

3-29-2012

Improved Annual Average Daily Traffic (AADT) Estimation for Local Roads using Parcel-Level Travel Demand Modeling

Tao Wang

Florida International University, tao5431@gmail.com

DOI: 10.25148/etd.FI12050229

Follow this and additional works at: <https://digitalcommons.fiu.edu/etd>

Recommended Citation

Wang, Tao, "Improved Annual Average Daily Traffic (AADT) Estimation for Local Roads using Parcel-Level Travel Demand Modeling" (2012). *FIU Electronic Theses and Dissertations*. 623.

<https://digitalcommons.fiu.edu/etd/623>

This work is brought to you for free and open access by the University Graduate School at FIU Digital Commons. It has been accepted for inclusion in FIU Electronic Theses and Dissertations by an authorized administrator of FIU Digital Commons. For more information, please contact dcc@fiu.edu.

FLORIDA INTERNATIONAL UNIVERSITY

Miami, Florida

IMPROVED ANNUAL AVERAGE DAILY TRAFFIC (AADT) ESTIMATION FOR
LOCAL ROADS USING PARCEL-LEVEL TRAVEL DEMAND MODELING

A dissertation submitted in partial fulfillment of the

requirements for the degree of

DOCTOR OF PHILOSOPHY

in

CIVIL ENGINEERING

by

Tao Wang

2012

To: Dean Amir Mirmiran
College of Engineering and Computing

This dissertation, written by Tao Wang and entitled Improved Annual Average Daily Traffic (AADT) Estimation for Local Roads using Parcel-Level Travel Demand Modeling, having been approved in respect to style and intellectual content, is referred to you for judgment.

We have read this dissertation and recommend that it be approved.

Mohammed Hadi

L. David Shen

Zhenmin Chen

Albert Gan, Major Professor

Date of Defense: March 29, 2012

The dissertation of Tao Wang is approved.

Dean Amir Mirmiran
College of Engineering and Computing

Dean Lakshmi N. Reddi
University Graduate School

Florida International University, 2012

DEDICATION

I dedicate this dissertation to my family for their support and love.

ACKNOWLEDGMENTS

My sincere gratitude and appreciation go first to my major advisor, Dr. Albert Gan, for his continued guidance, encouragement, inspiration, and support during my doctoral study at Florida International University. What I have learned from Dr. Gan will be invaluable to my future career and the rest of my life.

I am also grateful to the members of my dissertation committee, Dr. Mohammed Hadi, Dr. David Shen, and Dr. Zhenmin Chen for serving on my committee, and for their comments, suggestions, and interest in my research. Special thanks go to Dr. Fang Zhao, who taught me travel demand analysis which forms the foundation to this dissertation. She devoted her whole life to teaching and research, and has set a professional and respectable example for me.

I would like to extend my deep appreciation to Ms. Shanshan Yang, Dr. Yan Xiao, Ms. Jinyan Lu, Ms. Meng Ma, Dr. Hongbo Chi, and Dr. Wanyang Wu for their assistance and support. I would also like to thank Dr. Priyanka Alluri for her help to improve my dissertation. Special thanks go to my friend Ms. Linna Tang, who always cares about my graduation and encouraged me to finish my dissertation.

Furthermore, I would like to acknowledge financial support for this research from the Florida Department of Transportation Research Center.

Finally, I would like to thank my family for their consistent encouragement and love. Without their patience, understanding, and support, the completion of this work would have been impossible.

ABSTRACT OF THE DISSERTATION

IMPROVED ANNUAL AVERAGE DAILY TRAFFIC (AADT) ESTIMATION FOR
LOCAL ROADS USING PARCEL-LEVEL TRAVEL DEMAND MODELING

by

Tao Wang

Florida International University, 2012

Miami, Florida

Professor Albert Gan, Major Professor

Annual Average Daily Traffic (AADT) is a critical input to many transportation analyses. By definition, AADT is the average 24-hour volume at a highway location over a full year. Traditionally, AADT is estimated using a mix of permanent and temporary traffic counts. Because field collection of traffic counts is expensive, it is usually done for only the major roads, thus leaving most of the local roads without any AADT information. However, AADTs are needed for local roads for many applications. For example, AADTs are used by state Departments of Transportation (DOTs) to calculate the crash rates of all local roads in order to identify the top five percent of hazardous locations for annual reporting to the U.S. DOT.

This dissertation develops a new method for estimating AADTs for local roads using travel demand modeling. A major component of the new method involves a parcel-level trip generation model that estimates the trips generated by each parcel. The model uses the tax parcel data together with the trip generation rates and equations provided by the *ITE Trip Generation Report*. The generated trips are then distributed to existing traffic count sites using a parcel-level trip distribution gravity model. The

all-or-nothing assignment method is then used to assign the trips onto the roadway network to estimate the final AADTs. The entire process was implemented in the Cube demand modeling system with extensive spatial data processing using ArcGIS.

To evaluate the performance of the new method, data from several study areas in Broward County in Florida were used. The estimated AADTs were compared with those from two existing methods using actual traffic counts as the ground truths. The results show that the new method performs better than both existing methods. One limitation with the new method is that it relies on Cube which limits the number of zones to 32,000. Accordingly, a study area exceeding this limit must be partitioned into smaller areas. Because AADT estimates for roads near the boundary areas were found to be less accurate, further research could examine the best way to partition a study area to minimize the impact.

TABLE OF CONTENTS

CHAPTER	PAGE
1. INTRODUCTION	1
1.1. Research Background	1
1.2. Problem Statement	2
1.3. Research Objective	4
1.4. Organization	4
2. LITERATURE REVIEW	6
2.1. Introduction	6
2.2. Traditional Factor Approach	6
2.3. Regression Modeling	11
2.4. Travel Demand Modeling	19
2.5. Image Processing	20
2.6. Machine Learning	22
2.6.1. Artificial Neural Network Approach	22
2.6.2. K-Nearest Neighbor Approach	26
2.6.3. Support Vector Regression Machines Approach	27
2.7. URS Method	28
2.8. Summary	29
3. METHODOLOGY	32
3.1. Introduction	32
3.2. Traditional Travel Demand Forecasting Model	32
3.2.1. Introduction	32
3.2.2. Trip Generation	35
3.2.3. Trips Distribution	37
3.2.4. Mode Choice	38
3.2.5. Trip Assignment	39
3.3. Parcel-level Travel Demand Analysis Model	40
3.3.1. Introduction	40
3.3.2. Network Modeling	43
3.3.3. Parcel-level Trip Generation	44
3.3.4. Parcel-level Trip Distribution	52
3.3.5. Parcel-level Trip Assignment	56
3.6. Evaluation Method	57
3.4. Summary	59
4. MODEL DEVELOPMENT	62
4.1. Introduction	62
4.2. Model Development Overview	62
4.3. Network Modeling	67
4.3.1. Data Preprocessing	67
4.3.2. Build Cube Network File	72

4.3.3. Create Centroid Centers and Connectors	75
4.3.4. Calculate Free Flow Travel Time Skim Matrix	77
4.4. Parcel-level Trip Generation	78
4.5. Parcel-level Trip Distribution	80
4.6. Parcel-level Trip Assignment	81
4.7. Summary	84
5. MODEL OUTPUT AND EVALUATION	85
5.1. Introduction	85
5.2. Evaluation Method	85
5.3. Single Study Area Comparison	86
5.4. Multiple Study Areas Comparison	89
6. SUMMARY, CONCLUSIONS AND RECOMMENDATIONS	106
6.1. Summary of Research Approach and Results	106
6.2. Conclusions	107
6.3. Recommendations for Future Research	110
REFERENCES	112
APPENDIX A	116
VITA	122

LIST OF TABLES

TABLE	PAGE
Table 3-1 Examples of Land Use Types Matching	50
Table 4-1 Input and Output Files	66
Table 4-2 Parcel Attributes and ITE Trip Generation Independent Variable Matching...	79
Table 5-1 Performance of the Three Methods for the Study Area	88
Table 5-2 Performance of the Methods without Invalid Evaluation Count Sites	89
Table 5-3 Variance Measure of the Performance	90
Table A-1 Land Use Type Matching, Independent Variables, and Rate/Equation	116

LIST OF FIGURES

FIGURE	PAGE
Figure 2-1 Example of a Simple Feedforward Neural Network.....	23
Figure 3-1 TAZs in Broward County, FL.....	34
Figure 3-2 Creation of Centroid and Centroid Connectors.....	35
Figure 3-3 Flowchart of Parcel-level Travel Demand Analysis Model.....	41
Figure 3-4 Comparison of Parcels and TAZs	45
Figure 3-5 Example of ITE Trip Generation Report	47
Figure 4-1 System Components and Procedure.....	64
Figure 4-2 Model Steps in Cube	65
Figure 4-3 Dividing the Count Sites into Estimation and Evaluation Groups.....	69
Figure 4-4 Merging the Parcels and the Traffic Count Sites	70
Figure 4-5 Centroid Connectors Incorrectly Connecting Parcels to Intersections	71
Figure 4-6 Splitting the Roadway Polylines at the Parcel Access Points.....	72
Figure 4-7 Build Highway Network from Line Shape File Dialog	73
Figure 4-8 Mismatching of the Built Network with Original Shape File.....	74
Figure 4-9 Display True Link Shape Dialog.....	75
Figure 4-10 Automatic Centroid Connectors Generation Dialog.....	76
Figure 4-11 Example of Centroid Centers and Connectors Added	77
Figure 4-12 Components of FFT Skim Matrix Calculation.....	78
Figure 4-13 Components of Parcel-level Trip Generation Step	79
Figure 4-14 Components of Parcel-level Trip Distribution Step.....	81
Figure 4-15 Components of Parcel-level Trip Assignment Step	82

Figure 4-16 Calculating the Final AADTs.....	83
Figure 5-1 Evaluation Traffic Count Sites in the Study Area.....	87
Figure 5-2 Study Area No. 1.....	91
Figure 5-3 Study Area No. 2.....	92
Figure 5-4 Study Area No. 3.....	93
Figure 5-5 Study Area No. 4.....	94
Figure 5-6 Study Area No. 5.....	95
Figure 5-7 Study Area No. 6.....	96
Figure 5-8 Study Area No. 7.....	96
Figure 5-9 Study Area No. 8.....	97
Figure 5-10 Study Area No. 9.....	98
Figure 5-11 Study Area No. 10.....	99

CHAPTER 1

INTRODUCTION

1.1. Research Background

Annual Average Daily Traffic (AADT) is the average 24-hour volume of a roadway segment over a full year. AADT is used in many transportation analyses including estimation of the economic feasibility of highway projects, estimation of highway user revenues, computation of other highway statistics such as vehicle miles traveled (VMT) and crash rates, development of improvement and maintenance programs, etc.

The most accurate method for obtaining the AADT of a roadway segment is to install an Automatic Traffic Recorder (ATR) which can provide continuous traffic count coverage at selected locations with some sensor devices such as inductive loops and microwave radar sensors to count the total volumes continuously. However, as the installation and maintenance of permanent counters are expensive, the number of permanent counters is limited. For example, there are only a total of about 300 permanent counters installed along the state roads in Florida. Therefore, it is economically infeasible to apply this method of AADT estimation on a widespread basis.

An alternative approach to estimating AADT is to use portable counts, also called short-term, seasonal, or coverage counts, with different types of portable devices such as pneumatic road tubes and microwave radar sensors. The collected short-term volumes on the interested roads are then used to calculate Average Daily Traffic (ADT). ADT is the average daily volume at a given location over a defined time period of more than one

day and less than one year. It can be converted to AADT by using some adjustment factors, as follows:

$$AADT = ADT \times AF \times SF \times GF \quad (1-1)$$

where AF is the axle correction factor, SF is the seasonal adjustment factor, and GF is the annual growth factor. This factor approach with portable counters is more economically feasible than the permanent count method, but is still too costly to cover the large number of roads. Given this shortcoming, other AADT estimation methods have been widely researched.

1.2. Problem Statement

As aforementioned, one application of AADT is to calculate crash rates. As part of the new Highway Safety Improvement Program (HSIP) of the Federal Highway Administration (FHWA), states are required to submit an annual report describing no less than 5% of their highway locations on all public roads which exhibit the most severe safety needs. To submit this 5% report annually, the Florida Department of Transportation (FDOT) requires that AADTs be available for all roads, including both state and local roads. However, FDOT currently can estimate AADTs using both short- and long-term traffic counts for only state roads. Research is needed to identify and develop methods to estimate AADTs for local roads in Florida.

In addition to the 5% report, AADT is also required by a new safety analysis system developed by FHWA known as *SafetyAnalyst* (2011). The system aims at providing state and local highway agencies with a comprehensive set of tools to enhance their programming of site-specific highway safety improvements. FDOT plans to take

full advantage of the new capabilities of *SafetyAnalyst* to enhance the safety improvement programs for not only the state roads, but also the local roads that are critical to the overall performance of the state's roadway system.

In 2007, FDOT contracted with the University of South Florida (USF) to develop regression models to estimate AADTs for local roads (Lu et al., 2007). A preliminary evaluation based on Miami-Dade and Broward County data showed that the corresponding errors of the USF method exceeded 100% and 200%, respectively.

Other methods of AADT estimation have also been researched by other researchers. They include image processing, travel demand modeling, and machine learning algorithms, such as Artificial Neural Network (ANN), K-Nearest neighbor, and support vector regression machines, and so on. However, the existing methods have their limitations. For example, the image-based method attempts to estimate AADT based on the traffic volume data extracted from satellite images, aerial photos, or LiDAR (Light Detection and Ranging) data, but it is difficult to retrieve and estimate volume for local roads accurately, because the traffic on local roads is usually sparse and infrequent compared to major roads. Machine learning methods may appear sophisticated, but they usually try to improve the traditional factor approach and still need short-term traffic counts collected with portable count sites, which is unpractical for local roads. In addition, none of these methods can provide satisfying estimation results for local roads.

FDOT recently contracted with URS Corporation to improve the AADT estimation for local roads. The URS method divides the street network in a Traffic Analysis Zone (TAZ) into multiple tiers according to the road levels, and accumulates the trips of each road estimated from parcel data and employee data by trickling down the

created tier structure to estimate the final AADTs for the local roads within a TAZ. The novelty of the URS method is that it generates the trips at the parcel-level, and then assigns the trips by simulating the flow of the river system. While the URS method improves upon the performance of the USF method, significant errors still remain in the estimation. A preliminary evaluation based on Miami-Dade and Broward County data showed that this method still produced estimation errors of 78% and 71%, respectively.

1.3. Research Objective

The main objective of this research is to attempt to develop an improved method of estimating AADT for local roads by applying travel demand modeling techniques, including trip generation, trip distribution, and traffic assignment at the parcel level.

1.4. Organization

This dissertation is comprised of six chapters. Chapter 1 introduces the research background, describes the problem to be solved, and sets the research objective to be achieved.

Chapter 2 provides a comprehensive review of the existing AADT estimation methods. The main purpose of this review is to research the state-of-the-art AADT estimation methods and determine if the existing AADT estimation methods are suitable for local roads.

Chapter 3 describes the methodology of the proposed parcel-level travel demand analysis model to estimate AADT for local roads. The traditional zone-level four-step travel demand forecasting model is first introduced, and the four model steps of the proposed parcel-level travel demand model, including network modeling, parcel-level

trip generation, parcel-level trip distribution, and parcel-level trip assignment, are then described in detail.

Chapter 4 describes the implementation of the parcel-level travel demand analysis model. The application of an ArcGIS tool called ModelBuilder to perform data preprocessing and post-processing for the model is introduced, and the development of each model step with Cube is described in detail.

Chapter 5 evaluates the performance of the parcel-level travel demand analysis model for local road AADT estimation. The AADTs estimated from traffic count data for local roads are used as the ground truth data. Results from two existing methods are compared with those from the proposed method.

Finally, Chapter 6 draws conclusions, summarizes the main contribution, and provides recommendations for future research.

CHAPTER 2

LITERATURE REVIEW

2.1. Introduction

Different approaches for AADT estimation can be found in the literature. They include:

- Traditional factor approach
- Regression modeling
- Travel demanding modeling
- Image processing
- Machine learning
- URS method

In this chapter, each of these approaches and the related literature in AADT estimation are reviewed in detail. Their advantages and limitations are also summarized.

2.2. Traditional Factor Approach

To estimate AADT on road segments with short-term counts, the traditional factor approach uses adjustment factors, which are calibrated from continuous Automatic Traffic Recorder (ATR) data, to convert the short-duration volume data collected (usually over a period of 48 hours) from the short-term counts. The effectiveness of this approach is based on the fact that it accounts for variations in traffic over different time scales such as time of day, day of week, and season (month of the year). It has been

widely applied throughout the U.S., and is recommended by the guidelines of AASHTO (1992) and the Traffic Monitoring Guide (TMG) of FHWA (2001).

The procedure for using the traditional factor approach to estimate AADT can be divided into two steps. The first step is to calculate the adjustment factors using the continuous traffic data recorded on the ATR sites. The second step is to apply the adjustment factors calibrated to estimate AADT values for road segments with short-term counts. The commonly used adjustment factors include axle correction factors, seasonal adjustment factors, and annual growth factors. To estimate AADT accurately, the appropriate calculations of these factors are critical.

To obtain more accurate adjustment factors, factor groups can be created by grouping the short-term sites and associated ATR sites. In this way, the average adjustment factors for each group can be determined. Factor groups are usually divided according to the functional classification, geographical location, and the judgment of analysts. A report prepared by Cambridge Systematics, Inc. (1994) recommended that the number of ATR sites in each group should be between five and eight. Roess et al. (2004) pointed out that groups for daily factors and groups for seasonal factors do not have to be the same, although it is convenient if they match. A detailed discussion about the methodologies to create factor groups can be found in TMG (FHWA, 2001).

The axle correction factors are used to convert the number of axles to the number of vehicles. This correction is necessary only when the short-term counts measure axle impulses with a single road tube. To calculate axle correction factors, the data from the vehicle classification counters for the same days as the short-term traffic count are usually used. At each permanent counter site of a factor group, vehicle classification

counters can detect the number of the vehicles in each classification. The total number of axles for this site can be calculated by summing up the product of the number of axles and number of vehicles for each classification. Dividing this figure by its total number of vehicles will get the average number of axles per vehicle for the site, which is summed up for all sites in the factor group and divided by the number of counters. The result is the group mean axles per vehicle, and its inverse is the axle correction factor for the group. The calculations can be performed using the following formula:

$$AF_g = \left[\frac{\sum_g \frac{\sum_c (A_c \times V_c)}{\sum_c V_c}}{N_g} \right]^{-1} \quad (2-1)$$

where

AF_g = the axle correction factor for factor group g ,

A_c = the number of axles for vehicle class c at a permanent count site,

V_c = the number of vehicles for vehicle class c at a permanent count site, and

N_g = the number of permanent sites in factor group g .

The seasonal adjustment factors are used for the day-of-week and monthly adjustments. An example to show how the seasonal factors are calculated is given as follows:

$$SF_{ijk} = \frac{AADT_k}{MADT_{ijk}} \quad (2-2)$$

where

SF_{ijk} = the seasonal factor for the day-of-week j in month i at ATR site k ,

$AADT_k$ = the AADT of ATR site k , and

$MADT_{ijk}$ = the monthly average day of the week traffic for month i and day-of-week j at ATR site k .

Two basic steps are involved in computing the seasonal adjustment factors: computing the numerator, which is AADT, and the denominator, which depends on the procedure used.

The numerator AADT can be calculated with the continuous traffic data recorded by the ATR sites. There are two basic methods to calculate AADT. One is the simple average daily traffic of all days in a year, and the other is called the average of averages method, which was presented by AASHTO (1992). This method first calculates the seven values of monthly average day-of-week (MADW) traffic for each month. The results in 84 MADW values are then grouped by day-of-week and averaged across the twelve months to yield seven values of annual average days of the week (AADW) for the year. The last step is to calculate the arithmetic mean of the seven AADW values, which can be used as the estimation of AADT. Both Cambridge Systematics (1994) and TMG (FHWA, 2001) recommended this AASHTO method because it can provide a more accurate estimation than the simple average method for such cases as when some data are missing from a specified year at a given site.

The denominator of calculating seasonal adjustment factors depends on the temporal grouping procedures used. These procedures can be based on day-of-week, month, combined weekdays, or combination of day-of-week and month, etc. Cambridge Systematics, Inc. (1994) compared seven of these procedures and concluded

that a number of different factoring techniques can result in reasonably similar levels of AADT estimating accuracy as long as the procedure accounts for all types of variation present in the data. TMG (FHWA, 2001) recommended a procedure named “combined month and day-of-week factors,” which is also called “eight-four factors,” if all seven days of the week (i.e., including Saturday and Sunday) are involved for each month.

