x_i$. The threshold or bias (-ve of threshold) θ_q is externally applied, usually to lower the cumulative input to the activation function. The activation function shown in the Figure 3.1 helps define the output 'y_q' for a given input 'u_q'. From Figure 3.1, the output of the neuron 'q' can be written as, $y_q = f(v_q) = f(u_q - \theta_q) = f(\sum_{i=1}^{n} W_{qi}x_i - \theta_q)$. For no threshold scenario, $y_q = f(v_q) = f(u_q) = f(\sum_{i=1}^{n} W_{qi}x_i)$. The above explanation on neural networks is based on the work by Ham & Kostanic (2001).

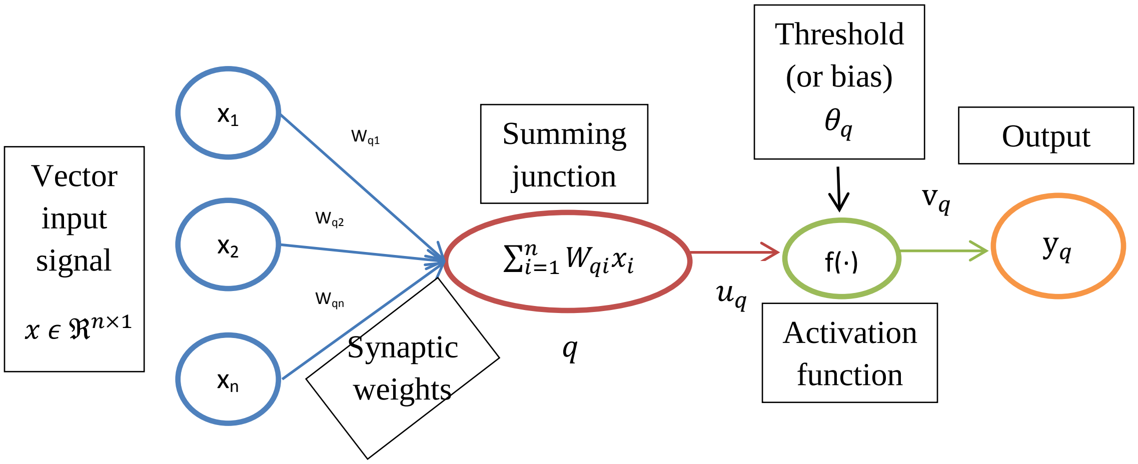

FIGURE 3.1: Nonlinear model of an artificial neuron

One of the most popular neural networks is a layered neural network with a back-propagation (BP) learning algorithm where the weights are adjusted based on least mean square error of the output. Neural networks could also be used for prediction purposes by using BP architecture. The use of BP architecture in the neural networks let the network learn an approximation of mapping (pattern) between inputs and outputs by updating its synaptic weights along with error minimization in order identify the implicit rules and relationship between the inputs and outputs. Along with the prediction, neural networks can also be used for various other applications such as optimization, forecasting, associative memory, function approximation, clustering, data compression, speech recognition, non-linear system modeling and control, pattern classification, feature extraction, solutions to matrix algebra problems and differential equations (Ham & Kostanic, 2001).

3.2 Neural Networks Applications in Transportation Area

Over the past few years, AI techniques have played a significant role in the design of sophisticated traffic management systems. Few of them include, Gilmore et al. (1993) on applications of neural networks in traffic management, Nakatsuji et al. (1994) on optimizing signal timing using AI techniques, Ledoux et al. (1995) and Yin et al. (2002) on modeling traffic flow using neural and fuzzy-neural networks, Hua et al. (1995) on intelligent traffic control systems, Smith et al. (1996), Wilde (1997), and Smith et al. (2002) on short-term traffic flow predictions using neural network approach, Ledoux(1996) on integrating neural networks and urban traffic systems, and Srinivasan et al. (2007) on intelligence based congestion prediction.

CHAPTER 4: METHODOLOGY

In this chapter, the databases and the methods used to evaluate the macroscopic relationship between land use characteristics and crashes, and microscopic relationship between land use characteristics and crashes / link level traffic volume are described. The entire methodology is divided into two parts based on macroscopic and microscopic level of analyses.

4.1 Macroscopic Models

The macroscopic models in this research are the models which help predict the dependent variable at an area level (TAZ level). In order to develop these models the database related to dependent and independent variables has to represent data at a TAZ level.

4.1.1 Database to Develop Models

The database to develop macroscopic models should contain crash data, land use data, street centerline network, and TAZ layer (with planning variables data, and trip productions and attractions) in a GIS format.

The crash data (shapefile - points) was overlaid on the TAZ layer to extract and estimate the total number of crashes and the number of crashes by severity (fatal crashes, injury crashes, and property damage only - PDO crashes) for each TAZ in GIS environment. The data related to planning variables (demographic / socio-economic characteristics which include population, number of household units and employment), trip attractions and productions for each TAZ are directly available in the GIS format. In case, if the data related to planning variables are obtained in the form of censes blocks in GIS format, the census data (shapefile – polygons) will have to be overlaid on the TAZ layer using "intersect" feature in GIS environment to estimate planning variables in each TAZ.

The land use data (shapefile – polygons) was overlaid on the TAZ layer using "intersect" feature in GIS environment to estimate the area of each land use characteristic for each TAZ. The street centerline network (shapefile – lines) was overlaid on the TAZ layer using "intersect" feature in GIS environment to summarize the total length (center-lane miles) for each TAZ. Table 4.1 briefly describes the land use characteristics considered in this research.

TABLE 4.1: Description of land use variables

Land Use Category	Description
Mixed use development (MUDEV)	Areas with residential and compatible non-residential uses less than 10 acres to serve the residents of the planned community.
Mixed use district (MUDIS)	Areas with residential and compatible non-residential uses greater than 10 acres to serve the residents of the planned community.
Urban residential (UR)	Single-family to higher-density residential development nearer to the employment core.
Industrial (IND)	Areas with manufacturing, processing, and assembling of parts, distribution centers, and transportation terminals; specialized industrial.
Business (BUS)	Areas with retailing of merchandise to serve a large trade area, warehousing, wholesaling, etc.
Urban residential commercial (URC)	Areas with residential, retail, office, recreational, and cultural uses.
Multi-family	Areas with a variety of housing types; 12 to 43 dwelling units per acre.
Office district (OD)	Areas conducive to establishment and operation of offices, institutions, and commercial activities not involved in sales.
Single-family (SF)	Areas with primarily single-family housing; 3 to 8 dwelling units per acre.

Institutional (INS)	Major cultural, educational, medical, governmental, religious, athletic, and other institutions.
Neighborhood service district (NSD)	Mixed-use areas focusing on neighborhood retail and service activities.
Right-of-way (ROW)	Right of way areas of interstates, major and minor thoroughfares and other roads.
Commercial center (CC)	Shopping centers and individual retail establishments larger than 70,000 ft^2 of floor area.
Innovative (INN)	Non-traditional and new type of land use.
Planned unit development (PUD)	Area with a variety of type, design, and arrangement of structures.
Rural district (RURD)	Areas rural in nature.
Research district (RESD)	Areas with high research, development and technology manufacturing operations and professional employment.
Manufactured house (MH)	Areas with homes manufactured in a factory, transported as a whole unit and used at the site to build a house.
Residential mobile (RM)	Areas with manufactured home and mobile home parks.

4.1.2 Selection of TAZ's to Develop Models

The data obtained for all the TAZ's in a given study area has to be consistent and should not contain any unknown variables or the variables with unknown values. Since the models are being developed at a TAZ level, incorporating the TAZs with these characteristics in the model may have a biased effect on the results obtained. Therefore, TAZs with unknown variables or variables with unknown values were excluded from the analysis and development of models.

For example, considering the land use data for the City of Charlotte, it was observed that a few TAZs have land use characteristics with unknown type or open area. Incorporating the TAZs with these characteristics in the model may have a biased effect on the results obtained. Therefore, TAZs with open land area or unknown type of land use characteristics were excluded from the analysis and development of models.

The final database now obtained was used to develop macroscopic models in predicting crashes. The database with the TAZ's that were excluded from the analysis and development of models are used to validate and test the models developed.

4.1.3 Models

The proposed methodology incorporates two different approaches in developing the models, modeling using statistical techniques and modeling using neural networks.

4.1.3.1 Statistical Models

The traditional approach in developing a statistical model to predict a dependent variable involves 1) examination of spatial autocorrelation between the crashes, 2) examination of the correlation between the independent variables and 3) selection of distribution function.

4.1.3.1.1 Spatial Autocorrelation

As crashes in each TAZ's are location specific, spatial autocorrelation has to be investigated to evaluate whether there is any influence of crashes in a TAZ on its adjacent TAZ's. Moran's I was calculated in GIS environment to measure spatial autocorrelation, wherein inverse distance method is used indicating the decrease in the influence with an increase in the distance from crash location. The distance is calculated as Euclidean distance in analysis process. Along with Moron's I value, Z-Score and P-values can also be calculated to evaluate the statistical significance.

In general, Moran's I closer to +1 indicates highly positive spatial autocorrelation, closer to -1 indicates highly negative spatial autocorrelation, and closer to 0 indicates zero spatial autocorrelation. To develop crash prediction models at zonal level, the crash data obtained should have zero spatial autocorrelation indicating that a crash or crashes in a TAZ do not influence crashes in adjacent TAZs (or TAZs in close proximity).

4.1.3.1.2 Examination of Correlation between Independent Variables

As stated earlier, the objective of this research includes developing crash estimation models and travel demand models as a function of land use characteristics at macroscopic and microscopic levels.

One needs to test the correlation between demographic / socio-economic characteristics (population, number of household units and employment), traffic indicators (trip productions and attractions), and on-network characteristics (center-lane miles by speed limit) and land use characteristics to minimize any possible bias that might arise due to eliminating these variables in the macroscopic models. Statistical tests were conducted by computing Pearson correlation coefficient to examine the correlation between population, number of household units, employment, trip productions, trip attractions, and center-lane miles in each TAZ and land use characteristics.

Similarly, Pearson correlation coefficient was computed to examine the correlation between all the independent variables considered in the microscopic models.

In this research, two variables were considered to be strongly correlated to each other if the computed Pearson correlation coefficient is less than -0.2 or greater than +0.2 (significance value less than 0.01 for the considered data). It is expected that all other predictor variables may have a strong correlation with at least one of the land use characteristics (see Section 5.3) as the land use characteristics can explain all other predictors.

Therefore, omitting these variables that are correlated to land use characteristics will not only minimize any possible multicollinearity but help examine the direct effect of land use characteristics on crashes and travel demand. The rational for minimizing the number of variables and restricting to land use characteristics in the model is that the resultant model is more likely to be numerically stable and can be more easily generalized.

4.1.3.1.3 Selection of Distribution Function to Develop Statistical Models

A mathematical relationship between the dependent variable and independent variables is the basis of statistical modeling process. The nature of data plays a vital role in developing a model. As discussed in the "Literature Review" section, linear as well as non-linear distributions (Poisson with log-link, Negative Binomial with log-link, and log-normal) were used to develop count based models in the past. Poisson distribution is generally used to model count data if the mean is equal to its variance. However, if data are over-dispersed, the variance will be greater than mean. In such a case, a Negative Binomial distribution is used in the modeling process. These Negative Binomial distribution models can take into account the effect of unobserved heterogeneity due to omitted variables or variables that were not considered among TAZs.

Equation (4.1) shows the relation between the variance and mean in case of a Negative Binomial distribution.

$$\text{Variance} = \sigma^2 = \mu + \alpha\mu^2 \qquad \text{... Equation (4.1)}$$

where, σ is the standard deviation of crashes, μ is the estimated mean number of crashes and α is the Negative Binomial dispersion parameter.

The variance has to be computed based on the dependent variable for the study TAZs and road links and are compared with their respective means. If data is observed to be over-dispersed (α greater than zero) and not spatially correlated, a Negative Binomial distribution will be more suitable to estimate dependent variable. Let ‘d_i’ represents the vector consisting of the independent variables data of TAZ ‘i’ and ‘c_i’ represents the number of crashes in TAZ ‘i’. The general form of Negative Binomial model is given by Equation (4.2) (Miaou (1994)).

$$p(C_i = c_i) = \frac{\Gamma\left(c_i + \frac{1}{\alpha}\right)}{\Gamma(c_i + 1)\Gamma\left(\frac{1}{\alpha}\right)} \left(\frac{1}{1 + \alpha\mu_i}\right)^{\frac{1}{\alpha}} \left(\frac{\alpha\mu_i}{1 + \alpha\mu_i}\right)^{c_i}$$

$$c_i = 0, 1, 2, 3 \ldots \qquad \text{…Equation (4.2)}$$

where,

C_i is the independent variable following Negative Binomial distribution,

$$\mu_i = expectation\ of\ C_i = E(C_i) = g(d_i) = \ e^{\left(\ \beta_o + \sum_{j=1}^{n} X_{ij}\beta_j\right)}$$

$$j = 1, 2, 3, \ldots .n \qquad \text{... Equation (4.3)}$$

n is the number of observations and $g(d_i)$is the functional form of Negative Binomial distribution.

Independent variables with Wald Chi-Square value less than 1 or the level of significance greater than 0.05 (95 percent confidence level) were considered to have a statistically insignificant effect on dependent variables. These parameters were examined for each independent variable to eliminate those that have a statistically insignificant effect on dependent variable.

The quasi-likelihood under independence model criterion, QIC, (to select the model with best working correlation structure) and the corrected version of quasi-likelihood under independence model criterion, QICC, (to choose the best subset of predictors) were used to assess the goodness of fit. The lower the QIC and QICC, the better is the goodness of fit. Also, the difference between QIC and QICC should be generally low for a good model.

Depending on the data and its distribution function, macroscopic crash estimation were developed with land use characteristics as independent variables. The models developed were validated using the test database that was not used in the development of models.

4.1.4 Back-propagation Neural Network Model

A multi-layer feed-forward back-propagation network using back-propagation (BP) learning algorithm, which can also be referred as Back-propagation Neural Network (BPNN) model, was developed. Typically, a BPNN model consists of three layers: input layer, hidden layer and output layer. The databases used to develop statistical models were used to develop BPNN model to maintain consistency for performance evaluation.

The independent variables used in the statistical models have to be given as the input vector to the network. So, the input layer has the number of neurons each of which corresponds to an independent variable in the model. The output layer has the number of neurons equal to number of outputs. For example, the output layer of macroscopic crash prediction model has three neurons each corresponding to the total number of crashes, the total number of injury crashes and the total number of PDO crashes as the dependent variables. Tangent sigmoid function and purelin function are used as transfer functions for hidden layer and output layer with Bayesian-Regulation Back-propagation function as training function.

The Bayesian regularization minimizes a linear combination of squared errors and weights. It also modifies the linear combination so that at the end of training the resulting network has good generalization qualities. One can avoid costly cross validation by using Bayesian regularization. It is particularly useful in this scenario where a part of the available data is reserved for validation. Regularization also reduces/eliminates the need for testing different number of hidden neurons for a problem. This Bayesian regularization takes place within the Levenberg-Marquardt algorithm. BP is used to calculate the Jacobian ‘J’. Each variable is adjusted according to Levenberg-Marquardt as follows (MATLAB, 2012):

$$\text{Error } (E) = \frac{1}{n}\sum_{i=1}^{n}(\hat{c}_i - c_i)^2 \qquad \text{... Equation (4.4)}$$

$$\text{Error gradient } (g) = J^t E \qquad \text{... Equation (4.5)}$$

$$\text{Hessian Matrix } (H) = J^t J \qquad \text{... Equation (4.6)}$$

$$\text{Cost Function } (C) = \beta * E_d + \alpha * E_w \qquad \text{... Equation (4.7)}$$

where, E_d and E_w are sum-squared errors and sum-squared weights respectively.

$$(H + \lambda \mathrm{I})\delta = g \qquad \text{... Equation (4.8)}$$

where, $\hat{c}_i$ and c_i are predicted and observed crashes, ‘I’ is an identity matrix and λ is the damping factor and is adjusted based on sum-squared errors.

The Equation (4.8) is solved to calculate ‘δ’ and the weights are updated based on the value of ‘δ’. MATLAB was used to build the BPNN and train the network. The BPNN was trained such that the error term shown in Equation (4.4) and the cost function shown in Equation (4.7) are minimized. The readers can refer to Liang (2003), Liang

(2005), MATLAB (2012) and Xie et al. (2007) for more detailed description of Bayesian regularization, Levenberg-Marquardt algorithm, and BPNN. The BPNN was trained using the same database that was used to develop statistical models and validated using the same test database.

The performance of the network was evaluated by calculating mean square errors. The number of neurons in the hidden layer is not limited to a fixed number. So, appropriate number of neurons was evaluated by changing the number of neurons in the hidden layer until the network performs well after training.

4.2 Microscopic Models

The microscopic models in this research are the models which help predict the dependent variable (link volume or crashes) at a link level. In order to develop these models the database related to dependent and independent variables has to represent data at a link level.

4.2.1 Database to Develop Models

The database to develop microscopic models should contain crash data, land use data and street centerline network, and traffic counts obtained from the permanent traffic count stations in a GIS format.

The permanent traffic counts may not be available for all the links in the network. So, the permanent traffic count data layer was overlaid on street centerline network layer to extracts the links with their respective traffic counts. The land use layer was then overlaid on the links which have permanent traffic counts. Spatial proximity tools in GIS environment were used to calculate the distance from all land use polygons to the count stations. For example, if the study area has permanent traffic counts for ‘N’ links and the study area is divided into ‘M’ number of land use polygons, a distance matrix of size ‘N x M’ is obtained. Each cell in the matrix indicates the distance between corresponding road link and the land use polygon. As the influence of land use characteristic on the road link is inversely proportional to the square of distance (Stewart (1948)), the areas of land use characteristics are divided by the square of distance and the respective land use characteristics summed together to evaluate the influence of each type of land use characteristic in the entire study area on the given road link. This is mathematically represented using Equation (4.2).

$$A_{ic} = \sum_{j=1}^{n} \frac{a_{ji}}{{d_{jc}}^{2}} \qquad \text{... Equation (4.2)}$$

where, A_{ic} is the total area of influence of land use characteristic of type ‘i’ on the road link ‘c’, n is the total number of land use polygons of type ‘i’ in study area, a_{ji} is the area of each individual land use polygon ‘j’ of type ‘i’ and d_{jc} is the distance between road link ‘c’ and the land use polygon ‘j’ of type ‘i’. Figure 4.1 shows the influence of each land use polygon on a given point on the road link represented by the darkness of the color.

The crash data was then overlaid on the street network layer with permanent counts. Spatial analysis tools in GIS environment were used to evaluate the total number of crashes based on crash type on each road link which have permanent counts.

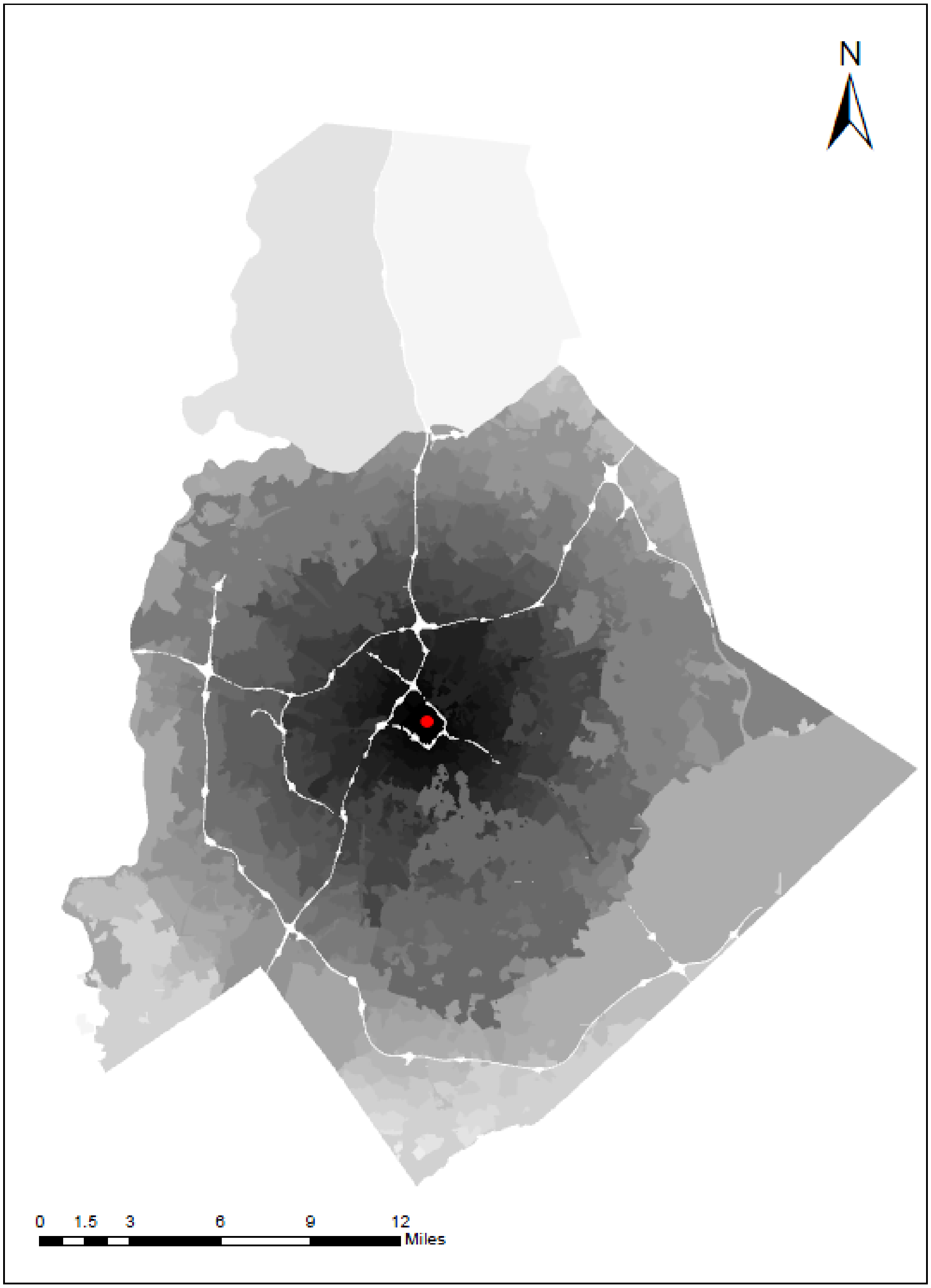

FIGURE 4.1: Spatial representation of the influence of land use polygons on a given road link

4.2.2 Statistical and Neural Network Model

Since the nature of the data is same as that of macroscopic models, the following statistical and BPNN models were developed by implementing the same procedure to develop macroscopic models.