The annual growth factors are needed when the historical traffic data are used to estimate AADT, since agencies rarely conduct traffic counts every year. The factors are usually the ratio of the AADT estimates of the current year to the preceding year. The sites from which these AADT estimates can be obtained are either ATR sites or short-term sites. While the ATR sites clearly provide better estimates of AADT, short-term sites provide a larger sample of sites, which means that more region-specific growth factors can be developed. Furthermore, the errors caused by short-term sites tend to be self-correcting over time (Cambridge Systematics, Inc., 1994).

After the necessary adjustment factors are calculated for a factor group, they can be used to estimate the AADT values for the road segments with short-term sites in the same group by simply multiplying short-term counts by the factors. In general, it can be represented with the following formula:

$$AADT_{gi} = ADT_{gi} \times AF_i \times SF_g \times GF_g \quad (2-3)$$

where

$AADT_{gi}$ = the annual average daily traffic at location i of factor group g ,

ADT_{hi} = the average daily (vehicle/axle) traffic at location i of factor group g ,

AF_i = the applicable axle correction factor for location i (if needed),

- SF_g = the applicable seasonal adjustment factor for group g , and
 GF_g = the applicable annual growth factor for group g (if needed).

The traditional factor approach to estimating AADT has been applied throughout the U.S. Although AASHTO (1994) and TMG (FHWA, 2001) have provided guidelines for this approach, different states have adopted slightly different procedures according to their individual circumstances. However, the basic principles of the approach are the same as those presented herein.

2.3. Regression Modeling

Regression analysis is a popular statistical tool to model and analyze the relationship between a dependent variable and one or more independent variables. Cook and Weisberg (1999) define regression analysis as a means to understand “as far as possible with the available data how the conditional distribution of the response y varies across subpopulations determined by the possible values of the predictor or predictors.” Hence, regression analysis is widely used for the purposes of description, prediction, and inference. More specifically, it is used to describe the distribution of a variable under a number of different conditions, predict the distribution of a variable in the future, and make inferences from a sample to a population. A number of techniques for carrying out regression analysis have been developed. Familiar methods such as linear regression and ordinary least squares regression are parametric, in that the regression function is defined in terms of a finite number of unknown parameters that are estimated from the data. Conversely, nonparametric regression refers to techniques that allow the regression function to lie in a specified set of functions, which may

be infinite-dimensional. Berk (2004) provides more detailed descriptions regarding regression analysis.

Regression analysis has been applied in several studies to estimate AADTs. At the state level, Deacon et al. (1987) produced a two-step modeling process to forecast highway volumes on the state highway systems in Kentucky.

Shon (1989) produced multiple regression models to estimate AADT according to the functional classification of the highways in Alabama. Different socio-economic characteristics were used as predictors for different functional classifications. State vehicle registrations and gasoline prices were used as predictors for principal arterials and interstate highways, year and county vehicle registrations were used for minor arterials, and year and gasoline prices were used for major collector roadways.

Cheng (1992) developed a regression model to estimate AADT on highway systems in Minnesota. Initially, independent variables were chosen from the road-log (RLG) database to be used as potential predictors. These included Route System (state roads or local roads), City Population, County Population, Location (urban or rural), Functional Classification (six functional classes for rural and eight for urban roads, respectively), Intersection Category, Special Road Section, Federal-aid System (if the road section receives federal aid), Access Control (uncontrolled, partially controlled, or fully controlled), Number of Through Lanes (in both directions), Type of Truck-route (eight truck-route classifications), Road Width (in feet, including sidewalks), and Surface Type (twenty-four categories). After analyzing each variable, some were dropped because they were either not useful or added significant complexity to the model. Ultimately, the number of the predictors was reduced to four: Route System, County

Population, Number of Through Lanes, and Location. It was found that Number of Through Lanes and AADT have a curvilinear relationship. The formula of the regression function is given as follows:

$$AADT = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_1 X_2 + \beta_4 X_3 + \beta_5 X_4 + \beta_6 X_2^2 \quad (2-4)$$

where

- X_1 = county population size,
- X_2 = total number of through lanes in both directions,
- X_3 = route system (state/non-state code), and
- X_4 = location (rural/urban code).

Mohamad et al. (1998) conducted a study to develop a linear regression model to estimate AADT on roadways in Indiana. Nine independent variables were considered initially: County Population, County Household, County Vehicle Registration, County Employment, County Per Capita Income, County Mileage, Location, Presence of Interstate Highway, and Accessibility (to the freeway for each road). After using the stepwise regression method to determine the independent variables which should be included in the model, four of them were chosen: Location, Accessibility, County Population, and County Mileage. The formula for the final AADT prediction model is given as follows:

$$\text{Log}(AADT) = 4.82 + 0.82 X_1 + 0.84 X_2 + 0.24 \text{Log}(X_5) - 0.46 \text{Log}(X_4) \quad (2-5)$$

where

- X_1 = location (1 = urban; 0 = rural),
- X_2 = accessibility (1 = easy access or close to the state highway; 0 = otherwise),

X_3 = county population, and

X_4 = total arterial mileage of county.

Xia et al. (1999) developed a regression model to estimate AADT for local roads in Broward County of Florida. The predictors used include number of lanes, area type, auto ownership, function classification, presence of non-state roads nearby, and service employment. The adjusted R^2 value was 0.5961, and prediction errors ranged from 1.31% to 57%. This model was later modified by Shen et al. (1999) by removing the service employment variable. The adjusted R^2 value was improved to 0.6069, with prediction errors ranging between 0.57% and 61.99%. Continuous efforts were made by Zhao and Chung (1999) based on the previous study. In this study, a larger data set was used, the old state roadway function classification system was replaced with the new federal function classification system, and a more extensive analysis of land use and accessibility variables was performed. Four models using different variables were developed, compared, and discussed. The best model used five predictors: Number of Lanes, Function Classification, Accessibility to Regional Employment, Direct Access (from a count station to expressway access points), and Employment in a Variable-sized Buffer surrounding a Count Station. This model has an adjustment R^2 value of 0.8180, and its Mean Squared Error (MSE) was 50.00.

The most relevant study regarding this topic was conducted by Lu et al. (2007). In the study, they developed a procedure to estimate AADT on all roads in Florida. The road segments were divided into three different types based on the number of traffic counts available to each street. The Type I streets include all freeways and major state highways where each road has at least one traffic count in each county. Minor state and

county highways and local streets consist of the Type II streets. The Type III streets include vehicle trails, freeway ramps, cul-de-sac, traffic circles, serve drivers, driveways, roads in parking area, and alleys. The linear regression models were developed to estimate the AADT values on Type II roads, which account for about 80% to 85% of the total streets. They also divided the counties in Florida into three groups based on the population in each county: rural area group (counties with population less than 100,000), small-medium urban area group (counties with population between 100,000 and 400,000), and large metropolitan area group (counties with population greater than 400,000). To estimate the AADT values on Type II streets, two distinct regression models, the state/county highway model and local street model, were created and applied to each county group, for a total of six complete regression models. Stepwise regression method was then used to select the variables for each model. The adjusted R^2 values and the Mean Absolute Percentage Error (MAPE) values were subsequently calculated. The final equations of the six prediction models with the adjusted R^2 and MAPE values are given as follows:

- Large Metropolitan Area, State/County Highway Model

$$\begin{aligned}
 AADT = & -848.8 + 13.541 \times VEHICLE + 1273.347 \times DIVIDED + \\
 & 2983.442 \times COMMERCIAL + 6259.677 \times LOCATION - \\
 & 8.845 \times LABORFORCE - 2839.185 \times AGRCULTRUE + \\
 & 421.252 \times NUMBEROFLANE + 1311.231 \times INSTITUTIONAL + \\
 & 129.069 \times INCOME + 796.601 \times 0.5MILE - \\
 & 782.648 \times RESIDENTIAL - 587.47 \times SEMIPUBLIC
 \end{aligned}$$

$$R_{adj}^2 = 0.186$$

$$MAPE = 46.81\%$$

- Large Metropolitan Area, Local Street Model

$$\begin{aligned}
AADT = & -2738.443 + 3.806 \times MUNICIPALITIES + 1349.659 \times DIVIDED - \\
& 452.459 \times RESIDENTIAL - 567.182 \times 1.5MILE + \\
& 2745.195 \times LOCATION + 259.492 \times NUMBEROFLANE + \\
& 1040.226 \times SEMIPUBLIC + 769.194 \times COMMERCIAL - \\
& 19.545 \times LABORFORCE + 17.369 \times POPULATION - \\
& 4.345 \times VEHICLE
\end{aligned}$$

$$R_{adj}^2 = 0.242$$

$$MAPE = 159.49\%$$

- Small-medium Urban Area, State/County Highway Model

$$\begin{aligned}
AADT = & 770.374 + 5566.145 \times LOCATION + 122.079 \times LABORFORCE + \\
& 2760.767 \times COMMERCIAL + 960.82 \times NUMBEROFLANE + \\
& 27.673 \times VEHICLE - 70.869 \times POPULATION + \\
& 0.994 \times SALES - 13.311 \times MUNICIPALITIES + \\
& 952.963 \times 1.5MILE - 431.282 \times RESIDENTIAL + \\
& 765.103 \times SEMIPUBLIC - 0.43 \times MILEAGE + 1072.666 \times INDUSTRIAL
\end{aligned}$$

$$R_{adj}^2 = 0.259$$

$$MAPE = 65.01\%$$

- Small-medium Urban Area, Local Street Model

$$\begin{aligned}
AADT = & 1533.94 + 2482.69 \times DIVIDED - 679.405 \times RESIDENTIAL + \\
& 2107.874 \times 1.5MILE + 2707.119 \times LOCATION + 18.468 \times VEHICLE - \\
& 14.468 \times POPULATION + 0.9437 \times MUNICIPALITIES + \\
& 3320.091 \times INDUSTRIAL + 1491.556 \times COMMERCIAL + 1464.231 \times \\
& INSTITUTIONAL + 2011.814 \times RECREATION
\end{aligned}$$

$$R_{adj}^2 = 0.166$$

$$MAPE = 65.35\%$$

- Rural Area, State/County Highway Model

$$\begin{aligned}
AADT = & 3015.747 + 3878.551 \times LOCATION + 17.722 \times VEHICLE + \\
& 57.072 \times MUNICIPALITIES - 1656.733 \times AGRICULTURE + \\
& 22.293 \times LABORFORCE - 1.931 \times SALES - \\
& 3312.919 \times RECREATION - 2324.493 \times INDUSTRIAL + \\
& 33.239 \times POPULATION - 748.708 \times RESIDENTIAL
\end{aligned}$$

$$R_{adj}^2 = 0.378$$

$$MAPE = 31.99\%$$

- Rural Area, Local Street Model

$$\begin{aligned}
AADT = & 1225.505 + 62.168 \times POPULATION + 1458.501 \times LOCATION - \\
& 1445.085 \times AGRICULTURE - 1017.873 \times RESIDENTIAL
\end{aligned}$$

$$R_{adj}^2 = 0.418 \quad MAPE = 46.79\%$$

The definitions of the independent variables used in the equations above are listed as follows:

- Socio-economic Variables
 - POPULATION: population in thousands;
 - MILEAGE: total mileage of highways in a county;
 - VEHICLE: total number of registered vehicles in thousands;
 - INCOME: the per capita income in thousands;
 - SALES: yearly retail sales in million;
 - MUNICIPALITIES: population within incorporated area in million; and
 - LABORFORCE: labor force within one county in thousands.
- Road Characteristics Variables
 - DIVIDED: if the roadway is divided, is 1; otherwise, 0;
 - NUMBEROFLANE: number of lanes in both directions;

- LOCATION: if the location is urban, is 1; otherwise, 0;
- 0.5MILE: if a road is within 0.5 mile from freeway, is 1; otherwise, 0;
- 1.0MILE: if a road is within 1 mile from freeway, is 1; otherwise, 0;
- 1.5MILE: if a road is within 1.5 mile from freeway, is 1; otherwise, 0;
- SEMIPUBLIC: if land use type is Public-Semipublic, is 1; otherwise, 0;
- COMMERCIAL: if land use type is Commercial, is 1; otherwise, 0;
- AGRICULTURE: if land use type is Agriculture, is 1; otherwise, 0;
- INSTITUTIONAL: if land use type is Institutional, is 1; otherwise, 0;
- RESIDENTIAL: if land use type is Residential, is 1; otherwise, 0;
- RECREATION: if land use type is Recreation, is 1; otherwise, 0; and
- INDUSTRIAL: if land use type is Industrial, is 1; otherwise, 0.

While all the applications of regression analysis given above used the traditional Ordinary Least Squares Regression (OLS-regression), Park (2004) applied Geographically Weighted Regression (GWR) to estimate AADT for highways in Broward County of Florida. Differing from OLS-regression, in which the model estimates the global parameters for the entire study area, GWR considers the influence of correlations among the variables over space, and estimates different parameters for different locations by weighting the observations inversely to their distance from the location where the AADT is estimated. Six independent variables were selected from 67 variables to develop the model: Number of Lanes, Speed, Regional Accessibility, Direct Access to Expressways, Density of Roadway Length, and Density of Seasonal Household. A comparison with the OLS-regression model was also done, and it was concluded that the GWR approach exhibited better performance.

2.4. Travel Demand Modeling

Travel demand modeling utilizes mathematical models to simulate “real world” transportation system and human travel behaviors. Traditionally, the “four-step process” has been used for travel demand analysis and, as its name implies, is composed of four steps: trip generation, trip distribution, mode choice, and trip assignment. The first step, trip generation, calculates the number of trips generated in each Traffic Analysis Zone (TAZ), which is the unit of geography commonly used in travel demand modeling. In the second step, trip distribution, the distribution of trips among the origin and destination zones is determined. The third step, Mode Choice, splits the trips between the origin and destination zones according to different modes of travel. Finally, trip assignment allocates the trips to routes by each travel mode.

Little research has been done in terms of applying the travel demand modeling approach to the estimation of AADT. Zhong and Hanson (2009) utilized traffic demand models to estimate AADT on low-class roads for two regions in the province of New Brunswick, Canada. Modifying the traditional four-step process, they omitted the third step of mode choice from their procedure. The Quick Response Method (QRM) (Sossiau et al., 1978) was also adopted for the trip generation step, and the traditional gravity model used for the trip distribution step. The final step, trip assignment, was implemented by using the STOCH method, which was first proposed by Sheffi (1985). The empirical results show that the average estimation errors can be limited to less than 40%, which is comparable to the results of other AADT estimating approaches.

While their research showed that this method has the potential to improve AADT estimation, due to the resolution limitations of the available census data, their method

was applied at the dissemination areas (DAs) level. DA is the smallest census unit in Canada. A DA is a small, relatively stable geographic unit composed of one or more neighboring dissemination blocks, with a population of 400 to 700 persons. Further research is needed to estimate the performance of this method as applied to smaller areas such as parcels level researched in this dissertation.

2.5. Image Processing

Estimating AADT with image-based data has been possible due to the collection of high-resolution satellite images, aerial photos, and LiDAR (Light Detection and Ranging) data by transportation agencies for planning and analysis purposes. McCord et al. (1995a) and McCord et al. (1995b) analyzed the feasibility of this approach and proved that 1-m resolution is necessary to count and classify cars and trucks with accuracy greater than 90%.

McCord et al. (2003) proposed the methodology of image-based AADT estimation and also compared this with the traditional ground-based factor method. To produce the AADT estimation on a road segment, the vehicle density is first obtained from the image and converted to a short-duration volume. The short-duration volume is then expanded to an hourly volume, daily volume, and finally, AADT, by multiplying some expansion factors. A comparison with the traditional ground-based factor approach indicated a small difference between the results of the two methods, which might imply that image-based estimation can augment traditional ground-based estimation and, therefore, that the combination of the two could lead to more accurate estimation. This combination of image-based and ground-based estimations was implemented in Jiang et

al. (2006). For ground-based data, they estimated AADTs for the current year by using seasonal factors and growth factors on coverage counts data in earlier years. For image data, they applied the method proposed in McCord et al. (2003) to estimate AADTs for road segments with a single, more recent image. The two AADT estimation results were then integrated by using a linear weighted combination according to their variances. An empirical study was conducted to simulate weighted estimation of AADTs on 122 Florida highway segments between 1994 and 2003, with the results showing that the accuracy of AADT estimation was markedly improved.

Jiang et al. (2007) verified the numerical results of Jiang et al. (2006) with a study of 12 aerial photos taken by Ohio DOT in 2005 for Ohio road segments equipped with ATRs. They compared both the combined estimation and traditional coverage count estimate to the “true” AADT determined by the ATRs data. The results showed that the combined estimation produced a lower average relative error, a higher proportion of estimates with relative error less than 0.10, and better estimates overall more than 50% of the time.

Another approach using image-based data to estimate AADT was researched by Jiang (2005). In this study, a Bayesian approach is used to combine the traditional ground-based data and the traffic data extracted from the images. A three-stage model was then developed to simulate the prior distribution of AADT and the probability distribution of short-term traffic counts conditional on AADT. This numerical investigation shows the benefits of image-based data in terms of improving the accuracy of AADT estimation.

2.6. Machine Learning

Mitchell (1997) defines machine learning as a computer program “to learn from experience E with respect to some class of tasks T and performance measure P, if its performance at tasks in T, as measured by P, improves with experience E.” The learning system utilizes certain learning algorithms to derive a description of a given concept based on a set of concept examples and background knowledge (Michalski et al., 1998). A number of machine learning algorithms have been used to perform the task of AADT estimation or provide helpful assistance to certain aspects of the task. This section reviews three typical approaches: the artificial neural network, k-nearest neighbor, and support vector regression machine.

2.6.1. Artificial Neural Network Approach

An Artificial Neural Network (ANN) is a computational model that is inspired by the structural/functional aspects of biological neural networks. It is an emulation of biological neural networks, and consists of simple artificial neurons connected by directed weighted connections. It may be thought of as simplified models of the networks of neurons that occur naturally in the animal brain (Gurney, 2009). The structure of an ANN is changed based on external or internal information that goes through the network during the training phase. Modern ANNs are non-linear statistical data modeling tools, and a well-trained ANN is usually used to model complex relationships between the inputs and the outputs of the network or to find patterns in the data.

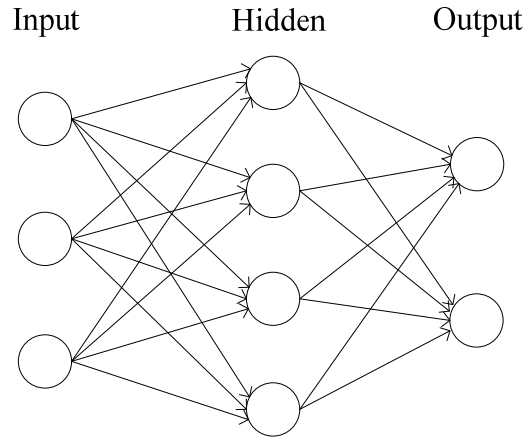


Figure 2-1 Example of a Simple Feedforward Neural Network

Figure 2-1 shows an example of a simple feedforward neural network from Wikibooks (2011). In this common type of ANN, there are three layers of units: the input layer, the hidden layer, and the output layer. The input layer is connected to the hidden layer directly, and the hidden layer is connected to the output layer directly. There is a weight value assigned to a connection between each pair of connected units, and the weight value can be adjusted during the learning phase. The activity of the input units represents the raw information that is fed into the neural network. The behavior of each hidden unit is determined by the activities of the input units and the weight values of the connections between the input and the hidden units. The activity of the output units depends on the activity of the hidden units and the weight values of the connections between the hidden units and the output units. “Feedforward” means the signals are allowed to travel one way only: from the input layer to the output layer. Feedforward network is simple and straight forward, since there are no loops in the network. On the contrary, more complex feedback networks can have signals travelling

in both directions, and they are more powerful and can be extremely complicated, because feedbacks (loops) are allowed in the network.

ANN is a type of non-linear processing system that is ideally suited for a wide range of tasks, especially tasks in which there is no existing algorithm for task completion (Wikibooks, 2011). When the system is set running, the activation levels of the input units are affixed to the desired values. After this, the activation is propagated, at each time step, along the directed weighted connections to other units. The activations of non-input neurons are computed using each neuron's activation function. The system might either settle into a stable state after a number of time steps, or in the case of a feedforward network, the activation might flow through to output units.

ANN can be trained to solve certain problems using a teaching method and sample data. In this way, identically constructed ANN can be used to perform various tasks depending on the training received. With proper training, ANN is capable of generalization, or the ability to recognize similarities among different input patterns, especially patterns that have been corrupted by noise. Detailed information about the theoretical foundations of ANN can be found in Anthony and Bartlett (1999).