- Statistical model to predict crashes at link level with areas of influence of each type of land use characteristics on a link as independent variables and the number of crashes on respective links as dependent variable.
- BPNN model to predict crashes at link level with areas of influence of each type of land use characteristics on a link as input vector and the number of crashes on respective links as output vector.
- Statistical model to predict link level travel demand with areas of influence of each type of land use characteristics on a link as independent variables and traffic counts on respective links as dependent variable.
- BPNN model to predict link level travel demand with areas of influence of each type of land use characteristics on a link as input vector and traffic counts on respective links as output vector.

CHAPTER 5: MACROSCOPIC MODELS

The City of Charlotte, Mecklenburg County, North Carolina was considered as the study area. Crash data, land use data, street centerline network and TAZ layer (with planning variables that are used to estimate trip productions/attractions) for the year 2005 was obtained in a GIS format from the City of Charlotte Department of Transportation (CDOT).

5.1 Spatial analysis and autocorrelation

As crashes in each TAZ are location specific, spatial autocorrelation was investigated to evaluate whether there is any influence of crashes in a TAZ on its adjacent TAZ's. Moran's I was calculated in GIS environment to measure spatial autocorrelation, where inverse distance method was used indicating the decrease in the influence with increase in the distance from crash location. The distance was calculated as Euclidean distance in analysis process. The Moran's I value was observed to be 0.07 (very low) with a Z-Score of 48.38 and a P-Value less than 0.01. In general, Moran's I closer to +1 indicates highly positive spatial autocorrelation, closer to -1 indicates highly negative spatial autocorrelation, and closer to 0 indicates zero spatial autocorrelation. Data considered in this research shows that crashes in a TAZ does not influence crashes in adjacent TAZs (or TAZs in close proximity).

5.2 Selection of TAZs

Data obtained showed that there were 1,057 TAZs in the study area. A total of 10,726 reported crashes occurred in these TAZs during the year 2005. It was observed that there were 24 fatal crashes, 3,522 injury crashes and 7,180 PDO crashes in the study area during the year 2005. As an example, Figure 5.1 shows the number of crashes in each TAZ during the year 2005 in the study area.

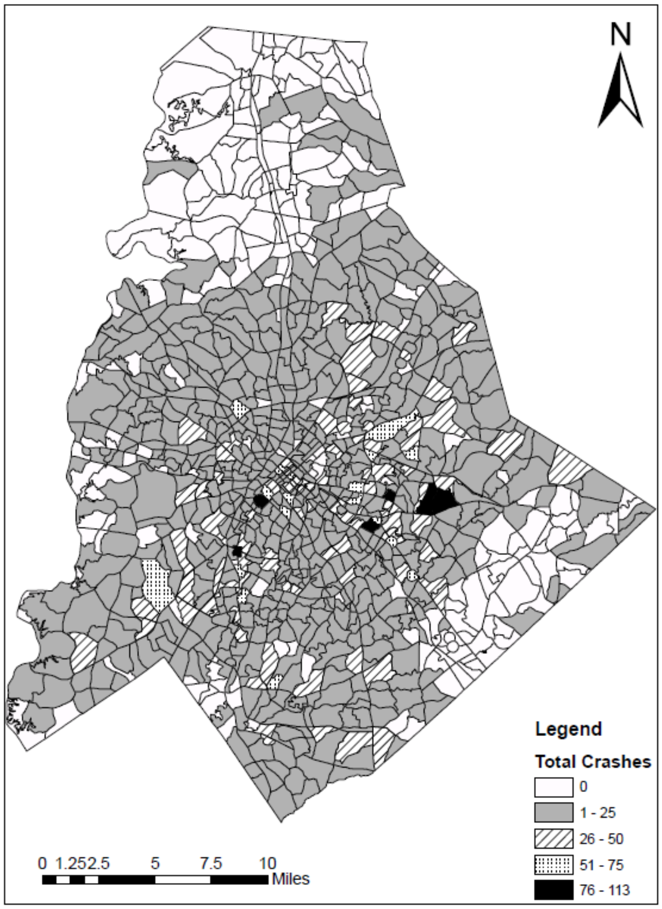

FIGURE 5.1: Spatial extent of total crashes per TAZ in Charlotte, North Carolina

Data for 765 TAZs with a total of 9,799 crashes (includes 20 fatal crashes, 3,227 injury crashes and 6,552 PDO crashes) were selected for the analysis and development of models, while data for 268 randomly selected TAZs (~35% when compared to 765

TAZs) was used to validate the models developed. A summary of land use characteristics considered in this research is shown in Table 5.1. The table also summarizes dependent and independent variables for the selected TAZs. The computed minimum, maximum, mean and standard deviation values using data for TAZs are also shown in the table.

TABLE 5.1: Summary of land use characteristics

Variable	Description	N	Minimum	Maximum	Mean	Standard deviation
TAZ	Number of TAZs	765				
Dependent variables						
TC	Total number of crashes	765	0.000	113.000	12.809	14.726
IC	Number of injury crashes	765	0.000	37.000	5.100	4.218
PDO	Property damage only crashes	765	0.000	76.000	10.127	8.565
Independent variables (in square miles)						
MUDEV	Mixed use development	765	0.000	0.221	0.002	0.011
MUDIS	Mixed use district	765	0.000	1.386	0.019	0.100
UR	Urban residential	765	0.000	0.155	0.001	0.009
IND	Industrial	765	0.000	3.658	0.057	0.181
BUS	Business	765	0.000	0.660	0.018	0.042
URC	Urban residential commercial	765	0.000	0.023	0.000	0.001
MF	Multi-family	765	0.000	0.328	0.029	0.050
OD	Office district	765	0.000	0.178	0.008	0.022
SF	Single family	765	0.000	1.919	0.256	0.297
INS	Institutional	765	0.000	0.882	0.006	0.041
NSD	Neighborhood service district	765	0.000	0.048	0.001	0.004
ROW	Right-of-way	765	0.000	0.086	0.001	0.006
CC	Commercial center	765	0.000	0.315	0.005	0.027
INNOV	Innovative	765	0.000	0.118	0.002	0.011
PUD	Planned unit development	765	0.000	0.472	0.008	0.046
RURD	Rural district	765	0.000	0.005	0.000	0.000
RESD	Research district	765	0.000	0.904	0.004	0.048
MH	Manufactured house	765	0.000	0.004	0.000	0.000

5.3 Examination of correlation between independent variables

As stated earlier, the primary focus of this step is to develop crash estimation models as a function of land use characteristics at a TAZ level. One needs to test the correlation between demographic / socio-economic characteristics (population, number of households and employment), trip productions and attractions, and network characteristics (centerline miles by speed limit) and land use characteristics to minimize any possible bias that might arise due to eliminating these variables. Statistical tests were therefore conducted by computing Pearson correlation coefficient to examine the correlation between population, number of households, employment, trip productions, trip attractions, and street centerlane miles in each TAZ and land use characteristics. In this research, two variables were considered to be strongly correlated to each other if the computed Pearson correlation coefficient is less than -0.2 or greater than +0.2 (significance value less than 0.01 for considered data).

Table 5.2 summarizes computed Pearson correlation coefficients between all the selected independent variables. Population of a TAZ was observed to be significantly correlated to single-family (SF) residential, multi-family (MF) residential, planned unit development (PUD) and innovative (INN) land use areas. The total number of household units in a TAZ was observed to be significantly correlated to single-family (SF) residential, multi-family (MF) residential and innovative (INN) land use areas. Employment or job count was observed to be significantly correlated to industrial (IND), office district (OD), mixed use development (MUDEV) and research district (RESD) land use areas.

Trips productions from and trip attraction to a TAZ were observed to be significantly correlated to industrial (IND), multi-family (MF) residential, office district (OD) and institutional (INS) land use areas.

The total number of centerlane miles was observed to be significantly correlated to mixed use district (MUDIS), single-family (SF) residential, multi-family (MF) residential and planned unit development (PUD) land use areas. The total number of centerlane miles with minor roads (local and collector roads with speed limit equal to 25 mph or 35 mph) was observed to be significantly correlated to mixed use district (MUDIS), single-family (SF) residential, multi-family (MF) residential and planned unit development (PUD) land use areas, while major roads (arterials and expressways with speed limit greater than or equal to 40 mph but less than 60 mph) was observed to be significantly correlated to business (BUS) land use. In general, total centerlane miles of major roads were found to be highly correlated to total centerlane miles of all roads.

The urban residential commercial (URC), rural district (RURD) and mixed use district (MUDIS) land use variables were observed to be significantly correlated to urban residential (UR), planned unit development (PUD) and single-family (SF) land use variables, respectively.

Therefore, omitting demographic, socio-economic and network characteristics as well as urban residential commercial (URC), rural district (RURD) and mixed use district (MUDIS) land use variables will not only minimize any possible multicollinearity but will help examine the direct effect of selected land use characteristics on crashes.

TABLE 5.2: Pearson Correlation Coefficients – Summary

Variable	Population	# households units	Employment	Trip productions	Trip attractions	Center-lane Miles			MUDEV	MUDIS	UR	IND	BUS	URC	MF	OD	SF	INS	NSD	ROW	CC	INN	PUD	RURD	RESD	MH
						All roads	Minor roads	Majors roads																		
Population	1.00																									
# household units	**0.95**	1.00																								
Employment	-0.06	-0.06	1.00																							
Trip productions	**0.32**	**0.31**	**0.81**	1.00																						
Trip attractions	**0.32**	**0.31**	**0.81**	**1.00**	1.00																					
All roads	**0.60**	**0.57**	0.05	**0.22**	**0.22**	1.00																				
Minor roads	**0.64**	**0.61**	0.03	**0.21**	**0.21**	**0.98**	1.00																			
Major roads	0.04	0.00	0.12	0.09	0.09	**0.45**	**0.26**	1.00																		
MUDEV	-0.09	-0.08	**0.24**	0.15	0.15	0.03	0.01	0.09	1.00																	
MUDIS	-0.01	-0.01	-0.04	-0.04	-0.04	**0.37**	**0.38**	0.09	-0.01	1.00																
UR	0.03	0.06	0.05	0.06	0.06	0.01	0.01	0.00	0.08	0.00	1.00															
IND	-0.13	-0.14	**0.43**	**0.21**	**0.21**	0.09	0.06	0.17	0.06	-0.01	-0.03	1.00														
BUS	0.03	0.04	0.15	0.17	0.17	0.07	0.03	**0.21**	0.01	-0.03	-0.04	0.01	1.00													
URC	0.02	0.08	0.00	0.02	0.02	-0.01	-0.01	0.01	0.08	-0.01	**0.34**	-0.02	-0.02	1.00												
MF	**0.54**	**0.60**	0.00	**0.21**	**0.21**	**0.30**	0.30	0.10	-0.06	-0.07	-0.03	-0.01	0.11	-0.04	1.00											
OD	0.05	0.07	**0.31**	0.33	**0.33**	0.10	0.07	0.19	0.13	-0.02	-0.02	0.01	0.14	-0.01	0.07	1.00										
SF	**0.37**	**0.33**	-0.18	-0.06	-0.06	**0.66**	0.68	0.16	-0.11	**0.25**	-0.06	-0.10	-0.08	-0.06	0.04	-0.02	1.00									
INS	0.18	0.03	0.12	**0.21**	**0.20**	0.09	0.07	0.11	-0.02	-0.02	-0.02	-0.01	0.04	0.01	0.05	0.02	-0.03	1.00								
NSD	-0.01	-0.01	0.07	0.03	0.03	0.09	0.09	0.06	-0.01	0.04	0.01	0.03	0.02	-0.01	0.04	0.12	0.13	0.08	1.00							
ROW	0.00	0.00	-0.03	-0.02	-0.02	0.11	0.07	0.18	-0.01	0.01	-0.01	-0.03	0.00	-0.01	-0.04	0.00	0.12	-0.01	-0.01	1.00						
CC	0.00	0.00	0.04	0.09	0.09	0.19	0.17	0.15	-0.01	0.16	-0.02	-0.04	0.06	0.00	0.01	0.03	0.04	0.04	-0.01	0.03	1.00					
INN	0.19	**0.22**	-0.04	0.05	0.05	0.11	0.12	-0.01	-0.02	-0.03	0.00	-0.06	-0.03	-0.01	0.07	0.00	0.12	-0.02	-0.03	-0.02	0.04	1.00				
PUD	**0.22**	**0.20**	-0.04	0.05	0.05	0.24	**0.26**	0.03	-0.03	0.05	-0.01	-0.04	0.00	-0.01	-0.01	0.00	0.11	-0.02	-0.02	0.11	0.08	0.09	1.00			
RURD	0.15	0.12	-0.01	0.03	0.03	0.12	0.13	0.00	-0.01	-0.01	0.00	0.03	-0.01	0.00	0.00	-0.01	0.05	-0.01	-0.01	0.00	-0.01	-0.01	**0.36**	1.00		
RESD	-0.02	-0.01	0.30	0.15	0.15	0.05	0.04	0.06	0.13	-0.01	-0.01	-0.03	-0.01	-0.01	0.02	0.03	-0.02	0.08	-0.01	-0.01	-0.02	-0.02	-0.02	0.00	1.00	
MH	-0.03	-0.03	-0.01	-0.02	-0.02	0.00	-0.03	0.11	-0.01	-0.01	0.00	0.06	-0.02	0.00	-0.02	-0.01	0.04	-0.01	-0.01	0.00	-0.01	-0.01	-0.01	0.00	0.00	1.00

5.4 Negative Binomial Model

Crash estimation models were developed separately considering the total number of crashes, the total number of injury crashes and the total number of PDO crashes as the dependent variable using SPSSv16 (SPSS (2008)). It was observed that the 765 TAZs data considered for modeling have only 20 fatal crashes (out of 9,799 crashes). Also, only 2.61% of all the TAZ's considered have fatal crashes. This percent is too low to even test zero-inflated models. So, a crash estimation model for fatal crashes was not developed due to fewer numbers of fatal crashes and statistically insignificant sample size observed in the study TAZs.

All land use characteristics shown in Table 5.1, excluding urban residential commercial (URC), rural district (RURD) and mixed use district (MUDIS), were considered as the independent variables. Though data was observed to be over-dispersed, linear, Poisson with log-link and log-normal distributions were also tested in addition to Negative Binomial with log-link distribution. QIC and QICC obtained for Negative Binomial with log-link distribution based models were observed to be the lowest implying that these models would result in lower specification errors than when compared to other models tested in this research. Therefore, results obtained considering Negative Binomial with log-link are only discussed in this report.

The statistical parameters along with the coefficient, Wald Chi-square and significance value for each variable in the preliminary model to estimate the total number of crashes per year in a TAZ are shown in Table 5.3. It was observed that a few land use characteristics have a significance value greater than 0.05 (at a 95 percent confidence level). In other words, these land use characteristics do not have a statistically significant effect on the total number of crashes in a TAZ.

The land use characteristics with a significance value greater than 0.05 were omitted and the model was re-run. The statistical parameters along with the coefficient, Wald Chi-square and significance value for each variable in the final model with only land use characteristics that have statistically significant effect on the total number of crashes are shown in Table 5.4. Similarly, crash estimation models were developed to predict injury and PDO crashes in a TAZ. Results obtained are also summarized and shown in Table 5.4.

The coefficient for all the land use characteristics other than single-family (SF) residential is positive for the models to estimate total number of crashes and PDO crashes. This implies that an increase in the area of these land use characteristics will lead to an increase in the total number of crashes. An increase in single-family (SF) residential land use area will result in a decrease in the total number of crashes and PDO crashes.

The coefficient for all the land use characteristics other than single-family (SF) residential and industrial (IND) is positive for the model to estimate the number of injury crashes in a TAZ. This implies that an increase in the area of these land use characteristics will lead to an increase in the number of injury crashes. An increase in single-family (SF) residential and industrial (IND) land use areas will result in a decrease in the number of injury crashes.

The institutional (INS), industrial (IND) and research district (RESD) land use characteristics have a significant effect in estimating the total number of crashes and/or injury crashes but not PDO crashes. This shows that, the presence of these characteristics in a TAZ might contribute to an increase in the severity of a crash (more injury crashes).

TABLE 5.3: Preliminary model parameters to estimate the total number of crashes in a TAZ.

Variable	Coefficient	Wald Chi-square	Significance value
(Intercept)	2.223	914.34	<0.001
Mixed use development (MUDEV)	8.213	9.41	0.002
Urban residential (UR)	3.813	3.82	0.051
Industrial (IND)	0.005	0.01	0.936
Business (BUS)	7.670	1,325.03	<0.001
Multi-family (MF)	4.038	99.92	<0.001
Office district (OD)	7.123	102.04	<0.001
Single family (SF)	-0.426	23.13	<0.001
Institutional (INS)	0.256	4.10	0.043
Neighborhood service development (NSD)	0.114	0.00	0.972
Right-of-way (ROW)	6.537	657.24	<0.001
Commercial center (CC)	-0.665	1.09	0.296
Innovative (INN)	0.457	1.08	0.298
Planned unit development (PUD)	0.403	48.89	<0.001
Research district (RESD)	0.334	4.05	0.044

TABLE 5.4: Summary of parameter estimates of the models

Variable	Coefficient	Std. error	Wald Chi-square	Significance value
Model to estimate the total number of crashes in a TAZ				
(Intercept)	2.227	0.071	981.128	<0.001
Mixed use development (MUDEV)	9.035	2.778	10.573	0.001
Urban residential (UR)	8.484	2.905	8.531	<0.001
Business (BUS)	7.590	0.242	987.116	<0.001
Multi-family (MF)	3.979	0.441	81.371	<0.001
Office district (OD)	7.016	0.701	100.280	<0.001
Single-family (SF)	-0.415	0.077	29.208	<0.001
Institutional (INS)	0.262	0.124	4.487	0.03
Right-of-way (ROW)	6.404	0.293	478.605	<0.001
Planned unit development (PUD)	0.401	0.082	23.889	<0.001
Research district (RESD)	0.311	0.175	3.164	0.07
Dispersion Parameter α = 1.24				
Quasi Likelihood under Independence Model Criterion (QIC) = 874.97				
Corrected Quasi Likelihood under Independence Model Criterion (QICC) = 891.19				
Model to estimate the number of injury crashes in a TAZ				
(Intercept)	1.124	0.067	279.521	<0.001
Mixed use development (MUDEV)	7.700	2.703	8.114	<0.001
Urban residential (UR)	9.525	2.717	12.294	<0.001
Industrial (IND)	-0.149	0.068	4.826	0.028
Business (BUS)	7.456	0.264	796.941	<0.001
Multi-family (MF)	4.378	0.362	146.420	<0.001
Office district (OD)	6.068	0.636	91.107	<0.001
Single-family (SF)	-0.430	0.075	32.690	<0.001
Institutional (INS)	0.477	0.104	21.040	<0.001
Right-of-way (ROW)	7.168	0.311	530.793	<0.001
Research district (RESD)	0.568	0.110	26.442	<0.001
Dispersion Parameter α = 0.48				
Quasi Likelihood under Independence Model Criterion (QIC) = 874.39				
Corrected Quasi Likelihood under Independence Model Criterion (QICC) = 891.69				
Model to estimate the number of PDO crashes in a TAZ				
(Intercept)	1.844	0.075	606.170	<0.001
Mixed use development (MUDEV)	9.588	2.800	11.725	0.001
Urban residential (UR)	7.376	3.180	5.379	0.02
Business (BUS)	7.276	0.232	987.177	<0.001
Multi-family (MF)	3.651	0.472	59.866	<0.001
Office district (OD)	7.309	0.780	87.848	<0.001
Single-family (SF)	-0.426	0.079	28.931	<0.001
Right-of-way (ROW)	6.098	0.359	289.195	<0.001
Innovative (INN)	0.966	0.584	2.736	0.09
Planned unit development (PUD)	0.586	0.068	74.950	<0.001
Dispersion Parameter α = 0.61				
Quasi Likelihood under Independence Model Criterion (QIC) = 875.03				
Corrected Quasi Likelihood under Independence Model Criterion (QICC) = 888.91				

The Negative Binomial dispersion parameter value (α) was lower for the model to estimate the number of injury crashes than when compared to the models to estimate the total number of crashes and the number of PDO crashes. This indicates that injury crashes are less dispersed when compared to the total number of crashes and PDO crashes. QIC and QICC were reasonably close for all the three final models.

5.5 BPNN Model

The BPNN models are built and trained using MATLAB Neural Network Toolbox (MATLAB, 2012). Default settings were used for all the parameters except for the number of epoch (set to 1000 iterations) and learning rate (set to 0.05). To evaluate the best number of neurons in the hidden layer and also the performance ratio, BPNN models with hidden neurons = 10, 11, 12…..34, 35, 36 with performance ratio's = 0.05, 0.15, 0.25….0.95 were tried. Tangent sigmoid function and purelin function are used as a transfer functions for both hidden layer and output layer respectively with 'trainlm' as training function which updates weight and bias values according to Levenberg-Marquardt optimization.. The network was trained using the database that was used to develop statistical models and validated using the test database. The performance of the network was evaluated by calculating errors. It is observed that, the BPNN with performance ratio of 0.95 and hidden layer with 10, 13, 17, 20 and 30 neurons performed well after training.

5.6 Comparison of Predictive Performance

To compare the predictive performance of the models developed, the data for the 268 TAZs that were randomly selected for validation are used. Four different criteria were used to evaluate the predictive performance of the models developed. They are:

- Average Error or Mean Absolute Deviation (MAD)

$$MAD = \frac{1}{n}\sum_{i=1}^{n}|\hat{c}_i - c_i|$$

- 50th Percentile Error = An error value below which 50% of the observations fall
- 85th Percentile Error = An error value below which 85% of the observations fall
- Root Mean Squared Error (RMSE)

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\hat{c}_i - c_i)^2}$$

where, ‘n’ is the sample size, $\hat{c}_i$ and c_i are predicted and observed number of crashes respectively.

For all the four evaluation criteria (MAD, 50th percentile error, 85th percentile error and RMSE), the values closer to zero indicate better model performance in prediction.

The predictive performances of Negative Binomial models and BPNN models in predicting total, injury and PDO crashes are summarized in Table 5.5. For all the three different types of crash prediction models (total, injury and PDO), the BPNN models outperformed Negative Binomial models. It was observed that, no single BPNN model outperformed other BPNN models when all the four performance evaluation criteria were considered. So, to evaluate a single best model, the average of the performance criteria are calculated and the BPNN model with lower average value was considered as the best model in predicting crashes at TAZ level.

From Table 5.5, the performance criteria of the models to predict total crashes i.e. MAD, 50th percentile error, 85th percentile error and RMSE are observed to be lower for

BPNN30 (30 neurons in the hidden layer) model (3.30, 1.28, 5.89 and 6.46 respectively) when compared to the Negative Binomial model (9.52, 9.27, 9.27 and 16.16). The performance criteria of the models to predict injury crashes are observed to be lower for BPNN13 (13 neurons in the hidden layer) model (1.02, 0.00, 2.13 and 2.28) when compared the Negative Binomial model (3.28, 3.08, 3.10 and 5.07). Similarly, the performance criteria of the models to predict PDO crashes are observed to be lower for BPNN20 (20 neurons in the hidden layer) model (2.31, 0.65, 4.35 and 4.75) when compared the Negative Binomial model (6.36, 6.32, 6.32 and 9.99).