ANN has been extensively applied to transportation research since the 1990s. Dougherty (1995) summarized the findings of research papers regarding the application of ANN to transportation. The subject areas with the most ANN application include driver behavior/autonomous vehicles, parameter estimation, pavement maintenance, vehicle detection/classification, traffic pattern analysis, traffic forecasting, etc. More applications of ANN in transportation can also be found in Himanen et al. (1998).

As an important aspect of transportation research, an ANN approach to AADT estimation has also been explored. Sharma et al. (1999) compared the ANN approach to the traditional factor approach with 48-hour short-term counts data for estimating AADT. A multilayered, feed-forward, and back-propagation neural network with supervised learning was designed to achieve this purpose. It was found that for a single 48-hour count, if ATR sites are grouped appropriately and the coverage counts are assigned to the ATR groups correctly, then the estimation errors of the traditional factor approach can be lower than that of the ANN approach. However, this investigation also indicated that, there was unfortunately little guidance on how to achieve a high enough ATR site grouping and accuracy of sample counts assignment to obtain reliable AADT estimates. It was also found that the accuracy of the ANN approach is comparable to the traditional factor approach when it is applied to two or more 48-hour counts taken during different months. Since the advantage of the ANN approach is that the groups of ATR sites and assignment of sample short-term counts are not required, the research recommends the ANN approach as a better choice.

While Sharma et al. (1999) focused on interstate and other high-volume roads, Sharma et al. (2000, 20001) applied the ANN approach to low-volume rural roads. In addition to some findings that verified those of Sharma et al. (1999), it also found that the 48-hour count duration is likely to produce much better estimation than the 24-hour count duration. Furthermore, 72-hour count duration may not necessarily offer an advantage.

Lam and Xu (2000) implemented a multi-layer feed-forward neural network with back-propagation algorithm to estimate AADT and determine the most appropriate length of counts. The case study was carried out by analyzing data on 13 trunk roads and

primary roads in Hong Kong, and the results showed that the neural network approach performed consistently better than the regression analysis approach in estimating AADT.

2.6.2. K-Nearest Neighbor Approach

The K-nearest neighbor algorithm (K-NN) is a data mining method for classification, although it can also be used for estimation and prediction. K-NN is among the simplest of all machine learning algorithms and is a type of instance-based learning in which the training data set is stored, thereby allowing a new unclassified record to be classified by comparing it to the most similar records in the training set (Larose, 2005). The similarity is measured by the distance between the records, with the new record assigned to the class most common among its K-nearest neighbors.

There is no obvious best solution to choose the value of K. As mentioned by Larose (2005), a K with a value that is too small may cause overfitting, while a K with a value that is too large tends to overlook locally interesting behavior. Thus, it is typically a small (but not too small) positive integer. If $K = 1$, then the object is simply assigned to the class of its nearest neighbor.

Since the K-NN algorithm is used mostly for classification, it can be utilized to assign short-term count sites to different ATR factor groups. Li and Fricker (2008) proposed a K-NN algorithm combined with GIS technology to carry out roadway classification. The attributes of a roadway count that are helpful for the classification were chosen, which include geographic spatial location, roadway link characteristics (Functional Class, Number of Lanes, and Posted Speed), and land use characteristics in the area surrounding the ATR. Various values of K from 5 to 9 were then tried and

compared, using data from 56 ATRs on the Indiana roadway network for 2004. They also compared the K-NN method with the traditional twenty-four and eighty-four factor approaches, which use each functional class as a factor group. The results showed that K-NN can produce better AADT estimates.

2.6.3. Support Vector Regression Machines Approach

Support vector machines (SVM) are a set of supervised learning methods. A support vector machine constructs a hyperplane or set of hyperplanes in a high or infinite dimensional space, which can be used for classification, regression, or other tasks. Support vector machines represent an extension to nonlinear models of the generalized portrait algorithm developed by Vladimir Vapnik. The SVM algorithm is based on the statistical learning theory and the Vapnik-Chervonenkis (VC) theory introduced by Vladimir Vapnik and Alexey Chervonenkis. A detailed description of the SVM algorithm is given by Vapnik (1995).

Based on SVM theory, Support Vector Regression Machines (SVR) were proposed by Drucker et al. (1996). While SVR uses the same principles as the SVM for classification, it also sets a margin of tolerance, ϵ , in approximation to SVM to predict the real number output, which has infinite possibilities and is very difficult to predict. SVR is the most common application form of SVMs. An overview of its basic ideas has been given in Smola and Schölkopf (1998).

SVR has been widely applied due to its remarkable characteristics. Castro-Neto et al. (2009) evaluated the performance of a modified version of SVR named SVR-DP (SVR with Data-dependent Parameters). This model was used in forecasting AADT one

year into the future based on the historical AADT values, which differs from the common type of current-year AADT estimation based on external predictor variables. The technique was first introduced by Cherkassky and Ma (2004). By computing the SVR parameters based on the distribution of the incoming training data, it can alleviate the problem of excessive data requirements and the time-consuming computation of adequate SVR parameters, which are crucial to the quality of SVR models. Castro-Neto et al. (2009) used AADT values collected between 1985 and 2004 for both urban and rural roads in 25 counties in Tennessee. The SVR-DP approach was compared with two other popular methods, Holt Exponential Smoothing (Holt-ES) and Ordinary OLS-regression. The results show that SVR-DP outperformed both of these models, although the Holt-ES also presented good performance.

2.7. URS Method

FDOT contracted with URS Corporation to improve the AADT estimation. The URS method divides the street network in a Traffic Analysis Zone (TAZ) into $N + 1$ (from 0 to N) tiers according to the road levels. Tier 0 segments represent roads that have an official FDOT AADT or segments in the Turnpike State model. Tier 0 segments are the boundary segments of the TAZ zones developed for the Turnpike State Model. Tier 1-N segments are roads inside a TAZ zone, and each TAZ is analyzed separately as a unit. The segments with the same Roadway ID are called a route. The segments of a route that touches a tier 0 segment were assigned a tier value of 1. The segments of a route that touches a tier 1 route were assigned a tier value of 2. The process repeats until every route and segment within the TAZ is assigned a tier value.

The AADT of a tier 0 segment will be the official FDOT AADT, but if a segment did not receive an official FDOT AADT, the Turnpike State model volume is used as the AADT.

To calculate the AADT for the non-state road segments in a TAZ, the routes are buffered and intersected with the parcel polygons and employment points to get the sum of housing units and employees associated with each route. The total number of housing units and employees within the TAZ can be summed. The total number of trips within the TAZ can be provided by the Turnpike State Model. The total number of trips divided by the total number of housing units and employees will generate a trip factor. Using this trip factor multiplied by the number of housing units and employees for each route, each route within the TAZ is assigned a volume.

Starting from the highest tier routes, each route's volume is trickled down to the connected lower tier routes which are called the mother routes. If there are multiple mother routes, the volume is split evenly and accumulated to each of the mother routes. The AADT of a route is the trips for that route plus the accumulation of the trips from the higher tiered routes that are connected to the route.

2.8. Summary

In this chapter, a comprehensive literature review has been conducted to investigate the current techniques and methods for AADT estimation. The major findings of the literature review are summarized below.

For AADT estimations, the traditional factor approach uses the permanent count sites to calibrate the adjustment factors, the short-term count sites to collect the

short-duration volume data, and converts the short-duration volume to the estimated AADT with the adjustment factors. This method may be the most accurate AADT estimation method and has been widely applied for state roads. However, it is obvious that it is economically infeasible to maintain the permanent count sites on local roads and also infeasible to use the portable count sites to cover all the local roads.

The regression modeling method uses the statistical methodology and tools to analyze the relationship between AADT and socio-economic variables such as population and the road characteristic variables such as number of lanes. This method has been most widely researched, but the main problem with this method is that it cannot capture passer-by trips. In addition, it does not perform well when the relationship between the independent and the dependent variable is nonlinear.

Travel demand modeling technique has seldom been researched in terms of AADT estimation. Zhong and Hanson (2009) was the only researched found and reviewed. While their research showed that this method has the potential to improve AADT estimation for low-class roads, further research is needed to estimate the performance of this method as applied to smaller areas such as parcels level researched in this dissertation.

The image processing method uses image-based data including the high-resolution satellite images, aerial photos, and LiDAR (Light Detection and Ranging) data to obtain vehicle density and then converts it to a short-duration volume which can be expanded to AADT by multiplying by expansion factors. The limitation of this method is that it is difficult to retrieve and estimate volume for local roads accurately,

because the traffic on local roads is usually sparse and infrequent compared to major roads.

The machine learning methods such as ANN, K-nearest neighbor algorithm, and SVR have also been reviewed, but it was found that these methods usually try to improve the traditional factor approach but still need to deploy portable count sites to collect short-term traffic count data, which has been proven to be unpractical for local roads. In addition, none of these methods can provide satisfying estimation results for local roads.

Lastly, the method recently proposed by the URS Corporation for FDOT was also reviewed. The URS method divides the street network in a TAZ into multiple tiers according to the road levels, uses the parcels and employee data in the road segment buffers to estimate the initial trips, and assigns the trips to the created roadway tire structure by trickling down to the connected parent routes. The idea of this method is based on the similarity between the roadway system and the river system, and its process is trying to simulate that of the river system. Theoretically, this AADT estimation method should be suitable for local roads, because it uses the most detailed parcel and employee data, and collects trips from the lowest level roads. However, the performance of this method needs further evaluation, so it is selected as one of the testing methods to compare with the method proposed in this dissertation.

CHAPTER 3

METHODOLOGY

3.1. Introduction

From the literature review, it can be concluded that the existing AADT estimation methods have limitations on estimating AADT on local roads. To estimate AADT more accurately for local roads, a parcel-level travel demand analysis model based on the traditional four-step travel demand forecasting model is proposed and implemented in this research.

In this chapter, the traditional four-step travel demand forecasting model is briefly introduced, and the methodology of the parcel-level travel demand analysis model for AADT estimation on local roads is then described in detail. Each step involved in the model is then explained at length. The method to evaluate the estimation results is also discussed.

3.2. Traditional Travel Demand Forecasting Model

3.2.1. Introduction

The primary objective of the traditional travel demand forecasting model is to predict the effects of various projects, policies, and programs on the highway and transit facilities. The impacts are usually quantified by traffic volumes and transit ridership. The model involves a series of mathematical models that simulate human travel behaviors in response to a given system of highway and transit alternatives.

Traditionally, it is also referred to as four-step travel demand model, as it involves the following four steps: trip generation, trip distribution, mode split, and trip assignment.

Before using the travel demand forecasting model for an urban area or a region, planners must clearly define the exact boundaries of the study area, i.e., the cordon lines. The study area generally includes all of the developed land and the undeveloped land that may be developed in the next 20 to 30 years. The establishment of the cordon line usually take into account the political jurisdictions, census area boundaries, and natural boundaries.

For modeling analysis, the study area is divided into Traffic Analysis Zones (TAZs). A TAZ is the basic unit used to quantify the activities, travel, and transportation characteristics of a physical location in the study area. Its size may vary, depending on the density or nature of the development area. A TAZ can be as small as a single city block in an urban area, or it can be larger than several square miles in a rural area. Figure 3-1 shows the TAZs in Broward County of Florida.

A study area may have multiple networks such as highway network and transit network, comprising of links and nodes. The links have associated data attributes including travel times, average speeds, capacity, number of lanes, direction, etc. The node attributes may include coordinates, type of intersection, etc.

A centroid is a special type of node that represents the “center of activity” in a TAZ. Centroids are connected to the surrounding roadways by a special type of links called centroid connectors. Centroids and centroid connectors are used to load the trips generated within a TAZ onto the highway network. The creation of centroids and centroid connectors are based on the zone boundaries and the street network. An

example is given in Figure 3-2 to illustrate the process of creating centroid and centroid connectors based on the connections of the street network.

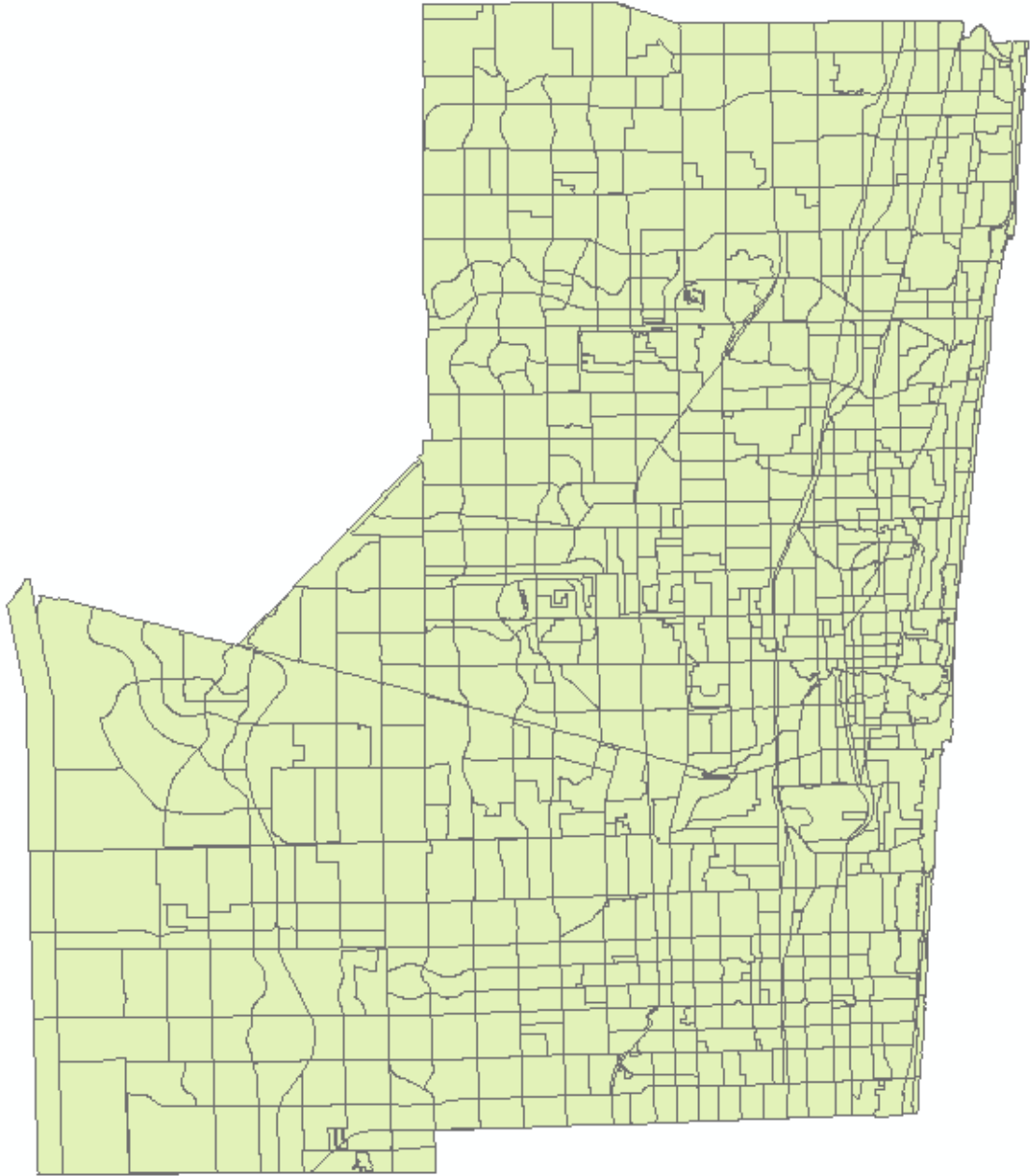


Figure 3-1 TAZs in Broward County, FL

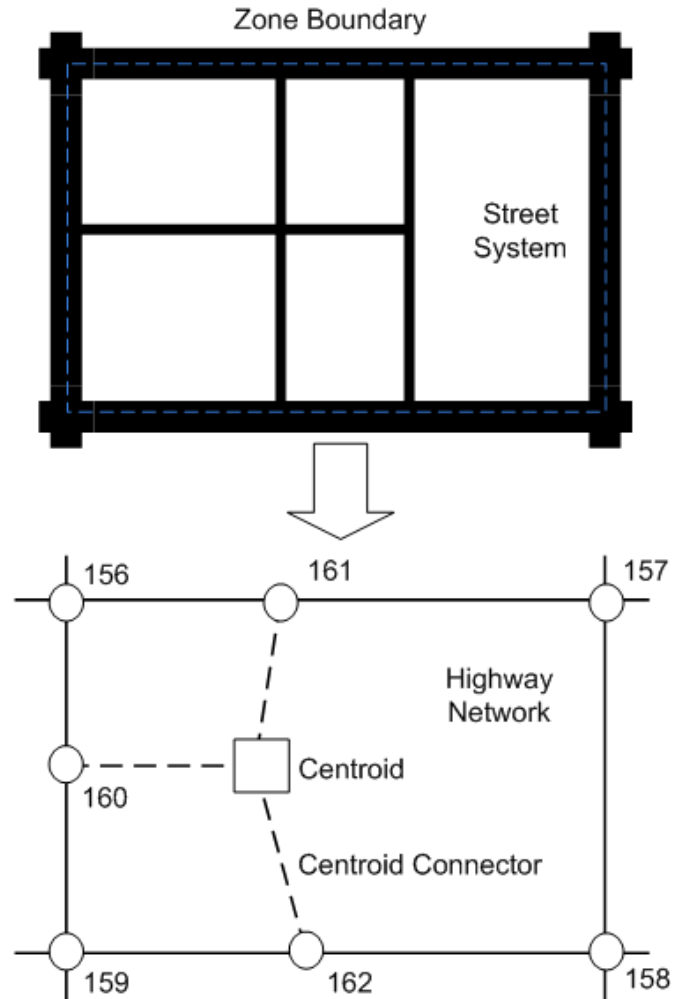


Figure 3-2 Creation of Centroid and Centroid Connectors

Once the transportation network with the centroids and centroid connectors is established, the four major model steps, trip generation, trip distribution, mode choice, and trip assignment, can be performed for the study area. Each of the steps is introduced separately in the following sections.

3.2.2. Trip Generation

The major objective of the trip generation step is to forecast the number of trips that each TAZ will produce or attract. The trips are categorized in different purposes

such as work, school, shopping, and social-recreational, etc. Trip purpose is a major factor affecting travel behaviors, so categorizing trips into different purposes can help build a more accurate travel demand forecasting model.

Oppenheim (1995) described trip generation in detail. A trip is a one-direction movement, which means when a person went from home to work in the morning and then returned home from work in the afternoon, a total of two trips were made. For modeling purposes, the trip origins and destinations are converted to trip productions and attractions. A trip production is defined as the home end of a home-based trip or the origin of a non-home-based trip, and a trip attraction is defined as the non-home end of a home-based trip, or the destination of a non-home-based trip.

Trip production is associated with households, so it is a function of household characteristics including house hold type, vehicle ownership, income, etc. Trip attraction is associated with commercial or industrial sites, so it is a function of the variables such as number of employees, total floor area, etc. Two common methods used to perform trip generation are multiple regression and cross-classification analysis.

The multiple regression method expresses trips as a function of one or more independent variables. Each variable is associated with a trip rate, which is estimated through a model calibration process using the trip survey data.

The cross-classification method, also known as category analysis method, stratifies trip rates based on household characteristics such as household size and vehicle ownership. Unlike the regression method which uses data aggregated to TAZ, the cross-classification method is a disaggregate method and uses input at the dwelling unit level. To estimate the trips generated by a TAZ, the number of households belonging to

each of the different strata is multiplied by the corresponding trip rate, and the total trips generated by a TAZ is obtained by summing the trips from each stratum.

3.2.3. Trips Distribution

After the trips generated by each TAZ are estimated, the trip distribution model can be used to distribute trips among the zones. The result is a set of trip interchanges for each pair of TAZs.

Oppenheim (1995) described trip distribution in detail. Trip distribution has traditionally been performed based on either the gravity model or the growth factor method. The gravity model was derived from Newton's law of gravity. It assumes the total number of trip interchanges between a pair of zones is directly proportional to the trip intensities of the two zones and inversely proportional to the separation between the two zones which is measured by travel impedance such as travel time. The model can be expressed using the following formula:

$$T_{ij} = \frac{A_j F_{ij} K_{ij}}{\sum_{m=1}^n (A_m F_{im} K_{im})} \times T_i \quad (3-1)$$

where,

T_{ij} = trips produced in zone i and attracted to zone j ;

T_i = total trip productions at zone i ;

A_j = total trip attractions at zone j ;

F_{ij} = separation between zones i and j , commonly known as friction factor;

K_{ij} = a socioeconomic adjustment factor between zones i and j ; and

n = number of zones.

The growth factor method predicts the future number of trip interchanges between two zones based on the base-year trip interchanges. This method is useful when information on travel impedance is not available or cannot be sufficiently estimated.