The above observations clearly indicate that the BPNN30 model, BPNN 13 model and BPNN20 model have better performance in predicting total, injury and PDO crashes, respectively in a TAZ when compared to Negative Binomial and other BPNN models.

Unlike Negative Binomial models, in the case of BPNN models, three different models are not required to predict total, injury and PDO crashes. The BPNN models are designed such that a single model would predict all the three types of crashes. Overall, to evaluate a best model in predicting total, injury and PDO crashes, from the above observations, the average of the performance criteria was considered and it is observed that the BPNN20 model performed well when compared to any other model developed. This means that the BPNN20 model has the best performance in predicting crashes in a TAZ.

TABLE 5.5: Performance criteria comparison for Negative Binomial and BPNN models (Macroscopic)

Total Crashes					
Model	MAD	50th Percentile Error	85th Percentile Error	RMSE	Average
Negative Binomial	9.52	9.27	9.27	16.16	11.06
BPNN10	4.32	3.17	5.28	6.76	4.88
BPNN13	4.14	2.39	7.10	6.41	5.01
BPNN17	3.75	2.20	6.00	6.37	4.58
BPNN20	3.55	0.82	6.04	7.36	4.44
BPNN30	3.30	1.28	5.89	6.46	4.23
Injury Crashes					
Model	MAD	50th Percentile Error	85th Percentile Error	RMSE	Average
Negative Binomial	3.24	3.08	3.10	5.07	3.62
BPNN10	1.45	0.99	1.70	2.37	1.63
BPNN13	1.02	0.00	2.13	2.28	1.36
BPNN17	1.92	1.68	1.89	2.51	2.00
BPNN20	1.21	0.56	1.93	2.25	1.49
BPNN30	3.19	3.45	3.45	3.58	3.42
PDO Crashes					
Model	MAD	50th Percentile Error	85th Percentile Error	RMSE	Average
Negative Binomial	6.36	6.32	6.32	9.99	7.25
BPNN10	2.98	2.20	3.40	4.67	3.31
BPNN13	2.59	1.21	4.95	4.30	3.26
BPNN17	2.74	2.22	2.82	4.34	3.03
BPNN20	2.31	0.65	4.35	4.75	3.02
BPNN30	2.60	1.55	3.55	4.59	3.07

5.7 Summary: Macroscopic Models

Results obtained from Negative Binomial macroscopic models show that mixed use development area, urban residential area, single-family residential area, multi-family residential area, business area and office district area are strongly associated with crashes in a TAZ. These land uses are generally high activity generators.

Land use characteristics such as institutional area (with major cultural, educational, medical, governmental, religious, athletic and other institutions), industrial area (with manufacturing, processing, and assembling of parts, distribution centers, and transportation terminals) and research district area (with high research, development and technology manufacturing operations and professional employment) were observed to play a statistically significant role only in estimating the total number of crashes and the number of injury crashes in a TAZ. From the results obtained, it can be inferred that these land use characteristics in a TAZ are strongly associated with an increase in severe crashes.

The coefficient of all the land use characteristics excluding single-family residential area is positive in the final model for total number of crashes in a TAZ. An increase in the area of single-family residential land use tends to lower the total number of crashes in a TAZ. While the presence of single-family residential area and industrial area tends to lower the number of injury crashes in a TAZ, the presence of single-family residential area only tends to lower the number of PDO crashes in a TAZ. It can, therefore, be inferred that presence of single-family residential area may have a neutralizing effect, possibly due to different behaviors adopted by drivers (such as cautious driving) or lower travel speed in these areas.

The coefficient for urban residential land use area and mixed use development land use area was observed to be the highest than when compared to any other land use characteristic area in case of all the models developed. This indicates that urban residential and mixed use development land use areas are strongly associated with the number of crashes in a TAZ. These land use areas are generally followed by business and office activity generators as areas that are strongly associated with higher number of crashes. TAZs with these mixed land use areas, which produce as well as attract trips of different modes of transportation throughout the day, need additional emphasis at planning and project implementation level to enhance traffic safety.

Validation data (data for 268 TAZ’s) were used to validate and compare the performance of the models developed. Both Negative Binomial and BPNN models performed well in predicting the crashes (total, injury and PDO). When MAD, 50th percentile error, 85th percentile error and RMSE are considered for performance evaluation, BPNN30, BPNN13 and BPNN20 models performed better in predicting total crashes, injury crashes and PDO crashes, respectively. As the performance criteria for the BPNN20 model are observed to be lower than any other model, one can infer that BPNN20 model has outperformed all other models in predicting crashes (total, injury and PDO) in a TAZ.

CHAPTER 6: MICROSCOPIC MODELS

The City of Charlotte, Mecklenburg County, North Carolina was considered as the study area. Crash data, land use data, street centerline network and permanent traffic count data for the year 2005 was obtained in a GIS format from the City of Charlotte Department of Transportation (CDOT).

6.1 Data Description

Data obtained showed that the permanent counts are available for 345 road links in the network. Data was extracted and gathered for each of these links. Of the 345 road links, data for 315 road randomly selected links was used for development of models while data for the remaining 30 randomly selected road links were used for validation and performance evaluation.

Along with the land use characteristics, as the traffic volume and crash severity may vary based on network characteristics, characteristics such as speed limit, presence of median, the number of lanes and functional classification were also considered in modeling process. The road links were divided into four different functional classes with volumes ‘<5,000’ as ‘1’, ‘5,001-10,000’ as ‘2’, ’10,001-20,000’ as ‘3’ and ‘>20,000’ as ‘4’.

A total of 25 independent variables were considered in developing crash prediction models, whereas, 24 variables were considered in developing travel demand models. Table 6.1 summarizes the minimum, maximum and mean value of traffic volumes, crashes and independent variables by links considered to develop microscopic models.

TABLE 6.1: Summary of Independent Variables

Variable	Description	N	Minimum	Maximum	Mean	Standard Deviation
Permanent Counts	# Counts	315				
Dependent Variables						
AADT	Annual Average daily Traffic	315	2,030.00	56,100.00	19,432.48	10,904.00
TC	Total number of crashes	310	0.000	39.000	8.797	8.208
IC	Number of injury crashes	310	0.000	14.000	2.803	2.878
PDO	Property damage only crashes	310	0.000	29.000	5.987	5.906
Independent Variables						
NL	Number of Lane	315	2.00	6.00	3.34	1.10
SL	Speed Limit (mph)	315	25.00	45.00	36.78	4.86
MEDIAN	Median	315	0.00	1.00	0.40	0.49
FC	Functional Class	315	1.00	4.00	3.13	0.91
MUDEV	Mixed Use Development (Influence)	315	0.01	11.09	0.44	1.20
MUDIS	Mixed Use District (Influence)	315	0.13	18.34	0.82	1.81
UR	Urban Residential (Influence)	315	0.01	4.40	0.19	0.50
IND	Industrial (Influence)	315	0.20	3,294.90	22.65	221.49
BUS	Business (Influence)	315	0.00	0.24	0.01	0.03
URC	Urban Residential Commercial (Influence)	315	0.30	130.24	4.16	8.87
MF	Multi Family (Influence)	315	0.05	136.09	1.96	8.21
OD	Office District (Influence)	315	4.05	2,486.69	20.94	142.77
SF	Single Family (Influence)	315	0.06	4.15	0.41	0.57
INS	Institutional (Influence)	315	0.01	4.91	0.13	0.47
NSD	Neighborhood Service District (Influence)	315	0.00	0.25	0.02	0.03
ROW	Right of Way (Influence)	315	0.31	560.76	5.35	31.60
CC	Commercial Center (Influence)	315	0.05	30.05	0.46	1.85
INNOV	Innovative (Influence)	315	0.01	5.74	0.13	0.40
PUD	Planned Unit Development (Influence)	315	0.06	88.56	0.56	5.00
RURD	Rural District (Influence)	315	0.00	0.04	0.00	0.00
RESD	Research District (Influence)	315	0.01	6.16	0.14	0.49
MH	Manufactured House (Overlay) (Influence)	315	0.00	0.00	0.00	0.00
RM	Residential Mobile (Influence)	315	0.01	1.85	0.10	0.19
UNKNOWN	Unknown (Influence)	315	0.63	4.63	1.34	0.67

6.2 Microscopic Crash Prediction Models

Two different techniques were incorporated in developing the models to predict link level crashes on the urban streets. When the count data exhibit over dispersion, Negative Binomial regression model remains the most widely used statistical model. In conjunction with the Negative Binomial model, a back-propagation neural network (BPNN) model was also developed using the same dataset. Both the models were compared for performance evaluation. The following sections briefly describe the Negative Binomial and neural network models developed in this research.

6.2.1 Negative Binomial Model

The Negative Binomial models were developed using SPSSv16 (SPSS (2008)). Before developing the models, one needs to test the correlation between independent variables to minimize any possible bias that might arise due to eliminating these variables in the development of models. Statistical tests were conducted by computing Pearson correlation coefficient to examine the correlation between the independent variables. Table 6.2 summarizes the Pearson correlation coefficients between each independent variable. In this research, two variables were considered to be strongly correlated to each other if the computed Pearson correlation coefficient is less than -0.3 or greater than +0.3 (significance value less than 0.01 for the considered data). The predictor variables that exhibited a strong correction with other predictor variables were omitted from the modeling process to minimize any possible multicollinearity.

In the modeling process, independent variables with Wald Chi-Square value less than 1 or the level of significance greater than 0.05 (95 percent confidence level) were considered to have a statistically insignificant effect on dependent variables. These parameters were examined for each independent variable to eliminate those that have a statistically insignificant effect on the dependent variable. Table 6.3 shows the parameter estimates of the Negative Binomial models developed to predict TC, IC and PDO crashes on urban road links. It can be observed from the Table 6.3 that the statistically significant independent variables at 95% confidence interval varied between the different models developed based on TC, IC and PDO models. These variations observed can be concomitant with the severity involved in the crash.

TABLE 6.2: Summary of Pearson correlation coefficients

	Pearson Correlation Coefficients																					
	NL	LTP	SL	MEDIAN	BUS	CC	IND	INNOV	INS	MH	MUDEV	MUDIS	MF	NSD	OD	PUD	RESD	RM	ROW	RURD	SF	UR	URC
NL	1.00																						
LTP	0.69	1.00																					
SL	0.16	0.20	1.00																				
MEDIAN	0.50	0.43	0.01	1.00																			
BUS	0.05	0.08	0.04	0.10	1.00																		
CC	0.09	0.10	-0.01	0.13	-0.02	1.00																	
IND	-0.08	-0.15	-0.02	-0.06	-0.01	-0.02	1.00																
INNOV	0.12	0.14	0.12	0.16	-0.01	0.05	-0.03	1.00															
INS	-0.15	-0.11	-0.10	-0.10	0.02	0.00	-0.04	-0.03	1.00														
MH	0.03	-0.10	0.21	-0.09	-0.03	-0.06	0.07	-0.07	0.09	1.00													
MUDEV	0.11	0.00	-0.12	0.00	-0.02	-0.06	-0.02	-0.05	-0.12	0.03	1.00												
MUDIS	0.05	0.04	-0.08	-0.06	-0.02	0.30	-0.03	-0.04	-0.05	-0.08	0.11	1.00											
MF	0.04	0.02	-0.02	-0.04	0.00	-0.05	-0.03	-0.01	-0.03	-0.06	-0.03	-0.07	1.00										
NSD	-0.01	-0.10	-0.03	-0.09	-0.02	-0.01	-0.01	-0.05	0.13	0.16	0.45	0.00	-0.01	1.00									
OD	-0.01	-0.06	-0.11	-0.01	0.00	-0.02	-0.02	-0.01	0.02	-0.05	0.06	-0.01	0.03	0.01	1.00								
PUD	0.03	0.06	0.10	0.08	-0.01	0.01	-0.01	0.07	-0.02	-0.04	-0.03	-0.02	-0.03	-0.02	-0.01	1.00							
RESD	-0.06	0.02	-0.08	-0.08	-0.01	-0.02	-0.03	-0.07	0.29	-0.18	-0.05	-0.02	-0.02	-0.02	-0.03	-0.02	1.00						
RM	-0.14	-0.05	-0.01	0.05	-0.01	-0.02	0.00	-0.03	0.05	-0.09	-0.10	-0.08	-0.05	-0.07	-0.04	-0.02	0.02	1.00					
ROW	-0.04	0.06	-0.02	0.07	0.00	-0.03	-0.04	-0.06	-0.13	-0.24	-0.07	-0.11	0.01	-0.06	-0.02	-0.02	-0.03	0.17	1.00				
RURD	-0.03	0.03	-0.01	0.04	-0.01	0.01	0.00	-0.04	-0.01	-0.09	-0.04	-0.04	-0.03	-0.01	-0.02	0.01	0.09	0.25	0.00	1.00			
SF	0.02	-0.01	0.09	-0.06	-0.01	-0.02	-0.01	0.02	-0.01	-0.03	-0.03	-0.03	-0.02	-0.02	-0.02	0.04	-0.02	-0.02	-0.02	-0.01	1.00		
UR	0.06	-0.01	-0.15	-0.06	-0.02	-0.05	-0.02	-0.03	-0.10	-0.03	0.25	0.09	-0.02	0.10	-0.01	-0.02	-0.06	-0.10	-0.07	-0.04	-0.02	1.00	
URC	0.06	-0.01	-0.14	0.02	-0.02	-0.08	-0.03	-0.07	-0.14	-0.02	0.52	0.20	-0.02	0.07	0.02	-0.03	-0.06	-0.11	-0.07	-0.05	-0.03	0.49	1.00

TABLE 6.3: Summary of parameter estimates of Negative Binomial models

Parameter Estimates for Predicting Total Crashes								
Parameter	B	Std. Error	95% Wald Confidence Interval		Hypothesis Test		Goodness of Fit	
			Lower	Upper	Wald Chi-Square	Sig.	QIC	QICC
(Intercept)	1.6479	0.1738	1.3073	1.9885	89.91	<0.01	285.268	300.053
NL	0.1559	0.0456	0.0665	0.2453	11.67	<0.01		
BUS	-0.0003	0.0001	-0.0005	-0.0001	5.93	0.01		
CC	0.0393	0.0076	0.0245	0.0541	26.99	<0.01		
IND	-0.0013	0.0003	-0.0019	-0.0006	14.08	<0.01		
INNOV	-0.1315	0.0608	-0.2506	-0.0123	4.67	0.03		
MF	0.0100	0.0019	0.0062	0.0138	26.30	<0.01		
PUD	-0.0059	0.0017	-0.0092	-0.0026	12.27	<0.01		
RM	-0.4794	0.2602	-0.9894	0.0306	3.39	0.07		
SF	-0.0001	0.0000	-0.0002	0.0000	11.36	<0.01		
Parameter Estimates for Predicting Injury Crashes								
(Intercept)	0.9482	0.0859	0.7799	1.1164	121.97	<0.01	301.02	310.308
MEDIAN	0.2830	0.1164	0.0549	0.5112	5.91	0.02		
BUS	-0.0019	0.0005	-0.0029	-0.0008	12.17	<0.01		
INNOV	-0.2259	0.0967	-0.4155	-0.0364	5.46	0.02		
MF	0.0077	0.0024	0.0030	0.0125	10.28	<0.01		
OD	-0.0137	0.0071	-0.0277	0.0003	3.69	0.05		
PUD	-0.0068	0.0023	-0.0113	-0.0022	8.58	<0.01		
Parameter Estimates for Predicting PDO Crashes								
(Intercept)	1.2319	0.1841	0.8711	1.5928	44.77	<0.01	289.606	300.954
NL	0.1578	0.0483	0.0631	0.2524	10.68	<0.01		
CC	0.0475	0.0080	0.0319	0.0631	35.68	<0.01		
IND	-0.0022	0.0006	-0.0033	-0.0011	15.56	<0.01		
MF	0.0112	0.0019	0.0075	0.0149	35.75	<0.01		
PUD	-0.0064	0.0018	-0.0100	-0.0029	12.49	<0.01		
RM	-0.5572	0.2605	-1.0679	-0.0465	4.57	0.03		
SF	-0.0002	0.0001	-0.0003	-0.0001	13.66	<0.01		

6.2.2 BPNN Model

The BPNN models were built and trained using MATLAB Neural Network Toolbox (MATLAB, 2012). Default settings were used for all the parameters except for the number of epoch (set to 5000 iterations) and learning rate (set to 0.05). To evaluate the best number of neurons in the hidden layer and also the performance ratio, BPNN models with hidden neurons = 14, 15, 16…..48, 49, 50 with performance ratio's = 0.05, 0.15, 0.25….0.95 were tried. Tangent sigmoid function and purelin function were used as transfer functions for both hidden layer and output layer, respectively with Bayesian-Regulation Back-propagation function as training function. The network was trained using the database that was used to develop statistical models and validated using the test database. The performance of the network was evaluated by calculating errors. It was observed that the BPNN with performance ratio of 0.95 and hidden layer with 18, 22, 32, 33 and 50 neurons performed well after training.

6.2.3 Comparison of Predictive Performance

To compare the predictive performance of the models developed, the data for the 30 road links that was randomly selected for validation are used. Four different criteria were used to evaluate the predictive performance of the models developed. They are:

- Average Error or Mean Absolute Deviation (MAD)

$$MAD = \frac{1}{n}\sum_{i=1}^{n}|\hat{c}_i - c_i|$$

- 50th Percentile Error = An error value below which 50% of the observations fall
- 85th Percentile Error = An error value below which 85% of the observations fall
- Root Mean Squared Error (RMSE)

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\hat{c}_i - c_i)^2}$$

where, ‘n’ is the sample size, $\hat{c}_i$ and c_i are predicted and observed number of crashes respectively.

For all the four evaluation criteria MAD, 50th percentile error, 85th percentile error and RMSE, the values closer to zero indicate better model performance in prediction.

The predictive performances of Negative Binomial models and BPNN models in predicting total, injury and PDO crashes are summarized in Table 6.4. For all the three different types of crash prediction models (total, injury and PDO), the BPNN models performed well compared to the Negative Binomial models.

From Table 6.4, the performance criteria of the models to predict total crashes i.e., MAD, 50th percentile error, 85th percentile error and RMSE are observed to be lower for BPNN33 (33 neurons in the hidden layer) model (4.02, 3.39, 6.53 and 4.86 respectively) when compared the Negative Binomial model (4.31, 4.17, 7.00 and 5.10) and also other BPNN models developed. The performance criteria of the models to predict injury crashes are observed to be lower for BPNN18 (18 neurons in the hidden layer) model (1.62, 1.24, 2.45 and 2.05) when compared the Negative Binomial model (1.73, 1.56, 2.58 and 2.23) and also other BPNN models developed. Similarly, the performance criteria of the models to predict PDO crashes are observed to be lower for BPNN50 (50 neurons in the hidden layer) model (2.96, 2.96, 4.44 and 3.38) when compared the Negative Binomial model (3.18, 2.96, 4.44 and 3.38) and also other BPNN models developed.

The above observations clearly indicate that the BPNN33 model, BPNN 18 model and BPNN50 model have better performance in predicting total, injury and PDO crashes, respectively when compared to Negative Binomial and other BPNN models.

From Table 6.4, The MAD/average errors in predicting total, injury and PDO crashes using Negative Binomial models are 4.31, 1.73 and 3.18, respectively when compared to 4.02, 1.62 and 2.86 for BPNN33 (with 33 neurons in hidden layer) model. The 50th percentile errors in predicting total, injury and PDO crashes using Negative Binomial models are 4.17, 1.56 and 3.38, respectively when compared to 3.39, 1.22 and 2.74 for BPNN33 model. The 85th percentile errors in predicting total, injury and PDO crashes using Negative Binomial models are 7.00, 2.58 and 4.53, respectively when compared to 6.30, 2.47 and 2.44 for BPNN50 model. The RMSE in predicting total, injury and PDO crashes using Negative Binomial models are 5.10, 2.23, 3.54, respectively when compared to 4.86, 2.06 and 3.34 for BPNN33 model.

Unlike Negative Binomial models, in the case of BPNN models, three different models are not required to predict total, injury and PDO crashes, the BPNN models are designed such that a single model would predict all the three types of crashes. Overall, the performance criteria for the BPNN33 model are lower than any other model except for the 85th percentile error. This means that the BPNN33 has the best performance in predicting crashes.

TABLE 6.4: Performance criteria comparison for Negative Binomial and BPNN models (Microscopic)

Total Crashes				
Model	MAD	50th Percentile Error	85th Percentile Error	RMSE
Negative Binomial	4.31	4.17	7.00	5.10
BPNN18	4.02	3.92	6.79	4.92
BPNN22	4.01	4.05	6.44	4.91
BPNN32	4.00	3.54	6.78	4.99
BPNN33	4.02	3.39	6.53	4.86
BPNN50	4.00	4.22	6.30	4.86
Injury Crashes				
Model	MAD	50th Percentile Error	85th Percentile Error	RMSE
Negative Binomial	1.73	1.56	2.58	2.23
BPNN18	1.62	1.24	2.45	2.05
BPNN22	1.64	1.37	2.48	2.09
BPNN32	1.72	1.55	2.62	2.13
BPNN33	1.62	1.22	2.48	2.06
BPNN50	1.66	1.46	2.47	2.06
PDO Crashes				
Model	MAD	50th Percentile Error	85th Percentile Error	RMSE
Negative Binomial	3.18	3.38	4.53	3.54
BPNN18	2.92	3.34	4.80	3.44
BPNN22	2.89	2.76	4.86	3.39
BPNN32	2.88	2.87	4.89	3.39
BPNN33	2.86	2.74	4.64	3.34
BPNN50	2.96	2.96	4.44	3.38

6.2.4 Summary: Microscopic Crash Prediction Models

Analysis of microscopic crash prediction models showed that number of lanes and land use characteristics such as commercial centers and multi-family residential are positively correlated to the total number of crashes and PDO crashes on road links. Whereas, the influence of land use characteristics such as industrial, planned unit development, residential mobile and single family residential are negatively correlated to both total number of crashes and PDO crashes on road links, indicating that an increase in the influence of these characteristics tends to lower the PDO crashes there by reducing total number of crashes. Similarly, the presence of median and influence of multi-family residential land use are positively correlated to injury crashes. Whereas, the influence of land use characteristics such as business districts, innovative centers, office district and planned unit development are negatively correlated to injury crashes on road links.