3.2.4. Mode Choice

After the trip interchanges between each pair of TAZs are estimated in the earlier step, the mode choice step, also known as mode split, is performed to determine what transportation mode each traveler will use. The step estimates the percentage of people that use private automobiles, carpools, public transit, etc.

Oppenheim (1995) described mode choice in detail. The mode choice step can also be performed after the trip generation step and before the trip distribution step. This is called the pre-distribution mode choice model. The common practice is to perform the mode choice step following the trip distribution, called the post-distribution mode choice model.

The most common form of the mode choice model is the logit model. It assumes that the probability of the traveler choosing a particular mode is based on the relative values of number of factors including the characteristics of the traveler, trip characteristics, and the characteristics of the transportation mode. The logit model is defined as follows:

$$P_i = \frac{e^{U_i}}{\sum_{j=1}^n e^{U_j}} \quad (3-2)$$

where,

P_i = probability of choosing mode i ,

U_i = utility function of mode i , and

n = number of zones.

In addition to the mode choice, three additional sub-steps are usually performed as part of the mode choice step. They include converting the person trips to vehicle trips, combining the vehicle trip tables of different trip purposes into a single trip table, and converting the trip table from non-directional production-attraction format to directional origin-destination format.

3.2.5. Trip Assignment

Once the transportation modes that the travelers will choose have been determined, the trip assignment step will be performed to predict the routes that they will use.

Oppenheim (1995) provided detailed information regarding trip assignment step. A simple trip assignment method is called all-or-nothing assignment. It assumes all the travelers will choose the route with the shortest free flow travel time for a trip. This method will become unreliable during congestion, because congestion increases travel time. Therefore, another method called all-or-nothing with capacity restraint, also known as the equilibrium assignment method, is commonly used for trip assignment. In this method, the travel times are recomputed on basis of the loaded network and the trips are reassigned based on the new travel times. It is necessary to implement an iterative procedure in order to apply this type of assignment. In each iterative procedure, the travel times are recomputed at the end of each assignment and used as the input to the next assignment iteration. The procedure continues until some sort of equilibrium is

established in the system. The basis in this procedure is the utilization of some form of capacity-restraint function to adjust travel times. A number of such functions have been researched. The one developed by the Bureau of Public Roads (BPR) is defined as follows.

$$T = T_0 \times \left[1 + 0.15 \times \left(\frac{v}{c} \right)^4 \right] \quad (3-3)$$

where

- T = new travel time,
- T_0 = free flow travel time,
- v = assigned volume, and
- c = practical capacity.

3.3. Parcel-level Travel Demand Analysis Model

3.3.1. Introduction

Similar to the traditional travel demand forecasting model, the proposed parcel-level travel demand analysis model also involves a series of mathematical models to simulate human travel behaviors. However, unlike the traditional travel demand forecasting model, which attempts to simulate the choices that the travelers may make during the entire trip from the origination to the destination, the parcel-level model attempts to simulate choices that travelers may make in response to the given local streets system to access the major roads which are just the initial parts of the entire trip.

As shown in Figure 3-3, the process of the proposed parcel-level travel demand analysis model is straightforward. Four steps named network modeling, parcel-level

trip generation, parcel-level trip distribution, and parcel-level assignment will be performed separately, in sequence.

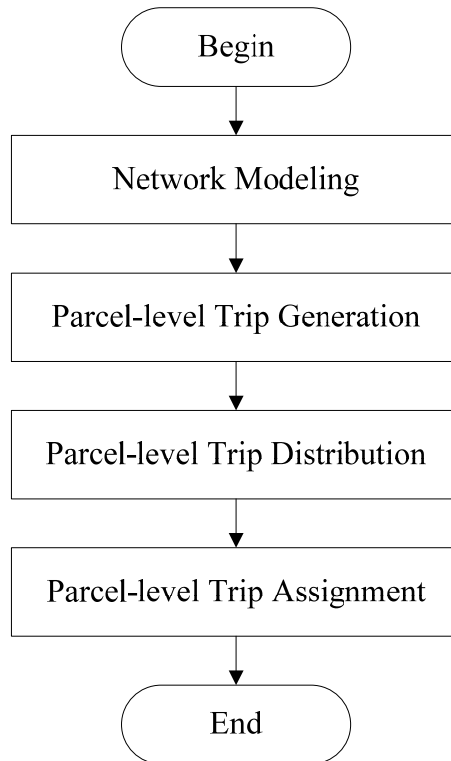


Figure 3-3 Flowchart of Parcel-level Travel Demand Analysis Model

The functionalities of each step involved in the parcel-level travel demand analysis model are listed as follows:

- Network Modeling defines the boundaries of the study area, prepare and preprocess the roadway network, parcel, and traffic counts data, and sets up the network representation of the roadway linked with parcels and traffic count sites.
- Parcel-level Trip Generation estimates the number of vehicle trips generated by each parcel in the study area. The estimation is calculated based on the land use type of each parcel and the respective ITE trip generation rate for that type.

- Parcel-level Trip Distribution determines where the trips generated by each parcel will go. It determines the number of trips between a parcel and a traffic count site based on traffic count data (or AADT estimated from the count data) and the shortest travel time between them.
- Parcel-level Trip Assignment predicts the routes the travelers will take to approach the traffic count sites, resulting in the estimated AADTs of local roads in the study area.

From the above discussion, it can be seen that the principle of the proposed parcel-level travel demand analysis model is similar to that of the traditional four-step zone-level travel demand analysis approach as both methods attempt to simulate human travel behaviors. However, there are also significant differences between them. While the traditional model performs the travel demand analysis on an abridged roadway network with only major roads, the proposed model simulates the trips on an unabridged roadway network including the local roads. Another major difference is that there is no mode choice step in the parcel-level model, since the parcel-level trip generation step will generate only vehicle trips and exclude the transit trips, which are usually represented by walk trips inside a zone traveling between public transit facilities such as bus stops located on major roads. A third difference is that while the traditional four-step zone-level travel demand model distributes trips among TAZs, parcel-level model will distribute the trips between the parcels and their nearby traffic count locations. The differences between these two approaches will be shown in greater detail in the sections below.

3.3.2. Network Modeling

In network modeling step, the data preparation process is comprised of the following three sub-steps:

- define the boundary of the study area;
- prepare and preprocess the required data including roadway network, parcels, and count sites; and
- link the parcels and the count sites to the unabridged roadway network.

As mentioned above, the boundary of the study area is commonly called the cordon line. When defining the cordon line, the same rules for the traditional zone-level travel demand analysis approach can be followed. To establish the cordon line, political jurisdictions, census area boundaries, and natural boundaries may be taken into account, and it is generally defined such that it intersects with as fewer roads as possible.

After the boundary of the study area is defined, the required data are prepared and preprocessed. The data include the unabridged roadway network data, the detailed parcel data, and the traffic count sites data. If necessary, some preliminary processing on the input data is performed in this step. For example, the traffic count site data can be divided into two groups based on the location of count site (if a count site is located on the major roads or the local roads) so that the major roads group will be used for AADT estimation, and the local roads group will be used for results evaluation.

Another important step in network modeling is to link parcels and traffic count sites to the unabridged network. Similar to the method adopted by zone-level travel demand analysis, some special nodes and links named parcel centroids and parcel

centroid connectors, respectively, can be used to represent parcels and their points of access to the surrounding roadways.

3.3.3. Parcel-level Trip Generation

Parcel-level trip generation is the process used to estimate and quantify the number of trips each parcel will produce and attract. In this research, this step will be implemented by using both the parcel data from the Department of Revenue (DOR) and the trip generation rates and regression equations from the *Trip Generation Manual* (8th edition, 2008), published by the Institute of Transportation Engineers (ITE).

The DOR parcel data describe the rights, interests, and value of properties and it defines the legal boundaries of land parcels in the deed to properties. Real estate tax parcels are typically graphic representations of the land ownership to support property taxing functions. Parcel data forms the basis for all land use and zoning decisions, and represents the location of residences, businesses, and public lands.

Parcels are the lowest geographical level land use. There are typically hundreds of parcels within a TAZ. The lowest level land use scale can provide more accurate and detailed geographical information to help conduct the microscopic transportation study such as AADT estimation for local roads in this research. An example that compares the sizes between parcels and TAZs is illustrated in Figure 3-4. In this figure, the thicker lines are the TAZ boundaries, and the thinner lines are the parcel boundaries.

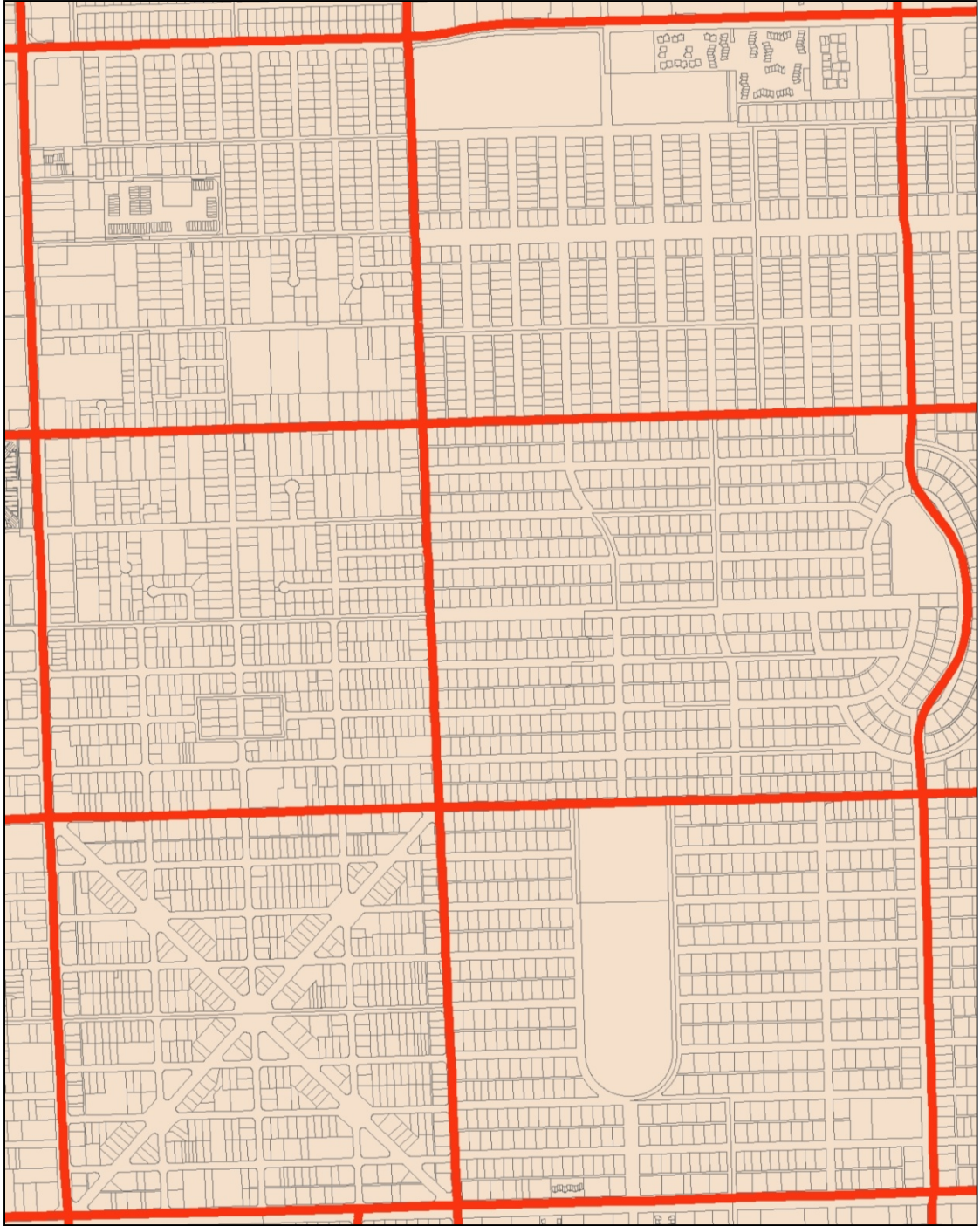


Figure 3-4 Comparison of Parcels and TAZs

The *ITE Trip Generation Report* (8th edition, 2008) is a multi-volume informational report which presents a summary of the trip generation data that have been

voluntarily collected and submitted to ITE. The data used to compile this information report is based on more than 4,800 individual studies conducted in the United States and Canada since the 1960s. These data were submitted voluntarily to ITE by various agencies. In the 8th edition of the report, trip generation rates and/or equations are provided for 10 main land use categories and 162 sub-categories. For a specific land use type, trip generation rates and regression equations (if available) are developed for daily traffic (average weekday, Saturday, and Sunday) and peak hour traffic (AM and PM peak hour for weekday, Saturday, and Sunday).

Figure 3-5 is an example of the statistical and descriptive information available for the majority of the land uses contained in the *ITE Trip Generation Report*. Data plots provide the most fundamental display of the variance within the database. Other important information provided in the report include the land use name, land use code, average trip rate, range of rates, independent variable, number of studies, regression equation, R^2 , etc. As shown in Figure 3-5, this report provides the weekday trip generation information for fast food restaurant with drive-through windows (with land use code 834) based on 1,000 square feet gross floor area, and its average trip rate is 632.125 trips per day.

It should be noted that great care should be taken when selecting the average trip rates and the regression equations to carry out the trip generation analysis. As shown in Figure 3-5, the text above the data plot warns that the data for this land use type should be used carefully because of the low R^2 . Therefore, the descriptions and statistical information provided for each land use should be carefully reviewed.

FAST FOOD RESTAURANT WITH DRIVE-THROUGH WINDOW (834)

Average Vehicle Trip Ends vs: 1,000 SQUARE FEET GROSS FLOOR AREA
On a: **WEEKDAY**

TRIP GENERATION RATES

Average Weekday Vehicle Trip Ends per 1,000 Square Feet Gross Floor Area				
Average Trip Rate	Range of Rates	Standard Deviation	Number of Studies	Average 1,000 Square Feet GFA
632.125	284.000–1359.500	*	8	3.0

DATA PLOT AND EQUATION

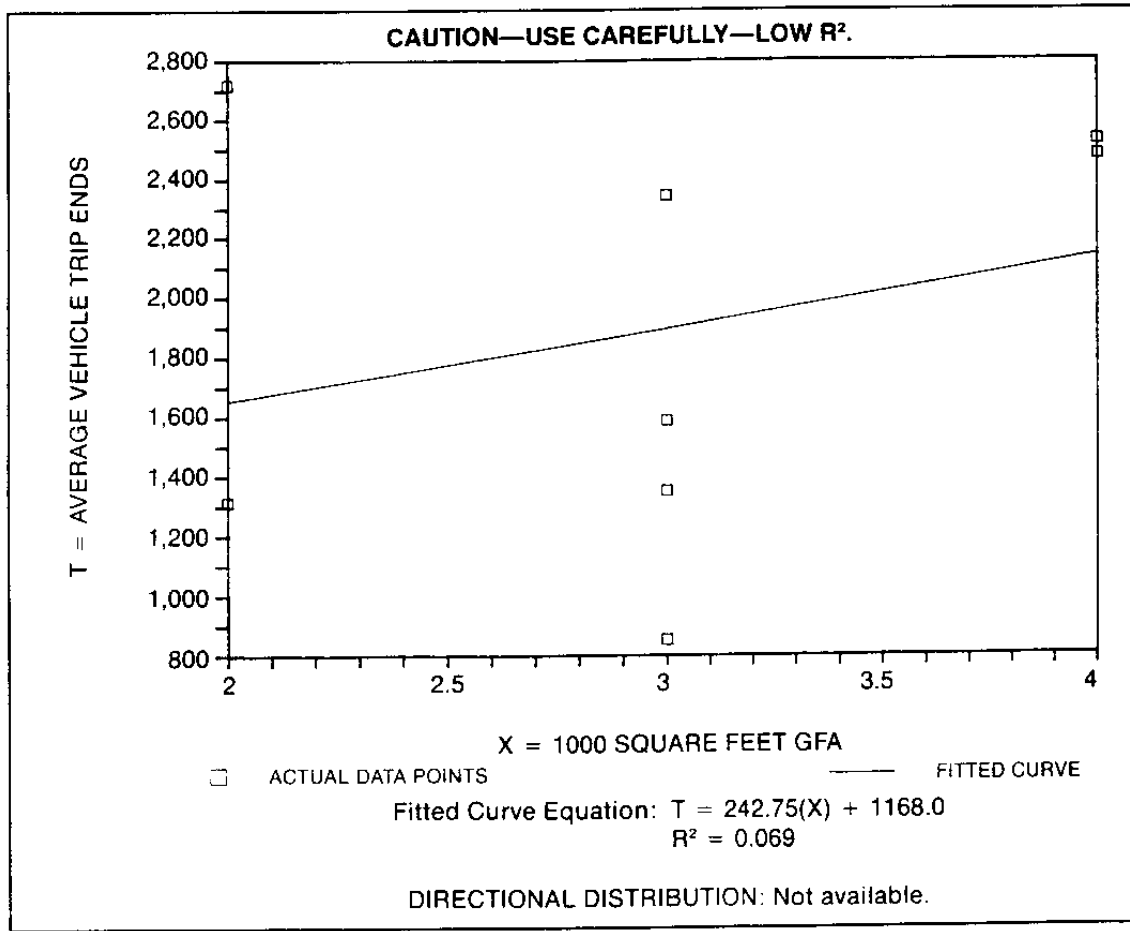


Figure 3-5 Example of ITE Trip Generation Report

Depending on the linear or logarithmic relationship between the independent variable and the dependent variable, there are two forms of regression equations used in the *ITE Trip Generation Report*, which are listed as follows:

$$\text{Linear : } T = aX + b \quad (3-4)$$

$$\text{Logarithmic : } Ln(T) = aLn(X) + b \quad (3-5)$$

where

T = number of vehicle trips generated by a parcel;

X = independent variable such as dwelling units, or gross floor area, etc.; and

a, b = parameters a and b.

Guidelines are provided in the *Trip Generation Handbook* which provides suggestions on selecting among weighted average trip rates, regression equations, and data plots in estimating the trip generation characteristics of a specific land use. Many professionals calculate trip generation characteristics with both the average rate and the regression equation, and then use the one that provides the highest estimate of the number of the trips in the analysis. However, this is not suggested in the handbook. ITE's suggested guidelines on using the average rates or the regression equations and when local data should be collected are listed as follows:

- Use the regression equations when:
 - regression equation is provided,
 - independent variable is within the range of data,
 - data plot has at least 20 points,
 - the R^2 is greater than or equal to 0.75,
 - equation falls within data cluster in plot, and

- standard deviation > 110 percent of the average rate.
- Use the average rate when:
 - at least three data points are available (ITE encourages local data be collected when three to five data points are provided),
 - independent variable is within the range of data,
 - standard deviation is less than or equal to 110 percent of the average rate,
 - the R^2 is less than 0.75 or no regression equation provided, and
 - average rate fall within data cluster in plot.
- Collect local data when:
 - study site is not compatible with the ITE land use code definition,
 - only one or two data points are provided,
 - independent variable does not fall within the range of data, and
 - neither average rate line nor fitted curve falls within data cluster at size of development.

In this research, the guidelines listed above are followed to the extent possible. In the case that a regression equation is lacking or not suitable for trip generation calculation, the average trip generation rates provided by the *ITE Trip Generation Report* will be used to calculate the parcel-level trips.

The *ITE Trip Generation Report* also includes the peak hour traffic information, but in this research, only daily traffic trip generation rates and regression equations will be needed. Due to the different travel patterns among the weekday, Saturday, and Sunday, the number of trips for weekday, Saturday, and Sunday will be calculated separately through the use of either a regression equation or the average trip generation

rate. The final estimated number of trips for a parcel is their average value, which can be calculated as follows:

$$T_{average} = \frac{T_{weekday} \times 5 + T_{Saturday} + T_{Sunday}}{7} \quad (3-6)$$

where,

$T_{average}$ = the final estimated number of trips generated by a parcel;

$T_{weekday}$ = average weekday trips generated by a parcel;

$T_{Saturday}$ = Saturday trips generated by a parcel; and

T_{Sunday} = Sunday trips generated by a parcel.

While parcel-level trip generation can be performed for most of the land use types defined in the parcel database through the steps introduced above, some special types may require further steps with the use of various other demographic or land use databases.

The *ITE Trip Generation Report* has more detailed land use types, so some land use codes in the parcel database encompass several ITE land use types. Table 3-1 lists two such examples.