Validation data (data for 30 road links) were used to validate and compare the performance of the models developed. Both Negative Binomial and BPNN models performed well in predicting the crashes (total, injury and PDO crashes). When MAD, 50th percentile error, 85th percentile error and RMSE are considered for performance evaluation, BPNN33, BPNN18 and BPNN50 models performed better in predicting total crashes, injury crashes and PDO crashes, respectively. As the performance criteria for the BPNN33 model are lower than any other model except for the 85th percentile error, one can infer that BPNN33 model has outperformed all other models in predicting crashes (total, injury and PDO).

6.3 Microscopic Travel Demand/AADT Estimation Models

Two different techniques were incorporated in developing the models to predict AADT on the urban streets. The variance has to be computed based on the dependent variable for the road links and are compared with their respective means. If data is observed to be over-dispersed (α greater than zero) and not spatially correlated, a Negative Binomial distribution will be more suitable to estimate dependent variable. The dispersion parameter ‘α’ for the AADT’s was observed to be 6,146. Therefore, Negative Binomial distribution was used to model AADT’s. In conjunction with the Negative

Binomial model, a neural network model was also developed using the same dataset. Both the models are compared for performance evaluation. The following sections briefly describe the Negative Binomial and neural network models developed in this research.

6.3.1 Negative Binomial Model

Since the nature of the data is same as that of microscopic crash prediction models, the Pearson correlation coefficients to examine the correlation between independent variables to predict AADT will be no different to the Pearson correlation coefficients summarized in Table 6.2. Independent variables with Wald Chi-Square value less than 1 or the level of significance greater than 0.05 (95 percent confidence level) were considered to have a statistically insignificant effect on the dependent variable (AADT or link volume). These parameters were examined for each independent variable to eliminate those that have a statistically insignificant effect on the dependent variable.

The statistical parameters for the model to estimate AADT on a given road link are shown in Table 6.5. The QIC and QICC were used to assess the goodness of fit.

TABLE 6.5: Final model parameters to estimate the AADT's on road links

Variable	Coefficient	Hypothesis Test	
		Wald Chi-Square	Significance
(Intercept)	7.67666	41,535.95	<0.01
FC	0.65295	2,715.13	<0.01
BUS	-0.00010	55.82	<0.01
CC	0.02077	8.49	<0.01
NSD	-0.05417	15.34	<0.01
PUD	0.00126	14.32	<0.01
RURD	6.14833	19.89	<0.01
SF	0.00007	5.98	0.01
UR	-0.04765	4.45	0.03
Quasi Likelihood under Independence Model Criterion (QIC) = 13.20			
Corrected Quasi Likelihood under Independence Model Criterion (QICC) = 30.80			

6.3.2 BPNN Model

The BPNN models are built and trained using MATLAB Neural Network Toolbox (MATLAB, 2012). Default settings were used for all the parameters except for the number of epoch (set to 5000 iterations) and learning rate (set to 0.05). To evaluate the best number of neurons in the hidden layer and also the performance ratio, BPNN models with hidden neurons = 14, 15, 16…..48, 49, 50 with performance ratio's = 0.05, 0.15, 0.25….0.95 were tried. The database used to develop Negative Binomial models was used to develop BPNN model to maintain consistency for performance evaluation.

All the 24 independent variables are given as the input vector to the network. So, the input layer has the number of neurons each of which corresponds to an independent variable in the model. The output layer has the number of neurons equal to number of outputs, which is equal to one (i.e. AADT). Tangent sigmoid function was used as a transfer function for both the hidden layer and the output layer with Bayesian-Regulation BP function as training function. The network was trained using the database that was used to develop statistical models and validated using the test database. The performance of the network was evaluated by calculating errors. The number of neurons in the hidden layer is not limited to a fixed number. So, appropriate number of neurons was evaluated by changing the number of neurons in the hidden layer until the network performed well after training. It was observed that the hidden layer with 14, 18 and 36 neurons performed well after training.

6.3.3 Validation and Performance Evaluation

The validation data (30 counts) are used to calculate outputs from both statistical model and neural network models. Similarly, AADT's were also calculated using the traditional four-step method for these 30 road links. Assuming that the weekend traffic is 25% less than the weekday traffic, the outputs from the four-step method, i.e., Annual Average Weekday Traffic (AAWT) are multiplied with 0.92 to evaluate AADT's. The predicted AADT outputs obtained from all the models developed, including four-step method, are compared with the observed counts to calculate errors. The percent error in estimation was calculated using the following Equation (6.1). Figure 6.1 represents the frequency of percent errors from the outputs of all the models in estimating AADT's.

$$Percent\ Error = \frac{Model\ Volume - Volume\ from\ Counts}{Volume\ from\ Counts} * 100 \qquad \text{...Equation (6.1)}$$

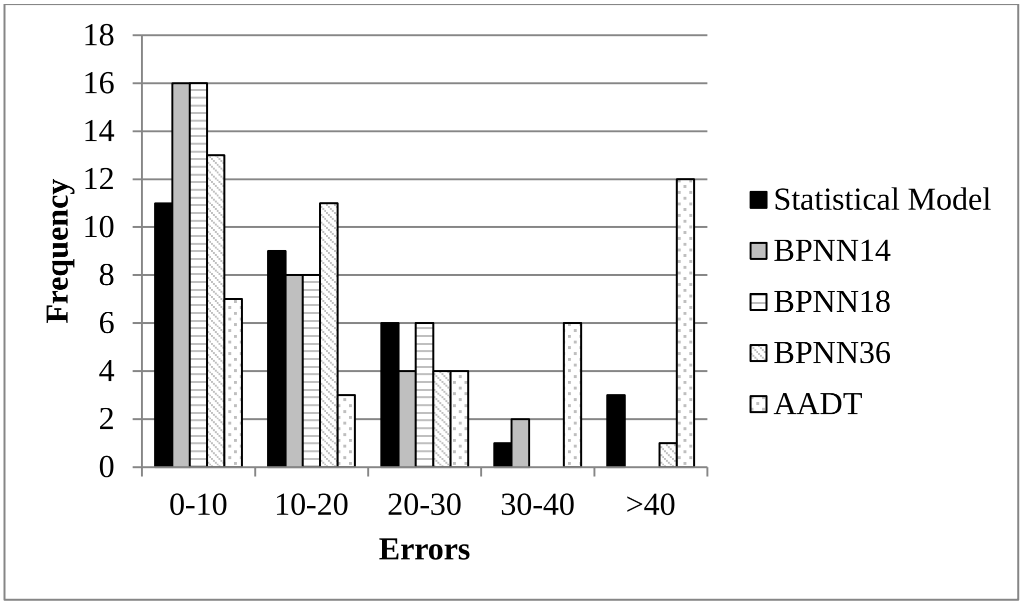

FIGURE 6.1: Frequency of percent errors

The percent error in predicting AADT's using statistical model developed vary from 0.19% to 45.14%. It was observed also that most of the largest errors in predicting AADT have occurred for the road links which had higher influence of business district land use. Similarly, for the BPNN14, BPNN18 and BPNN36 models, the percent errors range from 0.41% to 34.12%, 0.09% to 28.37% and 1.04% to 54.31%, respectively. However, when these ranges are compared with four-step method, 0.18% to 196.22%, the percent error range for four-step method was observed to be much higher than all the models developed for predicting AADT in this research.

TABLE 6.6: Summary of AADT estimation errors

Model	Percent Error in Estimating AADT			
	Average	50th Percentile	70th Percentile	95th Percentile
Statistical Model	16.70%	15.79%	22.10%	37.75%
ANN (14)	12.87%	10.91%	18.01%	26.95%
ANN (18)	13.25%	11.40%	17.37%	23.74%
ANN (36)	15.94%	12.56%	17.46%	41.13%
AADT (four step)	33.03%	32.13%	45.74%	66.68%

6.3.4 Summary: Microscopic Travel Demand/AADT Estimation Models

Analysis of microscopic travel demand/AADT estimation models showed that lane type and land use characteristics such as business district, commercial centers, neighborhood service district, planned unit district, rural district, single family residential and urban residential play a statistically significant role in estimating AADT's. Results indicate that the influence of business district, neighborhood service district and urban residential are negatively correlated to AADT, indicating a higher influence of these characteristics on a road link would result in lower traffic volume.

Permanent counts on 30 roads were used to validate the models developed and are also compared with traditional four-step method outputs. ANN 14 model performed better in estimating AADT when compared with any other model with a very low average error (12.87%) and 50th percentile error (10.91%). However, when 70th percentile and 95th percentile errors were considered in performance evaluation, ANN 18 model (17.37% & 23.74%) outperformed all other models.

The 95th percentile error values for statistical and ANN (14 & 18) models are around 38% and 25%, respectively when compared with 67% for four-step method. From the observed magnitude of the errors, one can infer that the proposed methodology that incorporated principle of demographic gravitation would yield better results in estimating AADT when compared to the traditional methods.

CHAPTER 7: CONCLUSIONS

In this research, macroscopic and microscopic models were developed with emphasis on land use characteristics to estimate travel demand and crashes. The microscopic models help evaluate the link level performance, whereas the macroscopic models help evaluate the overall performance of an area. The microscopic models can assist in design and operational analysis while macroscopic models can assist in planning (future developments and/or rezoning).

The proposed methodology involved scientific principles, statistical and artificial intelligent techniques. The method for developing macroscopic models differs from microscopic models. The areas of land use characteristics were considered in developing macroscopic models, whereas the principle of demographic gravitation has been incorporated in developing microscopic models. Statistical and back-propagation neural network techniques were used in developing models and are compared for performance evaluation.

Results obtained indicate that models based on Negative Binomial distribution yield better travel demand and crash estimates as the count data used in this research is observed to be over-dispersed. However, results from validation and performance evaluation indicate that neural network models yielded better results in estimating both crashes and travel demand (microscopic and macroscopic level) than any other method considered in this research and much better results are likely with larger data sets.

From the analysis of macroscopic crash prediction models, it can be concluded that strong associations could be established from crash estimation models developed based on land use characteristics to estimate the total number of crashes, the number of injury crashes and the number of property damage only crashes in a TAZ at a 95 percent confidence level. Maintaining a delicate balance between different land uses in a TAZ or area based on outcomes from this research would improve safety and maximize derived benefits.

From the analysis of microscopic models, it can be concluded that the proposed methodology that incorporated principle of demographic gravitation would yield better results in estimating travel demand/AADT and crashes when compared to the traditional methods. The neural network model yielded better results in estimating travel demand/AADT and crashes than any other method considered in this research and much better results are likely with larger data sets. The neural networks are fast and do not need any formulas or conditions. Their adaptive nature helps them adapt to the data variations and learn input characteristics yielding better results. However, unlike statistical models and traditional four-step method in case of estimating AADT, the neural network model is a black box model which has a predictive value solely based on observations but does not provide any explanation. Therefore, statistical models or a traditional four-step method are more appropriate when one wants to understand the role of explanatory variables.

Overall, from the results and observed magnitude of errors in estimating crashes, it can be concluded that the proposed methodology that incorporated the principle of demographic gravitation in estimating crashes not only yield better results, but also helps evaluate the influence of land use characteristics on AADT and crashes on the urban road links. Land use characteristics were found to have better predictive capability than other demographic, socio-economic or on-network characteristics considering in this research.

The developed methodology and results can be used to incorporate safety into long range transportation plans and land use decisions so as to minimize anticipated crashes in the future. The neural network application can be used for better predictions, whereas the statistical models could be used for mathematical formulation or explanation. The models developed using the methodology can also be used to examine the effect of changes in land use characteristics (new development or re-zoning) on safety, identify appropriate solutions to improve travel patterns and traffic safety, and also help planners to plan, propose and prioritize infrastructure projects for future improvements.