Table 3-1 Examples of Land Use Types Matching

Parcel Land Use Code	Parcel Land Use	ITE Land Use Code	ITE Land Use
072	Private School	534	Private School (K-8)
		536	Private School (K-12)
023	Financial Institutions	911	Walk-in Bank
		912	Drive-in Bank

In these cases, the ITE rates in the constituent land use type are averaged and weighted by an estimate of the relative presence of each category in the study area, which

can be obtained from other demographic or land-use databases. The calculation can be expressed as follows:

$$T = \sum_{i=1}^n F_i(X) \times P_i \quad (3-7)$$

where

T = number of vehicle trips generated by a parcel;

F_i = the ITE trip generation function (either regression equation or average trip rate, which can be regarded as a special linear regression without parameter b) for ITE land use type i ;

X = independent variable such as dwelling units, or gross floor area, etc.;
and

P_i = the percentage of the presence of ITE land use type i in the study area.

For most of the land use types, ITE trip generation rates are based on dwelling units or gross floor area, which are also the attributes of a parcel in the parcel database. Hence, for a majority of parcels, the trip generation can be calculated directly by using the parcel data. However, for some land use types, if the ITE rates use, as an independent variable, a size attribute that differs from the parcel data, these types need to be adjusted by the ratio between their mean values, and the calculation can be expressed as follows:

$$T = F(X \times R) \quad (3-8)$$

where,

T = number of vehicle trips generated by a parcel;

- F = the ITE trip generation function (either regression equation or average trip rate, which can be regarded as a special linear regression without parameter b);
- X = independent variable such as dwelling units, or gross floor area, etc.;
and
- R = the ratio of mean values between the ITE independent variable and the parcel size attribute.

3.3.4. Parcel-level Trip Distribution

Once the number of trips for each parcel has been generated, the next step is to distribute the trips to the nearby count sites; this is performed in the parcel-level trip distribution step.

While the traditional model distributes the trips generated by each TAZ to all the TAZs in the study area, the proposed model distributes the trips generated by each parcel only to the count sites on the major roads within a certain distribution range.

Similar to the zone-level trip distribution model, parcel-level trip distribution is also derived from Newton's law of gravity. It can be expressed as follows:

$$T_{ij} = \frac{\frac{A_j}{D_{ij}}}{\sum_{k=1}^n \frac{A_k}{D_{ik}}} \times T_i \quad (3-9)$$

where,

- T_{ij} = daily vehicle trips between parcel i and traffic count site j ,
- T_i = total vehicle trips generated at parcel i ,

A_j = AADT estimated from traffic count volume at traffic count site j ,

D_{ik} = the shortest free flow travel time between parcel i and traffic count site k , and

n = number of the nearby traffic count sites within a distribution range.

From the formula, it can be seen that parcel-level trip distribution is to distribute the trips between the parcels and the nearby traffic count sites within a distribution range in a manner that differs from zone-level trip distribution which distributes trips among all the zones. It assumes that the total number of vehicle trips between a parcel and a traffic count site is directly proportional to the trips generated by the parcel and the AADT value estimated from the traffic volume measured at the traffic count site, and is inversely proportional to the shortest free flow travel time between them.

In Figure 3-6 and Table 3-2, an example is given to illustrate the trip distribution calculation procedure. In Figure 3-6, it is assumed that there is a parcel which generates 200 trips per day, and the distribution range is the nearby major roads that surround the area. Within the distribution range, there are eight traffic count sites on the surrounding major roads, and their traffic count data (or AADTs estimated from the traffic count data) are shown in the figure. Table 3-2 lists the assumed free flow travel times from the parcel to the traffic count sites, the calculation procedure, and the calculated trips distributed to each traffic count site.

It should be noted that the process of parcel-level trip distribution introduced above only takes into consideration the case that there is only one centroid connector for each parcel. Theoretically, if a parcel has multiple centroid connectors, the parcel trips are suggested to be split, and the trip distribution for each connector will be the same as

the process described above. However, the parcels with multiple accesses to the roads are usually the large scale land use for business or education, and this type of parcels is typically located adjacent to major roads. Therefore, it should not affect the performance of the model if only one centroid connector is created for each parcel, since this research is to estimate AADTs for local roads alone.

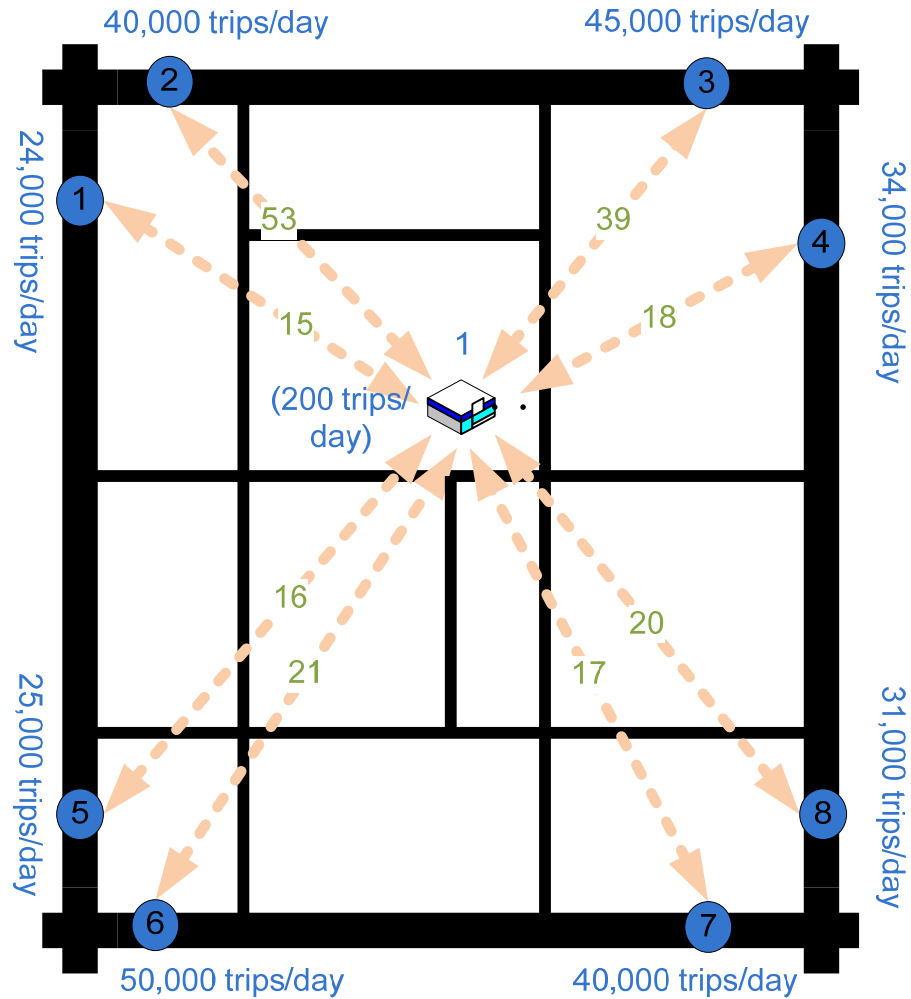


Figure 3-6: Example of Parcel-level Trip Distribution

Table 3-2: Example of Parcel-level Trip Distribution Calculations

i	j	T _i	A _j (Trips/Day)	D _{ij} (Seconds)	A _j / D _{ij}	T _{ij}
1	1	200	24,000	101	237.62	$\frac{237.62}{3081.63} \times 200 = 15$
1	2	200	40,000	49	816.33	$\frac{816.33}{3081.63} \times 200 = 53$
1	3	200	45,000	74	608.11	$\frac{608.11}{3081.63} \times 200 = 39$
1	4	200	34,000	123	276.42	$\frac{276.42}{3081.63} \times 200 = 18$
1	5	200	25,000	101	247.52	$\frac{247.52}{3081.63} \times 200 = 16$
1	6	200	50,000	153	326.80	$\frac{326.80}{3081.63} \times 200 = 21$
1	7	200	40,000	151	264.90	$\frac{264.90}{3081.63} \times 200 = 17$
1	8	200	31,000	102	303.92	$\frac{303.92}{3081.63} \times 200 = 20$
					$\sum_{j=1}^n \frac{A_j}{D_{ij}} =$ 3081 .63	Total Trips = 200

Since the parcel-level trip distribution is to distribute trips between the parcels and the traffic count sites, there are some requirements for the count sites data. Firstly, there should be enough traffic count sites to evenly cover the major roads in the study area. In addition, enough traffic count data on local roads are also suggested to be collected, so that they can be used to evaluate the model. Secondly, if there are no estimated AADT values available, daily traffic count data can also be used. In either case, it is required that all the traffic count sites adopt the same kind of daily volume measurement to maintain consistency.

While a traditional zone-level trip distribution model usually includes the effects of multiple travel impedance factors, such as travel time, cost, etc., parcel-level trip distribution will only consider the shortest free flow travel time. This is expedient because travel time is the major factor that determines the trips on the local roads, and travelers will choose the fastest path to access major roads in order to arrive at their destinations as soon as possible.

The free flow travel time will also be used to determine the distribution range. The trips generated by a parcel will be distributed to only the traffic count sites that can be reached within certain travel time, and the traffic count sites that are too far away from the parcel will not attract any trips, based on the fact that the trips on the local roads are mainly influenced by the nearby surrounding major roads. The distribution range is used to make sure that the trips of a parcel are distributed locally, and as a result, the trips will also be assigned locally in the following parcel-level trip assignment step.

3.3.5. Parcel-level Trip Assignment

Once the number of vehicle trips between the parcels and the nearby traffic count sites has been calculated by the parcel-level trip distribution step, the next step is to predict the routes that the travelers will take to approach these count sites. This step is referred to as the parcel-level trip assignment.

Zone-level trip assignment commonly uses an all-or-nothing with capacity restraint method, also known as the equilibrium assignment method. It is implemented through an iterative procedure in which travel time for each link on the network is calculated at the end of each assignment and used as the input to the next computation of

the minimum impedance routes. The iterative procedure continues until the model reaches equilibrium when further route changes will increase the travel time. This method is suitable for zone-level trip assignment because the trips among zones are usually assigned to the major roads where avoiding congestion to save travel time is the primary concern of travelers. In the case of parcel-level trip assignment, congestion seldom happens on local roads; as such, the simple all-or-nothing assignment method is adopted to assign trips, and only travel time is considered to affect the travelers' route selection.

After the trips for all of the parcels have been assigned, the sum of the trips assigned to each road segment is its estimated AADT.

3.6. Evaluation Method

Finding a good method to evaluate the accuracy for the estimation results of the proposed model is not easy. This is mainly because there are no permanent counters installed on local roads, thus no full year volume data available to calculate the true AADTs. Therefore, the methods introduced in this research will provide an approximate evaluation.

One way to evaluate the proposed method is to compare its results with those from the traditional factor method. This assumes that the traditional factor method with short-term traffic count data is more reliable and can be used as the ground truth data. Some local roads also have portable traffic counters, and the AADT values estimated with the traditional factor method are already available. Hence, those roads can be chosen from the study area as the evaluation locations.

To quantify the difference between the proposed method and the traditional factor method, the following three commonly used measures of accuracy will be used: Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and Mean Absolute Percentage Error (MAPE). Their calculations are expressed as follows:

$$MAE = \frac{\sum_{i=1}^n |F(i) - G(i)|}{n} \quad (3-10)$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (F(i) - G(i))^2}{n}} \quad (3-11)$$

$$MAPE = \frac{1}{n} \sum_{i=1}^n \left| \frac{F(i) - G(i)}{G(i)} \right| \quad (3-12)$$

where,

$G(i)$ = the ground truth AADT at location i ;

$F(i)$ = the estimated AADT at location i ; and

n = the total number of locations.

Results from the other AADT estimation methods are evaluated using the same ground truth data. The results from the USF regression method and the URS method will be chosen and compared with those of the proposed method.

Depending on the availability of enough traffic count data, the study areas will be selected from Broward County in Florida, which was found to have the most complete traffic count data for its local roads. Hundreds of traffic counters are deployed each year in this county to collect traffic volume. For all the evaluation locations on the local roads, the results of the three methods, the USF method, the URS method, and the

proposed method, will be compared with the ground truth AADT, separately, and the difference of each method from the ground truth AADT will be measured and compared.

To check the performance of the proposed method for the lowest level roads without any traffic count data available, the results of the three evaluation methods for this type of roads will also be checked and compared based on reasonableness.

3.4. Summary

In this chapter, the procedure of the traditional four-step travel demand forecasting model was first introduced as the background information. The methodology of the proposed parcel-level travel demand analysis model was then described in detail. Lastly, the method to be used to evaluate the proposed method was also presented.

Compared to the traditional travel demand forecasting model, the parcel-level travel demand model was simplified and optimized for estimating AADT on local roads, which is one of the major contributions of this study. The major differences between these two models are summarized as follows:

- The objectives of the two models are different. The major objective of the traditional travel demand model is to predict the changing traffic volume or transit ridership in the future caused by various projects, policies, and programs on the highway and transit facilities. The objective of the proposed parcel-level travel demand model is to estimate AADT for local roads in a study area based on the highway facilities and the traffic count data already collected on major roads.

- While the traditional model performs the travel demand analysis on an abridged roadway network with only major roads, the proposed model simulates the trips on an unabridged roadway network including the local roads.
- The trip generation methods of the two models are different. The traditional travel demand forecasting model separates the calculation of trip productions and attractions with the aggregated socioeconomic data at the zone level, but the parcel-level travel demand model uses the detailed DOR parcel data and ITE trip generation rates and equations to perform the parcel level trip generation without distinguishing between trip production and attraction. In addition, unlike the traditional model which has to convert the calculated person trips to vehicle trips, the parcel-level model calculates the vehicle trips directly. Further, the parcel-level model does not take into account the different trip purposes.
- The trip distribution procedures of the two models are different. While both models use the similar gravity model to perform trip distribution, the traditional travel demand model calculates the trip interchanges between each pair of zones, and the parcel-level model distributes the trips generated by a parcel to the traffic count sites.
- The parcel-level travel demand model does not have the mode choice step, a standard step for the traditional four-step travel demand forecasting model. Transit services are commonly provided on the major roads, so travelers usually access the nearby public transit facility such as bus stops by walking. Even if they access via automobile, the trips will be included by the parcel-level trip

generation step. Therefore, it is not necessary for the parcel-level travel demand model to include the mode choice step.

- The trip assignment methods of the two models are different. While the traditional travel demand model uses the equilibrium assignment method which involves multiple assignment iterations, the parcel-level model uses the simpler all-or-nothing assignment method.
- Unlike the traditional travel demand forecasting model, the result of which is a loaded network with bidirectional volume, the parcel-level travel demand model needs to perform additional post-processing on the loaded network to obtain the final AADT estimations on local roads.

CHAPTER 4

MODEL DEVELOPMENT

4.1. Introduction

As mentioned in Chapter 1, the goal of this research is to develop an improved method of estimating AADT for local roads by applying travel demand modeling techniques. In Chapter 3, the methodology of the proposed parcel-level travel demand analysis model has been described. This chapter describes the implementation of the proposed model in detail. An overview of the model implementation is first introduced, and the implementation of each step involved in the model is then explained.

4.2. Model Development Overview

A parcel-level travel demand analysis model was implemented to estimate AADT for local roads. The following two development tools have been adopted to implement this model: ArcGIS from Esri and Cube from Citilabs.

ArcGIS is a software suite consisting of a set of Geographic Information System (GIS) software products produced by Esri. ArcGIS has been widely used for creating, analyzing, and managing geographic information in many applications. In this research, ArcGIS 10.0 was used to perform both data pre and post-processing.

Cube is a travel demand modeling software product marketed by Citilabs. It has been widely used for transportation planning to analyze and estimate the impacts of a wide range of infrastructure improvements and operating policies. In this research, Cube 5.1.3 was used to develop the four model steps. Cube Voyager is a module of the Cube software suite. It provides a script-base structure allowing the implementation of

multiple model methodology including standard four-step model, discrete choice model, and activity-based model. It also provides a comprehensive library of functions for the modeling and analysis of passenger transportation systems. To implement the parcel-level travel demand model proposed in this research, Cube Voyager scripts were developed to customize the standard four-step model templates provided by Cube Voyager. Because the standard templates are designed for traditional zone-level travel demand modeling, it was necessary to customize them to simulate the parcel-level travel patterns with the zone-level implementation. A parcel defined in the proposed model can be treated as a small size Traffic Analysis Zone (TAZ) used in the traditional travel demand modeling.

Figure 4-1 shows the system components and the procedure used to estimate AADT. The procedure can be divided into the following sub-steps:

- ArcGIS is used to preprocess the input data for the model including the DOR parcel data, unabridged highway network data, and traffic count sites data.
- The preprocessed input data are imported into Cube, and the highway network is built from the unabridged roadway shape file.
- The built highway network is used by the network modeling step to calculate the free flow travel time skim matrix.
- The parcel-level trip generation step is performed by using the merged DOR parcel data and traffic count sites data as well as the trip generation rates and regression equations provided by the *ITE Trip Generation Report*.
- A parcel-level trip distribution gravity model is used to distribute the generated trips between the parcels and the nearby traffic count sites.

- The distributed trips are assigned to the network by using all-or-nothing assignment method in the parcel-level trip assignment step.
- The traffic volume data of the loaded network are exported, and ArcGIS is used again to calculate the final AADTs, which are then joined with the original roadway network to get the roadway network with AADTs.

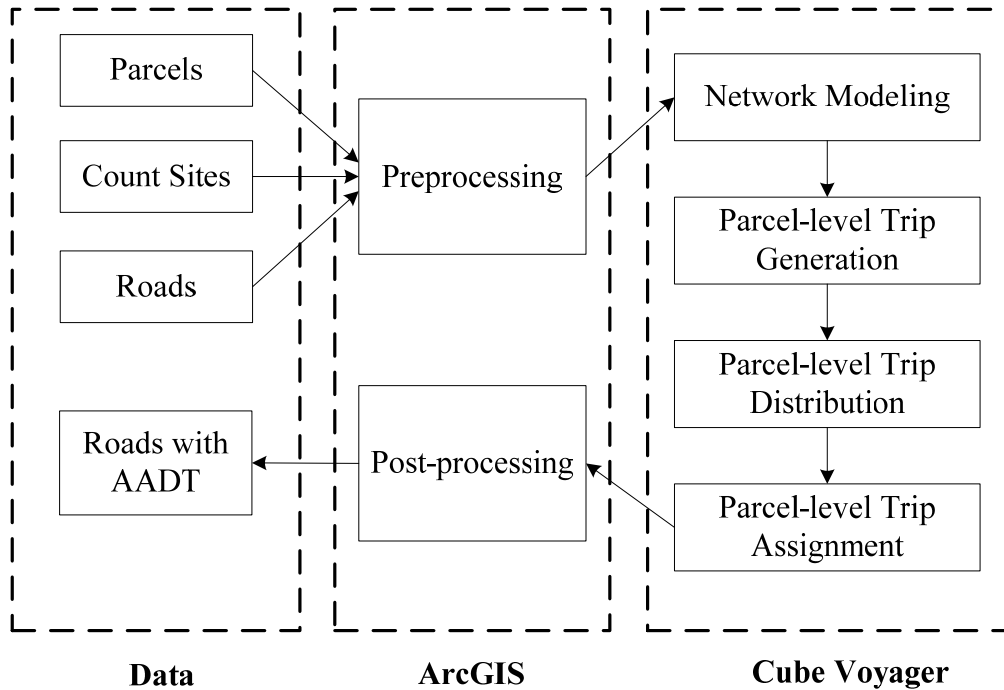


Figure 4-1 System Components and Procedure

The ArcGIS component was implemented with an ArcGIS application called ArcGIS ModelBuilder, which provides a visual programming environment allowing users to graphically link geoprocessing tools into models. While the models built with ModelBuilder can be executed directly in ArcGIS, they can also be exported to scripting language such as Python. The Python scripts can be called with the Cube Voyager Pilot program, so theoretically all the steps of the ArcGIS part can be integrated into Cube to simplify the running of the entire model. However, because this part has called some

geoprocessing tools that are supported only by ArcGIS 10.0, which is not compatible with the current version of Cube (5.1.3), integration of ArcGIS into Cube has not be implemented. Nevertheless, this incompatibility would not affect the results of the entire model.

Figure 4-2 shows the model steps and the input and output files for each steps implemented in Cube. It can be noted that the four model steps are integrated. When the model is run, the four steps are executed in sequence, and the output files of a previous step becomes the input files of a later step. If there were no compatibility problems as mentioned above, the ArcGIS part should have been combined with the Cube, and the steps shown in this Figure 4-2 would be all the steps involved in the entire model.

Parcel-level Travel Demand Analysis Model

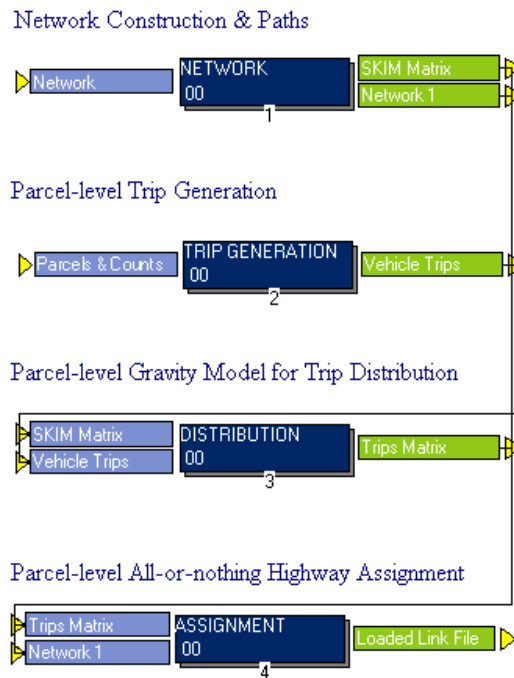


Figure 4-2 Model Steps in Cube

Table 4-1 summarizes the input and output files for each step. There are two input files for the Cube. One is the network file preprocessed by the ArcGIS, and the other is the DBF file for the merged parcels and traffic count sites shape file which is also generated by the ArcGIS. There is one output file generated by the Cube part, and it is the DBF link file with the traffic volume information exported from the loaded network assigned in the parcel-level trip assignment step. Among the steps, the output files of a preceding step may become the input files of a later step.

Table 4-1 Input and Output Files

Model Step	Input File	Output File
Network Modeling	Preprocessed Network File	Free Flow Time SKIM Matrix File
		Modified Network File
Parcel-level Trip Generation	Merged Parcels and Counts DBF File	Vehicle Trips DBF File
Parcel-level Trip Distribution	Free Flow Time SKIM Matrix File	Distributed Trips Matrix File
	Vehicle Trips DBF File	
Parcel-level Trip Assignment	Distributed Trips Matrix File	Link DBF File with Volume Exported from Loaded Network
	Modified Network File	

The following sections describe the steps in implementing the parcel-level travel demand model. Preprocessing of the input data with ArcGIS will be introduced in the Network Modeling step, and the calculation of the AADT values from the loaded network and the implementation of the evaluation with ArcGIS will be described under the Parcel-level Trip Assignment step.

4.3. Network Modeling

The implementation of the Network Modeling step includes the following four sub-steps:

- 1) Preprocess the input data in ArcGIS.
- 2) Build the Cube network file from the roadway shape file.
- 3) Create the centroid connectors in Cube.
- 4) Calculate the free flow time skims matrix in Cube.

The implementation of each of these steps is described in detail below.

4.3.1. Data Preprocessing

The preprocessing of the input data involved:

- Dividing the traffic count site point data into two groups, one for estimating AADT, and the other for evaluating the results.
- Buffering traffic count site points for AADT estimation and merging them with the DOR parcel polygon data.
- Splitting roadway polylines at the parcels' access points.

Division of traffic count sites was based on the level of the road at which a traffic count site is located. Count sites on the major roads were used for AADT estimation, and those on the minor local roads were used for results evaluation. This step is not required, but it is highly recommended if there are many traffic count sites located on the minor roads. This will not only help provide the required data for the results evaluation but also improve the accuracy of AADT estimation.

TAZ boundaries were used to locate traffic count sites on the major roads. Figure 4-3 shows the model used in ArcGIS to divide count sites into estimation and evaluation groups. As shown in Figure 4-3, the input data used were the TAZ polygon data and the traffic count site point data. The TAZ polygons were converted to the polylines, the TAZ boundaries, so that they can be processed by ArcGIS to create buffers on both sides of the TAZ boundaries. The traffic count sites located within the TAZ boundary buffers were erased first to retrieve those located on the local roads. The results were compared with the original traffic count sites, and differences were saved as the traffic count sites located within the TAZ boundary buffer and on the major roads. All the traffic count site points located on the major roads were used for AADT estimation, and the count sites located on local roads were used for result evaluation.

The traffic count sites for AADT estimation had to be merged with the parcels, because the trips were to be distributed between them, and they were treated like the TAZs in the traditional zone-level travel demand model. The traffic count site points were buffered first so that they have the same feature types with the parcel data. Not all the parcels were used. Depending on their land use types, very few or no trips could be generated by some parcels such as vacant residential, rights-of-way streets, roads, canals, camps, rivers, lakes, etc. A total of 42 parcel land use types are listed with "N/A" in the ITE land use column in Table A-1 of Appendix A and were not used. After the merging procedure was completed, a new field named "TAZ" was added, and its values range from one to the total number of merged parcels and count sites. This field is required because it ensures that the centroid connectors would be created successfully with Cube in the next step. Figure 4-4 shows this procedure.

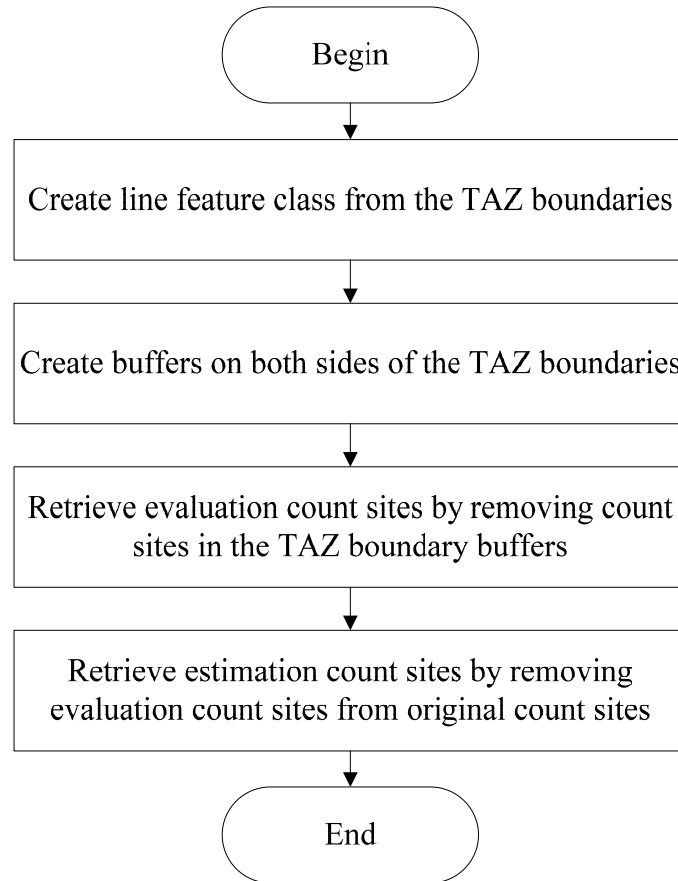


Figure 4-3 Dividing the Count Sites into Estimation and Evaluation Groups

It is also necessary to split the roadway polylines at the access points of the parcels so that the centroid connectors can be created correctly with Cube in the next step. Cube provides a functionality to automatically add centroid centers and centroid connectors, but the connectors can only be created between two nodes. This means that a centroid connector will always connect a centroid center node to its nearest intersection node. An example of a subarea with the centroid connectors created incorrectly is shown in Figure 4-5, in which the gray lines represent the added centroid connectors connecting the parcels to the closest intersections.

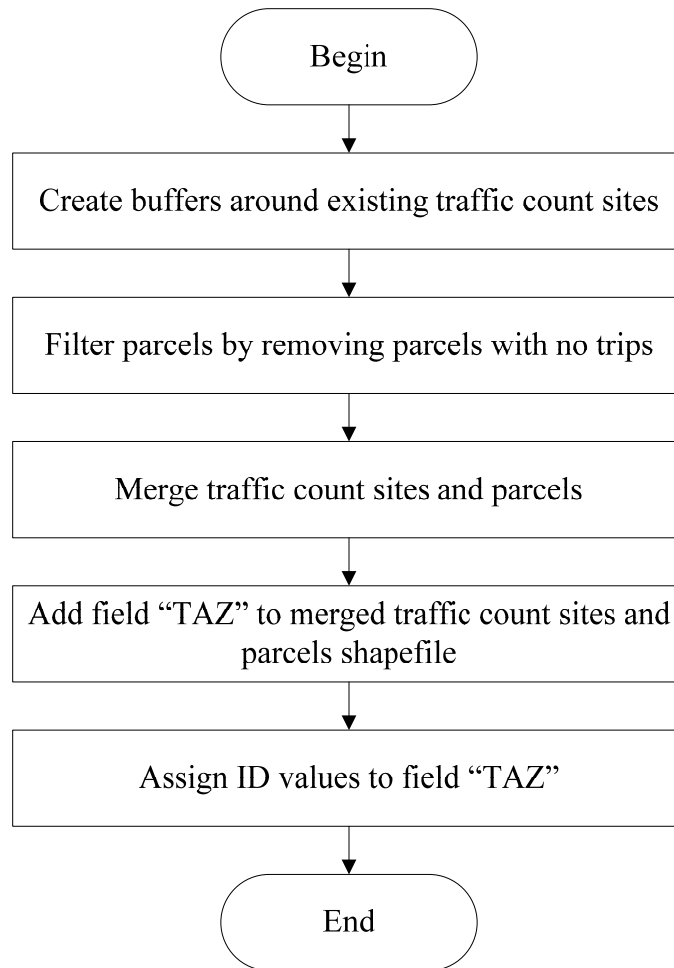


Figure 4-4 Merging the Parcels and the Traffic Count Sites

Connecting the parcels to the closest intersections will seriously reduce the accuracy of the AADT estimation. To prevent this from happening, the access points of the parcels on the roads can be estimated and inserted as nodes into the road segments, and the centroid connectors will then connect the centroid centers to the closest roads instead of the closest intersections. After the roadway polylines are split, three fields named “A”, “B”, and “FF_TIME” were added. The fields “A” and “B” were required for Cube to create the centroid connectors automatically and the field “FF_TIME” was

used to store the Free Flow Time calculated in the next step. The procedure is shown in Figure 4-6.



Figure 4-5 Centroid Connectors Incorrectly Connecting Parcels to Intersections

Preprocessing of the input data generated two shape files: the split roadway file and the merged parcels and count sites file. These generated files were later used to build the Cube network file and create the centroid connectors automatically on the network in the sub-steps to follow. The DBF file associated with the merged parcels and count sites shape file were later used as an input file to the parcel-level trip generation step. This file is similar to the TAZ file in the traditional zone-level travel demand model.

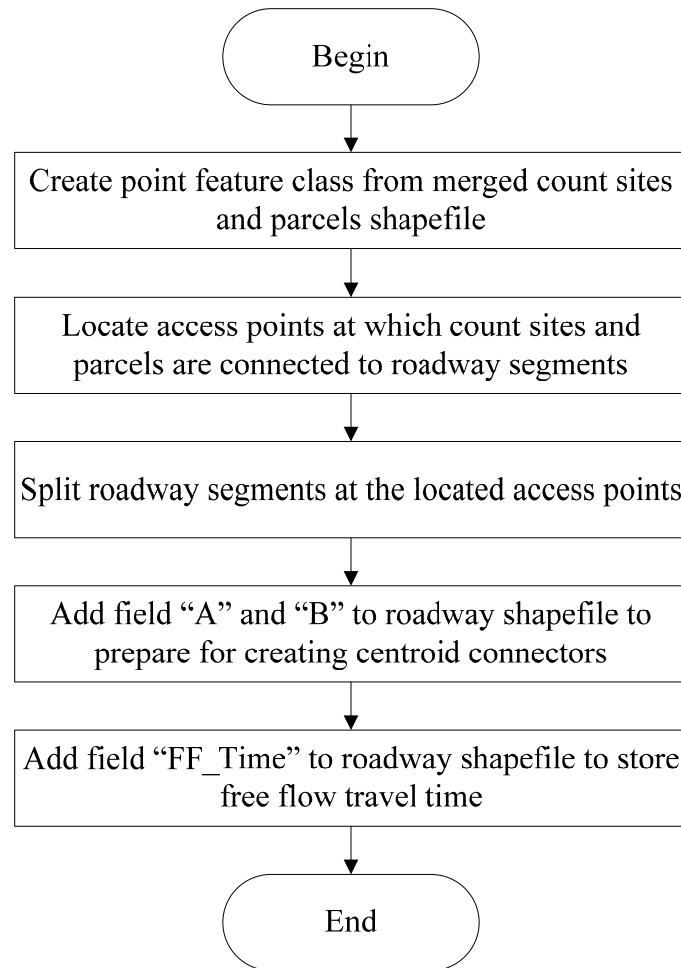


Figure 4-6 Splitting the Roadway Polylines at the Parcel Access Points

4.3.2. Build Cube Network File

In this research, the unabridged roadway network data are in the shape file format, but Cube models are based on the highway network file format which was defined by Citilabs, so the preprocessed roadway shape file has to be converted to the highway network file of Cube. Cube Base provides the functionality and interface to build a highway network file from either a shape file or a geodatabase's feature class. The dialog to build the Cube highway network from the line shape file or the line feature class is shown in Figure 4-7.

Build Highway Network from Line Shape File or Line Feature Class

Output Binary Network File or Geodatabase Network Name:
 Browse ...

Input Line Shape File or Geodatabase Line Feature Class:
 Browse ...

Use Node Numbers from Network, Point Shape File, or Geodatabase Point Feature Class:
 Browse ...

Node Number Field

Join Point and Line Shape Files using ID, FROM_ID and TO_ID fields

Please specify the A-Node, B-Node and the 1-way/2-way indicator fields from the line shape database.

A-Node Field Name

B-Node Field Name

Clear all values in the A-Node and B-Node fields first

Level Field Name

1-way/2-way Options:

All 1-way

All 2-way

Use indicator field:

Consolidate AB/BA Field Pairs Change All 0 Values to 2 (for 2-Way)

AB Field Mask (e.g. AB_* FT* *_AB)

BA Field Mask (e.g. BA_* TF* *_BA)

Distance Options:

Add Distance Field Scale

Do Not Add Distance Field

Node Grouping Limit

Starting New Node Number Highest Zone Number

Figure 4-7 Build Highway Network from Line Shape File Dialog

In this dialog, the output binary network file and the input line shape file can be defined. The node fields A and B have been added to the input line shape file in the last sub-step and can be chosen in this dialog. If the input line shape file has a field to indicate one-way or two-way, it can be chosen; otherwise, always two-way is chosen since AADT is non-directional. If the input line shape file does not have an attribute for the road segment distance, the Add Distance Filed option should be chosen so that the free flow time can be calculated in a later step. The highest zone number is the number

of the merged parcels and traffic count sites in the study area, and the “Starting New Node Number” should always be greater than the highest zone number and can be the highest zone number plus one. All other items can be based on the default values.

After the highway network is built, it was compared with the original shape file to check if the build network was the same as the original roadway layout, this was needed because Cube might simplify some curved road segments and convert them to straight lines. An example of this mismatching is shown in Figure 4-8. Fortunately, Cube provides a functionality to fix this problem. It is implemented by overriding the original shape file and the built highway network as two layers and correcting the difference between the two layers on the built highway network layer. The Display True Link Shape dialog is shown in Figure 4-9. By using the interface provided by Cube, the built network can be fixed by taking the actual shape of the roadway in the shape file.

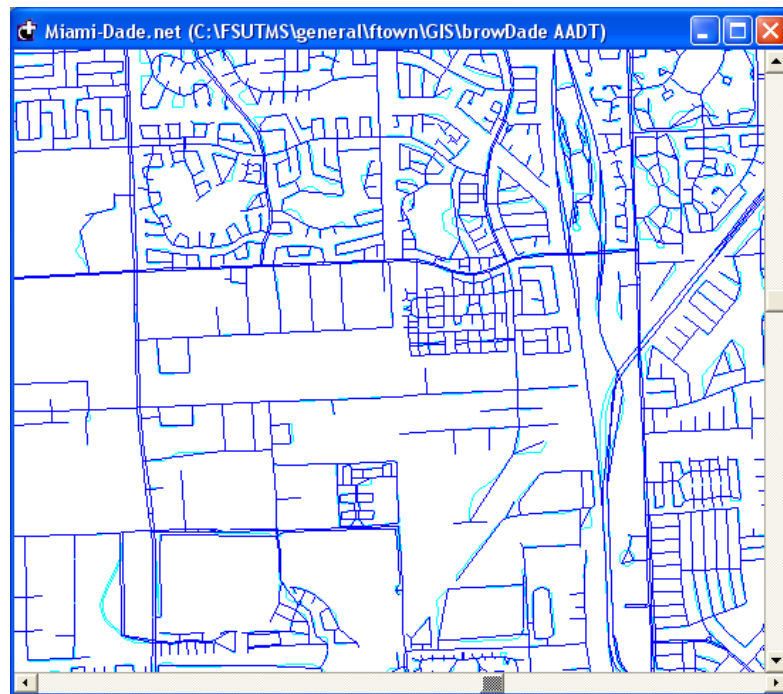


Figure 4-8 Mismatching of the Built Network with Original Shape File

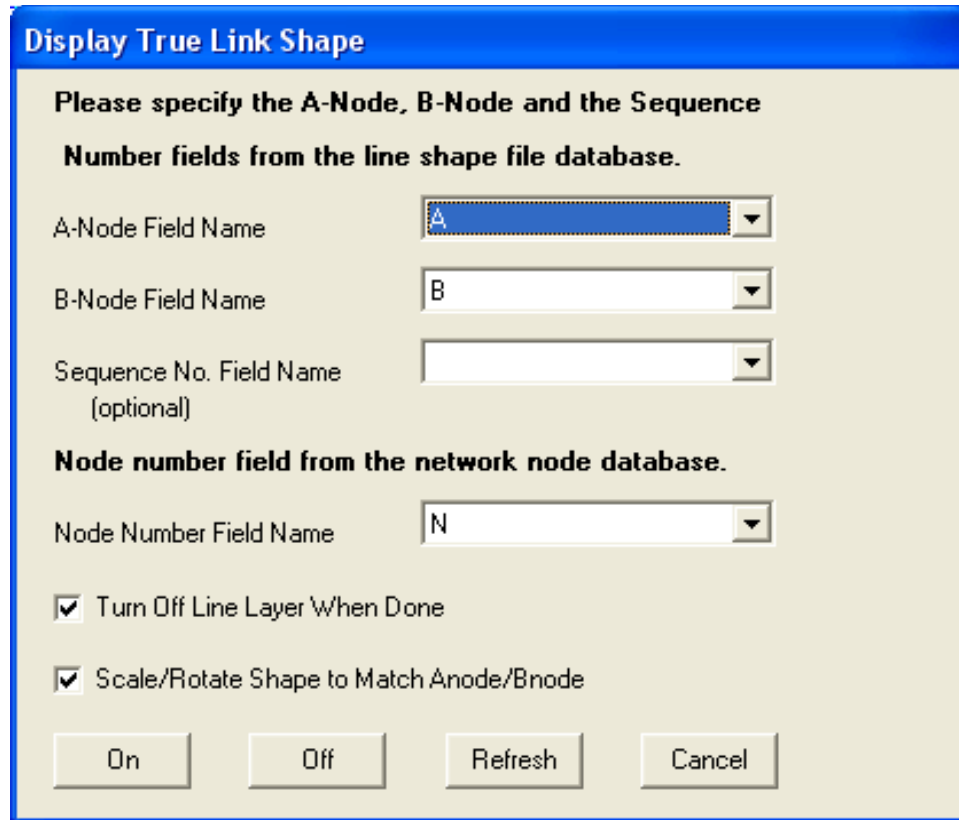


Figure 4-9 Display True Link Shape Dialog

The shape source data have now been converted to a highway network for use in modeling, and the next sub-step was to create the centroid centers and connectors for the highway network.

4.3.3. Create Centroid Centers and Connectors

Cube provides a functionality to add centroid centers and centroid connectors automatically. To use this function, the highway network must be loaded into the highway network layer with the correct number of zones specified, and the nodes inserted into the road segments in the data preprocessing step are also necessary for this function. The Automatic Centroid Connectors Generation dialog for the user to specify the

parameters of the generation is shown in Figure 4-10. In this dialog, some important parameters to generate the centroid centers and centroid connectors can be defined. The Maximum Number of Connectors to Generate option was set as one, and default values were used for other parameters. An example of a subarea with the added centroid connectors connecting the parcels and the closest roads is shown in Figure 4-11. The green polygons represent the parcel boundaries; the blue lines represent the roadway; and the gray lines are the added centroid connectors.