In travel demand models at microscopic level, the effect of mode choice was not considered in this research. The presence or access to public transportation systems could play a vital role in estimating travel demand, which needs an investigation. Further, research needs to be carried out to determine the correlation between the macroscopic and microscopic models that were developed in this research.

REFERENCES

Abdel-Aty, M. A., and A. E. Radwan. Modeling traffic accident occurrence and involvement. Accident Analysis and Prevention, Vol. 32, 2000, pp. 633-642.

Aguero-Valverde, J., and P. P. Jovani. Spatial analysis of fatal and injury crashes in Pennsylvania. Accident Analysis and Prevention, Vol. 38, 2006, pp. 618-625.

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

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

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

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.

Caliendo, C., Guida, M., and A. Parisi. A crash-prediction model for multilane roads. Accident Analysis and Prevention, Vol. 39, 2007, pp. 657-670.

Chin, H. C., and M. A. Quddus. Applying the Random Effect Binomial Model to Examine Traffic Accident Occurrence at Signalized Intersections. Accident Analysis & Prevention, Vol. 35(2), 2007, pp. 253-259.

Cichocki, A., and R. Unbehauen, Neural networks for optimization and signal processing, Wiley, Chichester, 1993.

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.

Federal Highway Administration (FHWA). A Message to Customers & Partners, Who We Are & What We Do, Publication No. FHWA-hcm-03-003. 2006. Available online at: http://www.fhwa.dot.gov/whoweare/message.htm

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.

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

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

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.

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.

Gilmore, J. F., K. J. Elibiary and M. Abe, Traffic management applications of neuro networks systems, Working notes, AAAI-93 Workshop on Al in Intelligent Highways Systems, 1993, pp. 85-95.

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.

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. Available Online at: http://www.ctre.iastate.edu/pubs/crossroads/122impact.pdf

Greibe, P. Accident Prediction Models for Urban Roads. Accident Analysis & Prevention, Vol. 35(2), 2003, pp. 173-185.

Ham, F. M., and I. Kostanic, Principles of neurocomputing for science and engineering, New York: McGraw-Hill, 2001.

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

Horowitz, A. J., and D. D. Farmer, Critical Review of Statewide Travel Forecasting Practice.In Transportation Research Record: Journal of the Transportation Research Board, No. 1685, Transportation Research Board of the National Academics, Washington, D.C., 1999, pp. 13-20.

Hua, J., and A. Faghri, Development of neural signal control system—toward intelligent traffic signal control, Transportation Research Record, Vol. 1497, 1995, pp. 53-61.

Ivan, J. N., Wang, C., and N. R. Bernardo. Explaining two-lane highway crash rates using land use and hourly exposure. Accident Analysis and Prevention, Vol. 32, 2000, pp. 787-795.

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.

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.

Kim, K., and E. Yamashita. Motor vehicle crashes and land use: Empirical analysis from Hawaii. Transportation Research Record, vol. 1794, 2002, pp. 73-79.

Kim, K., I. M. Brunner and Y. Yamashita, Influence of Land Use, Population, Employment, and Economic Activity on Accidents, Transportation Research Record, Vol. 1953, 2006, pp. 56-64.

Kim, K., P. Punt, and E. Yamashita, Measuring Influences of Demographic and Land Use Variables in Honolulu, Hawaii, Transportation Research Record, Vol. 2147, 2010, pp. 9-17.

Ladron de Guevara, F., Washington, S. P., and J. Oh. Forecasting Crashes at the Planning Level: Simultaneous Negative Binomial Crash Model Applied in Tucson, Arizona. In Transportation Research Record: Journal of the Transportation Research Board, Vol. 1897, Transportation Research Board of the National Academies, Washington, D.C., 2004, pp. 191 – 100.

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.

Ledoux, C., An urban traffic control system integrating neural networks, Eighth International Conference on Road Traffic Monitoring and Control London, 1996, pp. 197-201.

Ledoux, C., F. Boillot, S. Sellam and P. Gallinari, On the use of neural networks techniques for traffic flow modelling, Second World Congress on Applications of Transport Telematics and Intelligent Vehicle Highway Systems Yokohama, Japan, 1995.

Levine, N., Kim, K. E., and L. H. Nitz. Spatial analysis of Honolulu motor vehicle crashes: I. Spatial patterns. Accident Analysis and Prevention, Vol. 27, 1995, pp. 663-674.

Levine, N., Kim, K. E., and L. H. Nitz. Spatial analysis of Honolulu motor vehicle crashes: II. Spatial patterns. Accident Analysis and Prevention, Vol. 27, 1995, pp. 675-685.

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.

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.

Liang, F. An effective Bayesian neural network classifier with a comparison study to support vector machine. Neural Computation, Vol. 15, 2003, pp. 1959-1989.

Liang, F. Bayesian neural networks for nonlinear time series forecasting. Statistics and Computing, Vol. 15, 2005, pp. 13-29.

Lingras, P., Sharma, S. C., 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.

Ma, J., Kockelman, K. M., and P. Damien. A Multivariate Poisson-Lognormal Regression Model for Prediction of Crash Counts by Severity, Using Bayesian Methods. Accident Analysis & Prevention, Vol. 40(3), 2008, pp. 964-975.

MATLAB Neural Network Toolbox, 2012 (a). The MathWorks Inc., Natick, MA.

McCord, M., Yang, Y., Jiang, Z., Coifman, B., and P. Goel. Estimating AADT from Satellite Imagery and Air Photos: Empirical Results. . In Transportation Research Record: Journal of the Transportation Research Board, No. 1855, Transportation Research Board of the National Academics, Washington, D.C., 2003, pp. 136-142

Miaou, S. P. The relationship between truck accidents and geometric design of road sections: Poisson versus Negative Binomial regressions. Accident Analysis and Prevention, Vol. 26, 1994, pp. 471-482.

Miaou, S. P., and D. Lord. Modeling Traffic Crash Flow Relationships for Intersections: Dispersion Parameter, Functional Form, and Bayes versus Empirical Bayes Methods, In Transportation Research Record: Journal of the Transportation Research Board, Vol. 1840, Transportation Research Board of the National Academies, Washington, D.C., 2003, pp. 31-40.

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

Mitra, S., and S. Washington. On the Nature of Over-dispersion in Motor Vehicle Crash Prediction Models. Accident Analysis & Prevention, Vol. 39(3), 2007, pp. 459-468.

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.

Naderan, A., and J. Shahi. Aggregate crash prediction models: Introducing crash generation concept. Accident Analysis and Prevention, Vol. 42, 2010, pp. 339-346.

Nakatsuji, T., and T. Kaku, Development of a self-organizing traffic control system using neural network models, Transportation Research Record, Vol. 1324, 1995, pp. 137-145.

National Highway Traffic Safety Administration (NHTSA), National Center for Statistics and Analysis, 2008. Traffic Safety Annual Assessment-Highlights, Traffic Safety Facts, 2009, Washington DC. Available online at: http://www-nrd.nhtsa.dot.gov/pubs/811172.pdf.

Nilsson, N. J., Problem-Solving Methods in Artificial Intelligence, New York: McGraw-Hill, 1971.

Noland, R. B., and L. Oh. The effect of infrastructure and demographic change on traffic-related fatalities and crashes: a case study of Illinois county-level data. Accident Analysis and Prevention, Vol. 36, 2004, pp. 525-532.

Noland, R. B., and M. A. Quddus. A spatially disaggregate analysis of road casualties in England. Accident Analysis and Prevention, Vol. 36, 2003, pp. 973-984.

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

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

Quddus, M. A. Modelling area-wide count outcomes with spatial correlation and heterogeneity: An analysis of London crash data. Accident Analysis and Prevention, vol. 40, 2008, pp. 1486-1497.

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.

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.

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

Seaver, W. L., Chatterjee, A., and M. L. Seaver. Estimation of traffic volume on rural local roads. In Transportation Research Record: Journal of the Transportation Research Board, No. 1719, Transportation Research Board of the National Academies, Washington, D.C., 2000, pp. 121-128.

Selby, B., and M. K. Kockelman. Spatial Prediction of AADT in Unmeasured Location by Universal Kriging. In Transportation Research Board Annual Meeting 2011, Paper #11-1665.

Sharma, S. C., and R. R. Allipuram. Duration and Frequency of Seasonal Traffic Counts. Journal of Transportation Engineering, Vol. 119, 1993, pp. 344-359.

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.

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.

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.

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.

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.

Sharma, S. C., Lingras, P., Xu, F., and P. Kilburn. Application of Neural Networks to Estimate AADT of Low-Volume Roads. Journal of Transportation Engineering, Vol. 127, 2001, pp. 426-432.

Siddiqui, C., M. Abdel-Aty, and H. Huang, Aggregate Nonparametric Safety Analysis of Traffic Zones, Accident Analysis & Prevention, Vol. 45(0), 2010, pp. 317-325.

Smith, B. L., and M. J. Demetsky, Short-term traffic flow prediction: neural network approach, Transportation Research Record, Vol. 1453, 1996, pp. 98-104.

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

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

SPSS 16.0, Command Syntax Reference 2008, SPSS Inc., Chicago, IL.

Srinivasan, D., C. Wai Chan, and P. G. Balaji, Computational intelligence-based congestion prediction for a dynamic urban street network, Neurocomputing, Vol. 72, 2009, pp. 2710-2716.

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.

Stewart, J. Q. Demographic Gravitation: Evidence and Application. Sociometry, Vol. 11, 1948, pp. 31-58.

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

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.

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

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.

Ukkusuri, S., L. F. Miranda-Moreno, K. Ramadurai and J. Isa-Tavare, The Role of Built Environment on Pedestrian Crash Frequency, Safety Science, Vol. 50(4), 2012, pp. 1141-1151.

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.

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

Washington, S., van Schalwyk, I., Mitra, S., Meyer, M., Dumbaugh, E., Zoll, M., 2006. Incorporating safety into long-range transportation planning. NCHRP Report 546. http://onlinepubs.trb.org/onlinepubs/nchrp/nchrp_rpt_546.pdf (accessed 04.06.2011).

Wier, M., Weintraub, J., Humphreys, E., Seto, E., and R. Bhatia. An Area-Level Model of Vehicle-Pedestrian Injury Collisions with Implications for Land Use and Transportation Planning. Accident Analysis & Prevention, Vol. 41(1), 2009, pp. 137-145.

Wild, D., Short-term forecasting based on a transformation and classification of traffic volume time series, International Journal of Forecasting, Vol. 13, 1997, pp. 63-72.

Wood, G. R. Confidence and Prediction Intervals for Generalized Linear Accident Models. Accident Analysis & Prevention, Vol. 37(2), 2005, pp. 267-273.

Wood, G. R. Generalized Linear Accident Models and Goodness of Fit Testing. Accident Analysis & Prevention, Vol. 34(4), 2002, pp. 417-427.

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.

Xie, Y., Lord, D., and Y. Zang. Predicting motor vehicle collisions using Bayesian neural network models: An empirical analysis. Accident Analysis & Prevention, Vol. 39, 2007, pp. 922-933.

Yin, H., S.C. Wong, J. Xu and C. K. Wong, Urban traffic flow prediction using a fuzzy-neural approach, Transportation Research, Vol. 10C, 2002, pp. 85-98.

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.

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.

Zhong, M., Lingras, P., and S. Sharma. Estimation of missing traffic counts using factor, genetic, neural, and regression techniques. Transportation Research Part: C, Vol. 12, 2004, pp. 139-166.