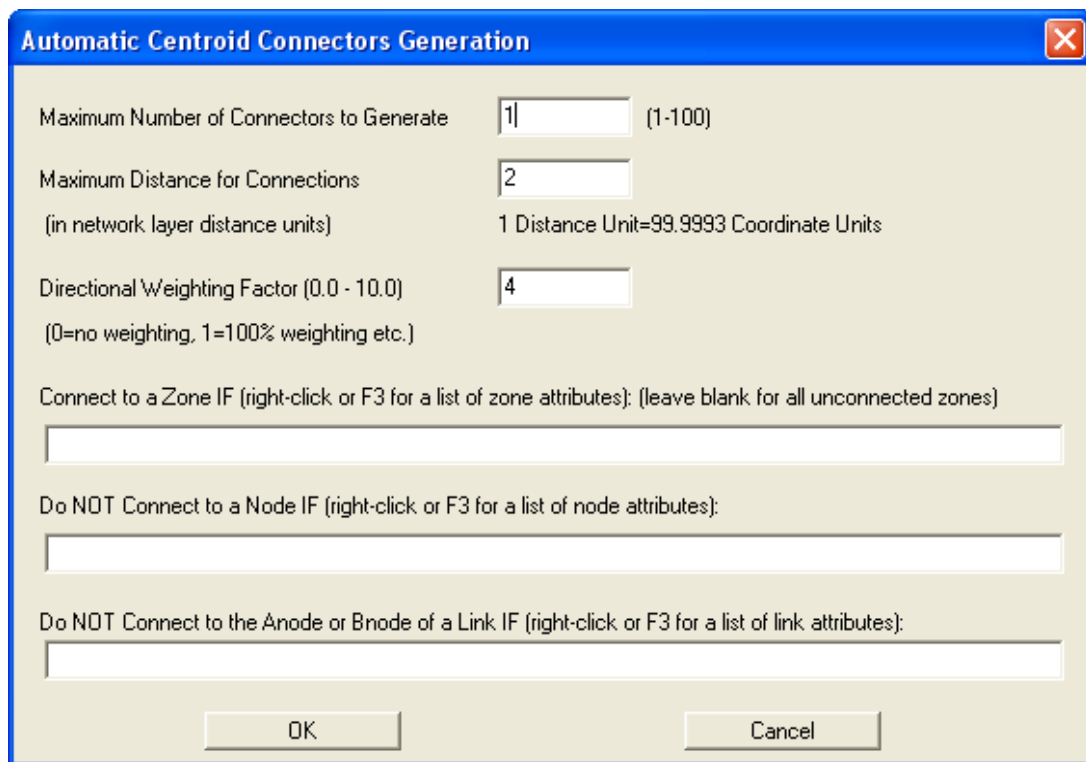


Figure 4-10 Automatic Centroid Connectors Generation Dialog

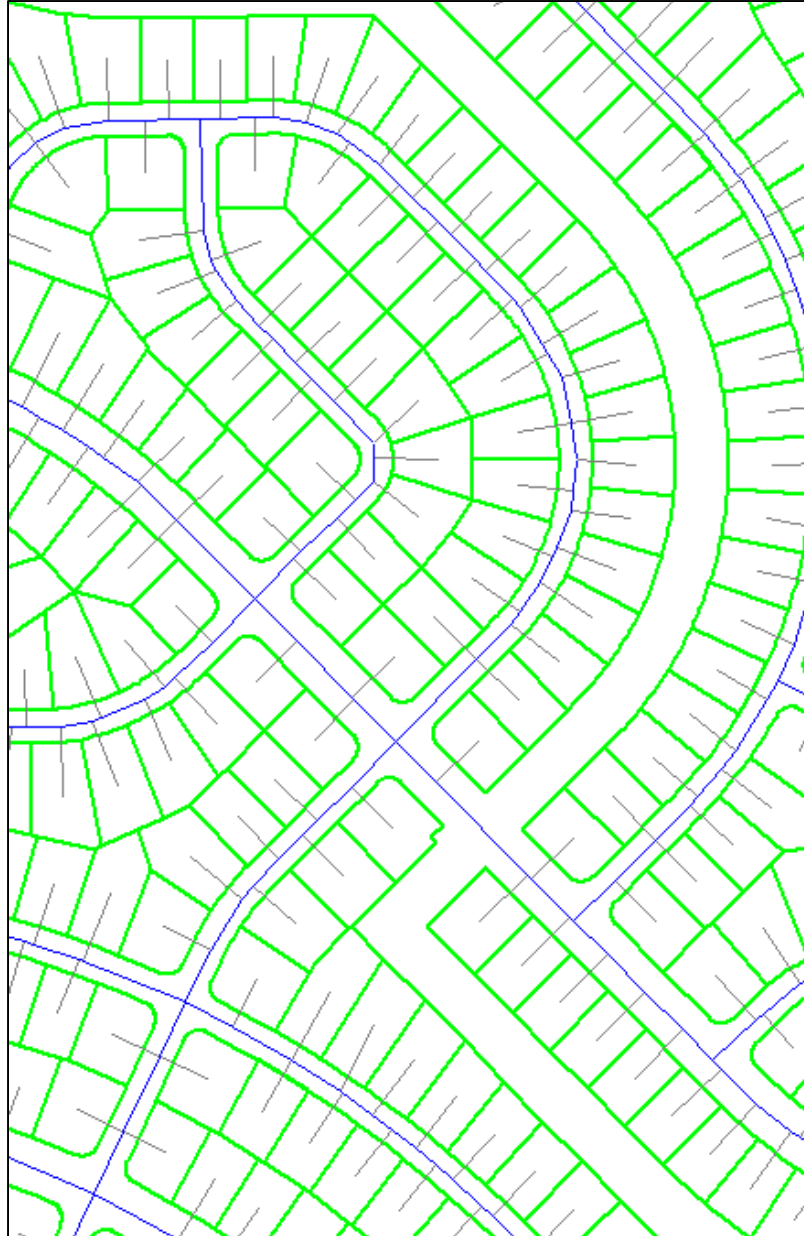


Figure 4-11 Example of Centroid Centers and Connectors Added

4.3.4. Calculate Free Flow Travel Time Skim Matrix

After building the Cube network file and correctly adding the centroid centers and connectors, the next step was to calculate the free flow time skim matrix, which contains the free flow travel times between each pair of parcels/count sites, although only the free

flow travel times between the parcels and the count sites were used by the model. The components of this sub-step are shown in Figure 4-12. It shows that the Cube Voyager Network program was called to calculate the free flow time on each roadway segment with its distance and speed values, and the Cube Voyager Highway program was called to generate the skim matrix file. This was implemented by using Cube Voyager script programming.

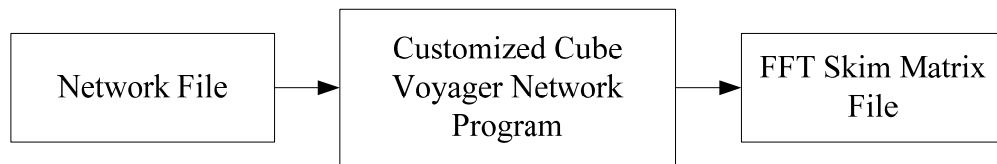


Figure 4-12 Components of FFT Skim Matrix Calculation

4.4. Parcel-level Trip Generation

The parcel-level trip generation step is to estimate the trips generated by each parcel in the study area. To implement this step, the DBF file associated with the merged parcels and count sites shape file generated by data preprocessing with ArcGIS was used as the input file, and Cube Voyager Matrix program was called and customized to calculate the trips based on each parcel’s land use type. The output file was also a DBF file containing fields such as “TAZ”, “Production”, and “Attraction”. The calculated parcel trips were saved in the “Production” field, and the attraction values of the parcels were zero. The count sites have zero production values, and their attraction values were the AADTs estimated from count data. The components of the parcel-level trip generation step are shown in Figure 4-13.

To calculate the trips of a parcel based on its land use type, it is necessary to match the two kinds of land use type classification from the DOR parcel data and the *ITE*

Trip Generation Report. For tax assessment purpose, DOR parcel data assign each parcel with a land use type by using a land use code. There are a total of 100 land use types that were classified. The *ITE Trip Generation Report* is based on even more detailed land use types with more than 162 specific land use types classified. Therefore, to implement the parcel-level trip generation step, it is important to match the two sources of land use type well.

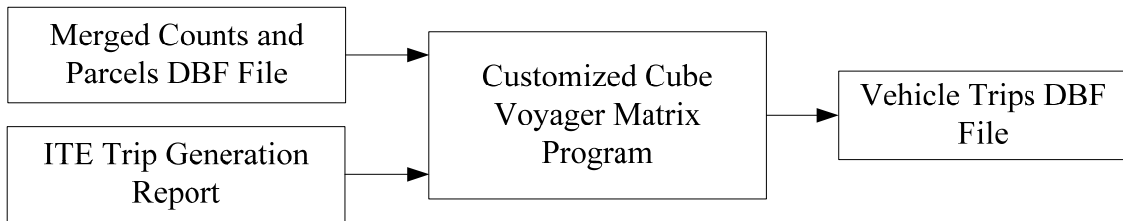


Figure 4-13 Components of Parcel-level Trip Generation Step

For each land use type, the *ITE Trip Generation Report* provides trip rates information based on several independent variables such as Gross Floor Area (GFA), employees, and dwelling units. The DOR parcel data have many attributes, three of which can be used to match the ITE Trip Generation independent variables. Table 4-2 lists the matching of parcel attributes and the independent variables in the *ITE Trip Generation Report*.

Table 4-2 Parcel Attributes and ITE Trip Generation Independent Variable Matching

Parcel Attributes	ITE Trip Generation Independent Variable
NORESUNTS	Dwelling Units
TOTLVGAREA	1000 Sq. Feet Gross Floor Area
ACRES	Acres

The *ITE Trip Generation Report* provides three methods of estimating trips. The data plots can only be used to graphically obtain a rough estimation of trips, so it is not practical to use for this model. The problem with the regression equations is that there are many instances when it will result in illogical estimation of trips if the independent variable is significantly less than the average-sized value. The parcel-level trip generation is based on each parcel, so the independent variables are usually much lower than the average-sized value. Therefore, the weighted average trip rates were used for most of the land use types.

Appendix A summarizes the matching of land use types of parcel data and the *ITE Trip Generation Report*, the selected independent variables, and the selection of the estimation method (average rate or regression equation) for each land use types. From Table A-1 in Appendix A, it can be noted that 42 parcel land use types can be dismissed as they generate either zero or an insignificant number of trips. All the dismissed land use types are listed as "N/A" in the "ITE Land Use" column, and their estimated parcel trips were zero.

4.5. Parcel-level Trip Distribution

The parcel-level trip distribution step distributes trips between the parcels and the traffic count sites. The input files of this step are the results of the two previous steps: the free flow skim matrix file generated by the network modeling step and the production and attraction DBF file generated by the parcel-level trip generation step. The Cube Voyager Distribution program was called to calculate the trips. The output file was a

Cube Matrix file containing the trips between each pair of parcels/count sites. The components of the parcel-level trip distribution step are shown in Figure 4-14.

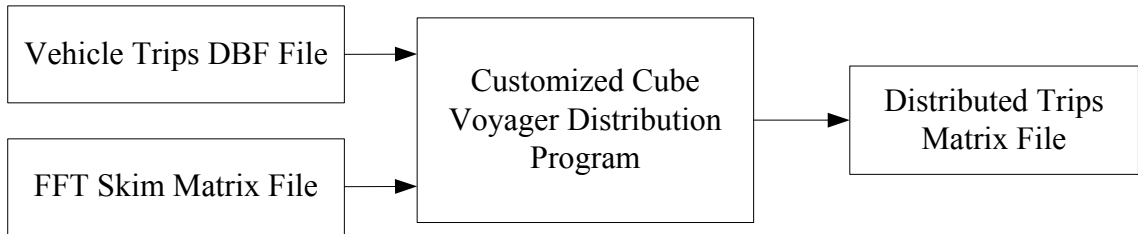


Figure 4-14 Components of Parcel-level Trip Distribution Step

It should be noted that it is not necessary that the trips generated by a parcel to be distributed to all the traffic count sites in the study area. The distribution range can be adjustable, and Cube provides a keyword, LOSRANGE, to specify the range of LOS values that are valid for use in the distribution process. For example, if the LOSRANGE is set as 0-10, the trips from a parcel will be distributed to the count sites which can be reached within 10 minutes, and there will be no trips to those count sites farther than 10 minutes. Theoretically, the travelers are most likely to access the closest higher level state-roads as soon as possible to reduce travel time, so the traffic counts close to a parcel tend to attract more trips, and the distribution range should be very small. To verify that, different distribution ranges were chosen and tested, and the final distribution range used was 5 minutes.

4.6. Parcel-level Trip Assignment

The parcel-level trip assignment step assigns trips between the parcels and the count sites onto the routes. The trips matrix file generated by the trip distribution step and the highway network file were used as the input files. The Cube Voyager Highway

program was called to perform the all-or-nothing assignment. In addition, the Cube Voyager Network program was called to extract traffic volume data for each road segment from the loaded network file and save them into a DBF file, which was later joined with the original roadway shape file to calculate the final AADT values in ArcGIS. The components of the parcel-level trip assignment step are shown in Figure 4-15.

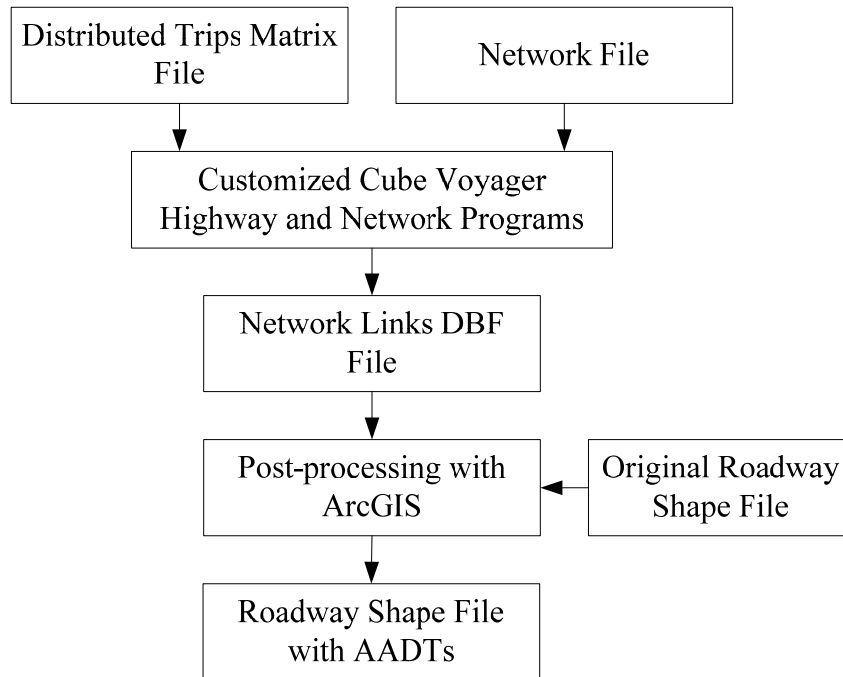


Figure 4-15 Components of Parcel-level Trip Assignment Step

Trip assignment is usually the most time consuming step for the traditional zone-level travel demand step. This is because the implementation of the equilibrium assignment method involves running several iterations of the assignment procedure with an adjustment for the travel time. However, because the all-or-nothing assignment involves only one iteration, the parcel-level trip assignment does not take a long time to execute even with a high number of parcels.

After exporting the loaded network to the DBF file containing traffic volume data, ArcGIS was used to calculate the final AADTs and link the results to the original shape file of the road network. Figure 4-16 shows the procedure for calculating the final AADT values.

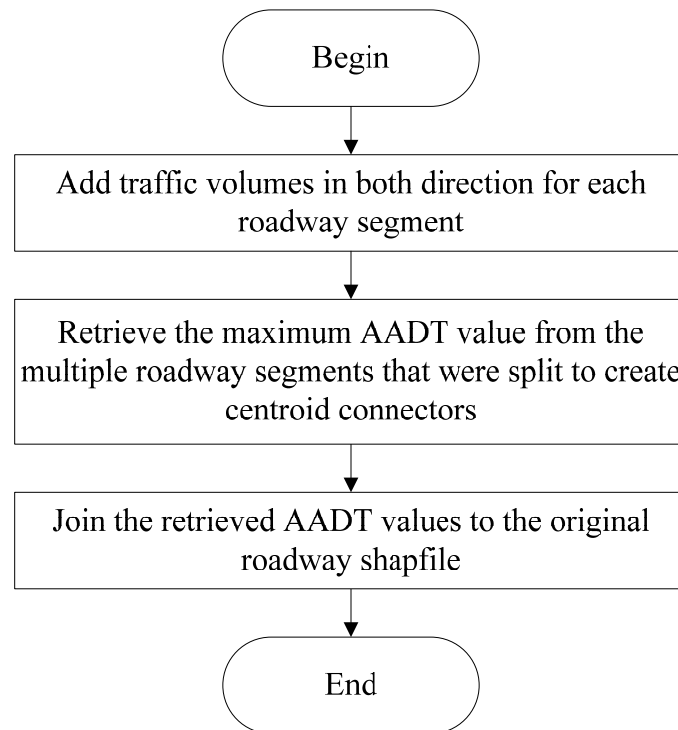


Figure 4-16 Calculating the Final AADTs

The volume data exported from Cube have two values, one for each direction. However, AADT is bidirectional traffic volume. As such, the two directional volume values were summed up for each road segment. In addition, because the roadway polylines were split at the access points of the parcels in the network modeling step, the calculated AADTs provide results that are too detailed. This means that there could be multiple AADTs each road segment depending on the number of access points. To address this issue, the maximum AADT value of the multiple road segments was used as

the final AADT for the original segment. After calculating the AADT values for the road network, the results were joined with the original roadway shape file.

4.7. Summary

This chapter has described the implementation of all the steps of the parcel-level travel demand model. ArcGIS and Cube were used as the development tools to implement the model. First, ArcGIS was used to preprocess the data. In the network modeling step, the data preprocessing was performed to classify traffic count data for AADT estimation and results evaluation, merge the parcels and the traffic count sites feature classes, and split the roadway segments by locating and inserting parcel access points. Second, Cube was used to build the network file from the highway shape file preprocessed with ArcGIS and create the centroids and centroid connectors on the network automatically. Cube Voyager scripts were written to calculate the free flow time skim matrix and develop trip generation, trip distribution, and trip assignment. The loaded network was exported from Cube, and ArcGIS was used to calculate the final estimated AADTs, and combine the results to the original roadway network file.

CHAPTER 5

MODEL OUTPUT AND EVALUATION

5.1. Introduction

In this chapter, the evaluation of the implemented parcel-level travel demand analysis model is described. An overview of the evaluation procedure is first introduced, and then each step of the evaluation procedure and the relevant results are analyzed.

5.2. Evaluation Method

Depending on the availability of traffic count data, the study areas were chosen from Broward County in Florida. Because Cube can only process a maximum of 32,000 zones at a time, the size of the study areas was limited to a maximum of 32,000 parcels and traffic count sites.

To evaluate the performance of the proposed method, results of the USF regression method and the URS method were compared with those of the proposed method. The comparison was performed at different levels. Firstly, the three methods were compared based on a selected study area. The traffic count sites located on the local roads in this area were selected as the evaluation count sites. The results of the three methods were compared with the AADT values estimated from traffic count data which were used as the ground truth AADTs. The overall estimation errors for this study area were also calculated and compared.

To measure the change of the three methods' performance with the change of the locations and the area types, the comparison was conducted for 10 selected study areas to

cover more locations in Broward County, and the standard deviations of the estimation errors for these study areas were calculated and compared.

To further compare the performance of the three methods, the overall estimation errors for the 10 study areas were also calculated and compared.

Lastly, one of the scenarios was chosen to show the performance of the three methods for the lowest level local roads without any traffic count sites. Subjective judgment was used to compare the results of the three methods. This type of comparison may not be very accurate as it is based on intuition and reasoning.

5.3. Single Study Area Comparison

The chosen study area was an area about 4.7×4.7 miles located at the center of the Broward County. As shown in Figure 5-1, the study area has a total of 19 evaluation count sites.

Table 5-1 lists the AADTs for all the estimation count sites estimated by the three methods, the ground truth AADTs, and their corresponding estimation errors. The results indicate that the URS method and the proposed method had similar performance for this study area, and they have much lower estimation errors than those of the USF method.

By further checking the AADTs estimated by the proposed method for the evaluation locations, it was found that three locations (sites 2, 6, and 19) have very high estimation errors exceeding 90%. From the locations shown in Figure 5-1, it can be seen that they are located near the boundary of the study area. It is very possible that the results were underestimated for these sites, because they were too close to the

boundaries causing the trips which should have passed them to be excluded. To verify that, those three evaluation count sites were removed. Table 5-2 shows the results, and it indicates that the performance of the proposed method was improved. Therefore, one of the limitations of the proposed method is that it provides less accurate AADT estimation for roads near the boundary areas.

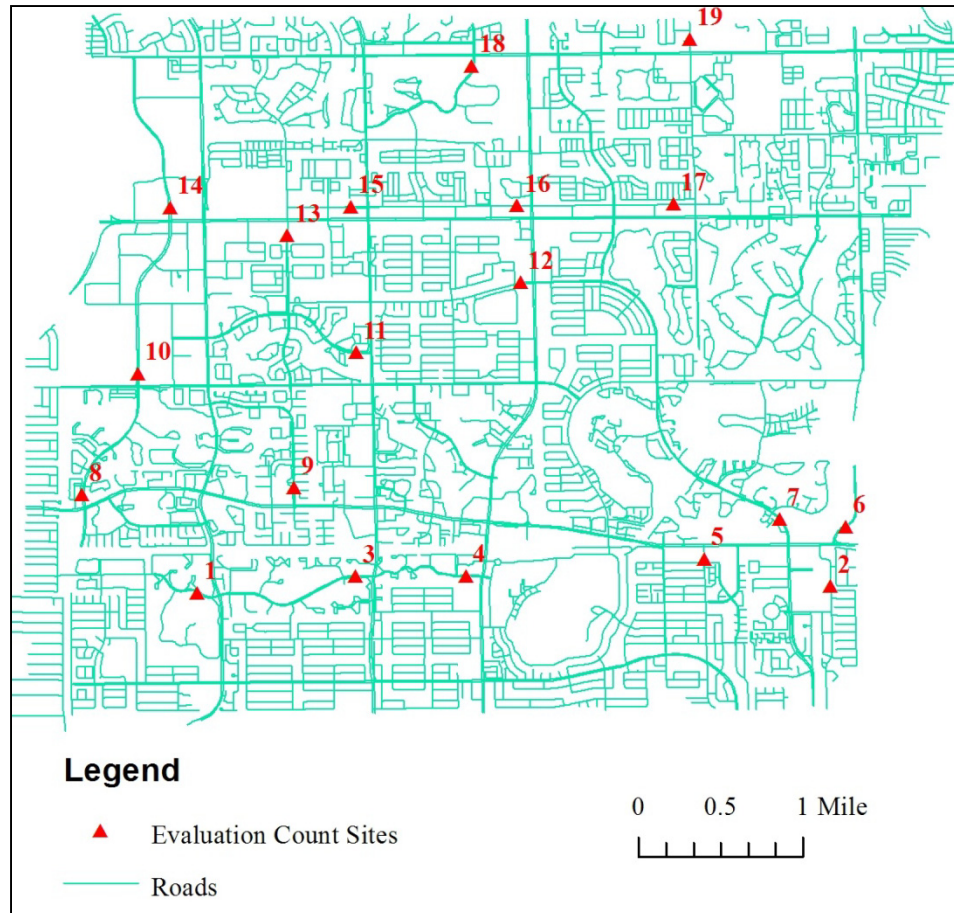


Figure 5-1 Evaluation Traffic Count Sites in the Study Area

Table 5-1 Performance of the Three Methods for the Study Area

Site No.	AADT by Count Data	AADT by USF Method	USF Method MAPE (%)	AADT by URS Method	URS Method MAPE (%)	AADT by Proposed Method	Proposed Method MAPE (%)
1	5,900	20,967	255.37	7,074	19.90	1,385	76.53
2	11,000	18,067	64.25	2,225	79.77	1,001	90.90
3	4,400	2,400	45.45	823	81.30	4,112	6.55
4	3,300	20,967	535.36	1,004	69.58	2,010	39.10
5	7,900	0	100.00	1,936	75.49	7,246	8.28
6	21,500	10,750	50.00	35,014	62.86	0	100.00
7	17,500	9,450	46.00	17,100	2.29	27,165	55.23
8	15,700	8,700	44.59	2,209	85.93	11,014	29.85
9	4,800	17,932	273.58	7,354	53.21	1,011	78.93
10	13,000	6,200	52.31	2,209	83.01	1,375	89.42
11	5,000	18,384	267.68	2,665	46.70	7,080	41.60
12	6,000	3,600	40.00	2,076	65.40	5,689	5.18
13	5,900	17,365	194.32	3,614	38.75	10,338	75.21
14	13,000	2,350	81.92	2,209	83.01	9,451	27.30
15	2,800	0	100.00	328	88.29	2,455	12.32
16	4,000	17,932	348.30	2,955	26.13	1,348	66.29
17	6,100	15,496	154.03	1,540	74.75	1,800	70.49
18	4,800	19,154	299.04	220	95.42	601	87.48
19	17,600	17,500	0.57	18,640	5.91	5	99.97
MAPE (%)			155.41			59.88	55.82

Table 5-2 Performance of the Methods without Invalid Evaluation Count Sites

Site No.	AADT by Count Data	AADT by USF Method	USF Method MAPE (%)	AADT by URS Method	URS Method MAPE (%)	AADT by Proposed Method	Proposed Method MAPE (%)
1	5,900	20,967	255.37	7,074	19.90	1,385	76.53
3	4,400	2,400	45.45	823	81.30	4,112	6.55
4	3,300	20,967	535.36	1,004	69.58	2,010	39.10
5	7,900	0	100.00	1,936	75.49	7,246	8.28
7	17,500	9,450	46.00	17,100	2.29	27,165	55.23
8	15,700	8,700	44.59	2,209	85.93	11,014	29.85
9	4,800	17,932	273.58	7,354	53.21	1,011	78.93
10	13,000	6,200	52.31	2,209	83.01	1,375	89.42
11	5,000	18,384	267.68	2,665	46.70	7,080	41.60
12	6,000	3,600	40.00	2,076	65.40	5,689	5.18
13	5,900	17,365	194.32	3,614	38.75	10,338	75.21
14	13,000	2,350	81.92	2,209	83.01	9,451	27.30
15	2,800	0	100.00	328	88.29	2,455	12.32
16	4,000	17,932	348.30	2,955	26.13	1,348	66.29
17	6,100	15,496	154.03	1,540	74.75	1,800	70.49
18	4,800	19,154	299.04	220	95.42	601	87.48
MAPE (%)			177.37	61.82		48.11	

5.4. Multiple Study Areas Comparison

To measure the change of the three methods' performance with different locations, a total of 10 study areas were selected from Broward County. These locations cover diverse areas and as many evaluation count sites as possible. The roadway layout and the locations of the traffic count sites for evaluation and estimation are illustrated in 10 maps as shown from Figure 5-2 to Figure 5-11 in sequence.

Table 5-3 lists the MAPEs of the three methods for the 10 study areas, and the standard deviation of the MAPEs. From the results, it can be noted that the proposed method has much lower MAPEs than the USF method for all 10 study areas, and it has fairly lower MAPEs than the URS method for 9 study areas. It can also be noted that the proposed method has lower standard deviation for the MAPEs of the 10 study areas than the other two methods, which means that its performance is least affected by the locations and the area types of the study areas.

Table 5-3 Variance Measure of the Performance

Area	USF Method MAPE (%)	URS Method MAPE (%)	Proposed Method MAPE (%)
1	216.67	57.16	48.02
2	314.14	68.72	52.16
3	177.37	61.82	48.11
4	345.49	87.90	66.09
5	175.80	35.82	49.08
6	405.20	77.79	60.08
7	114.43	66.12	62.27
8	186.21	63.32	49.10
9	181.39	65.30	55.67
10	157.22	42.18	39.15
Standard Deviation	94	15	8

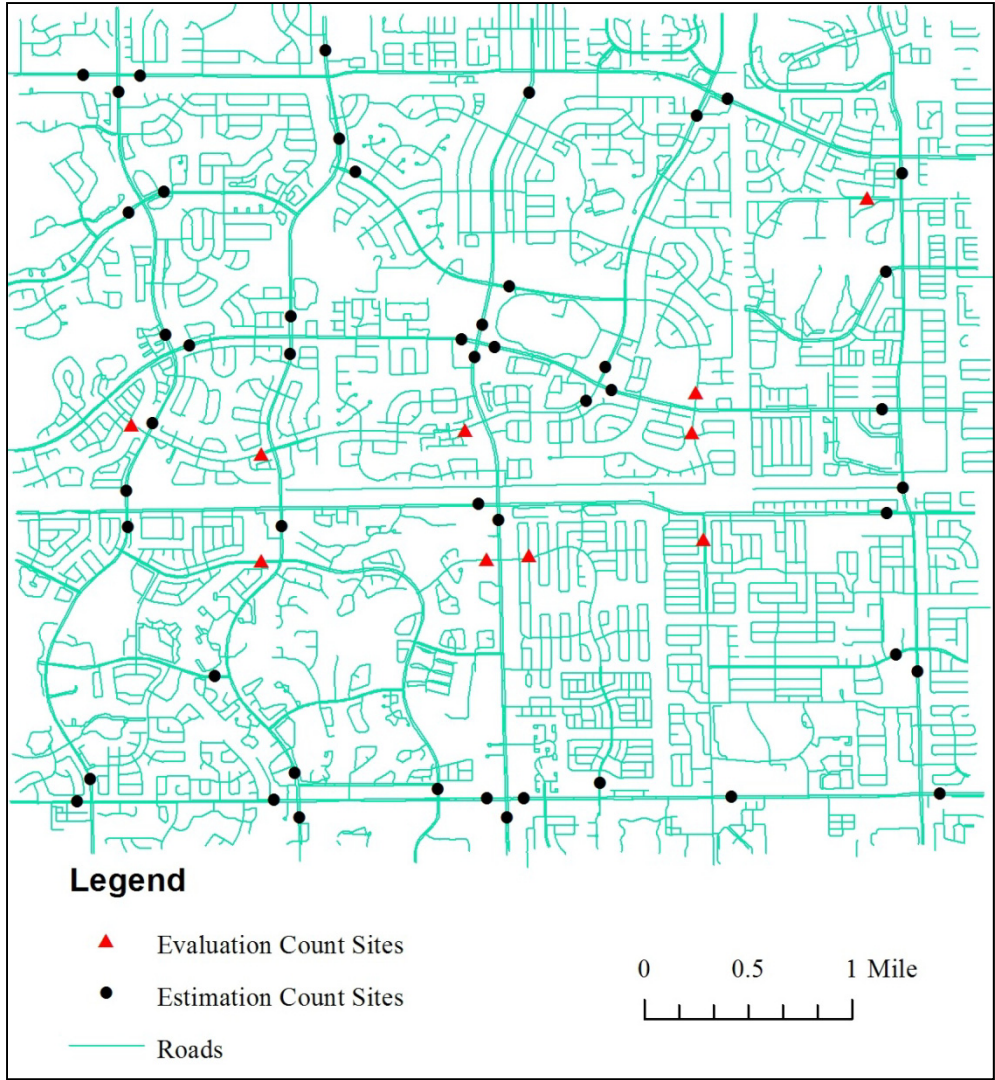


Figure 5-2 Study Area No. 1

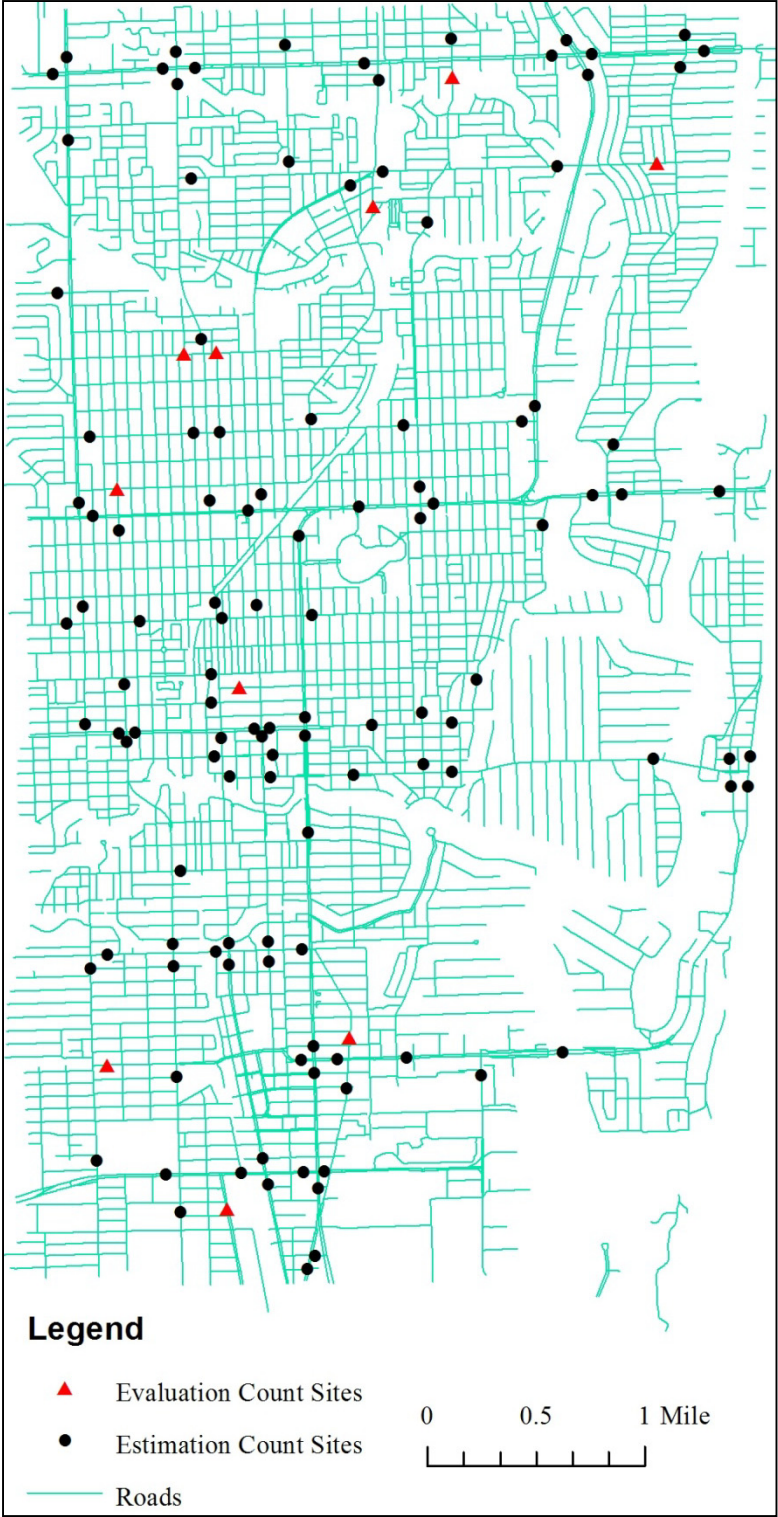


Figure 5-3 Study Area No. 2

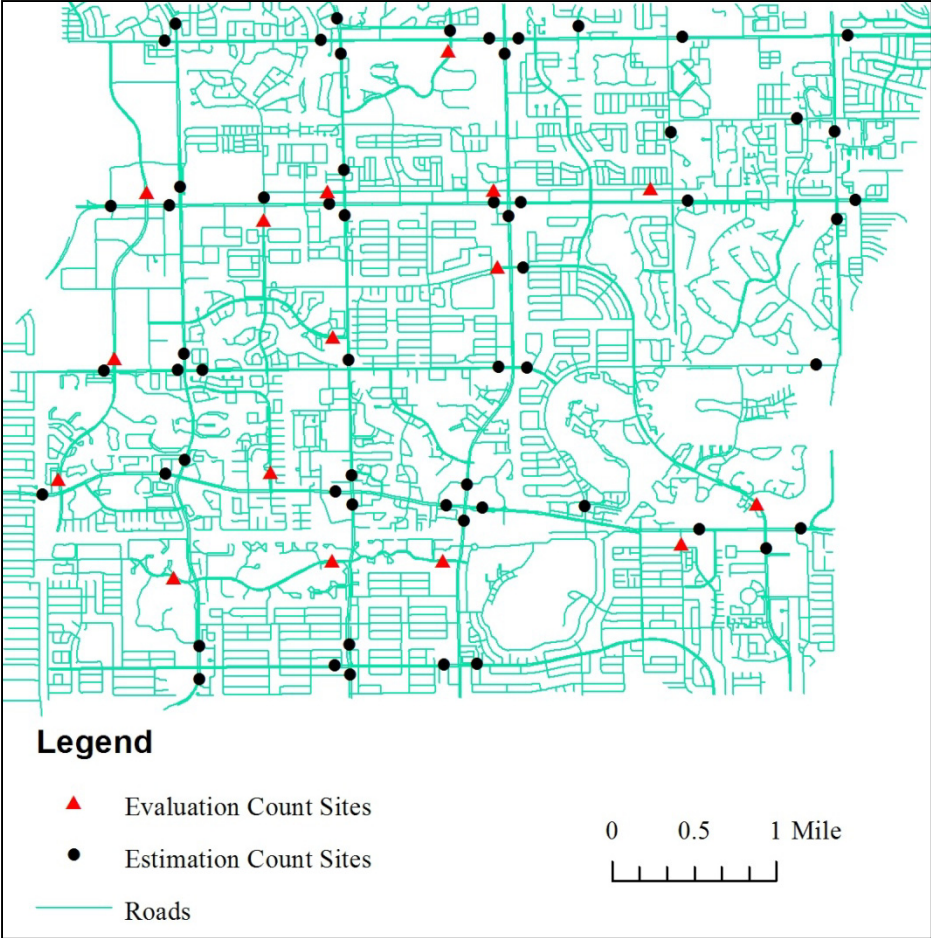


Figure 5-4 Study Area No. 3

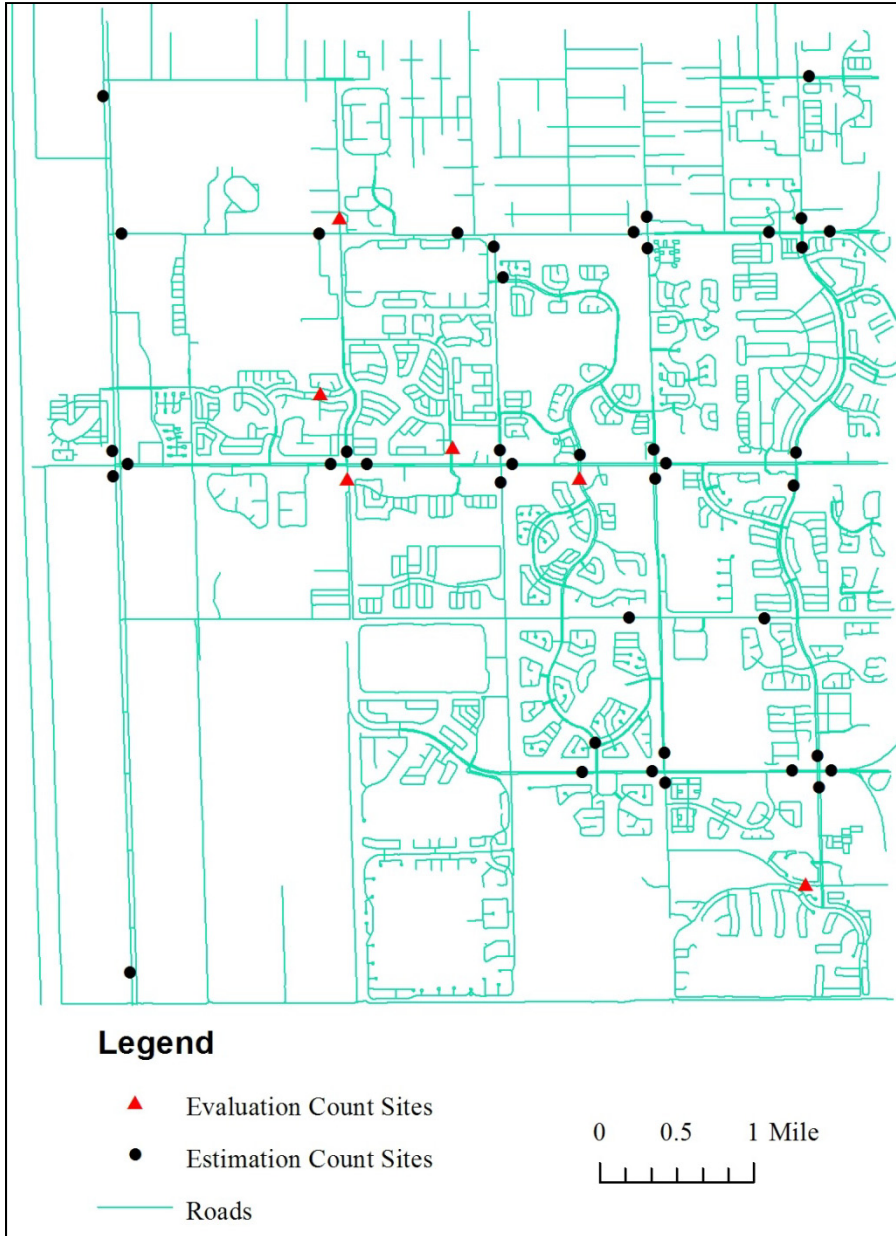


Figure 5-5 Study Area No. 4

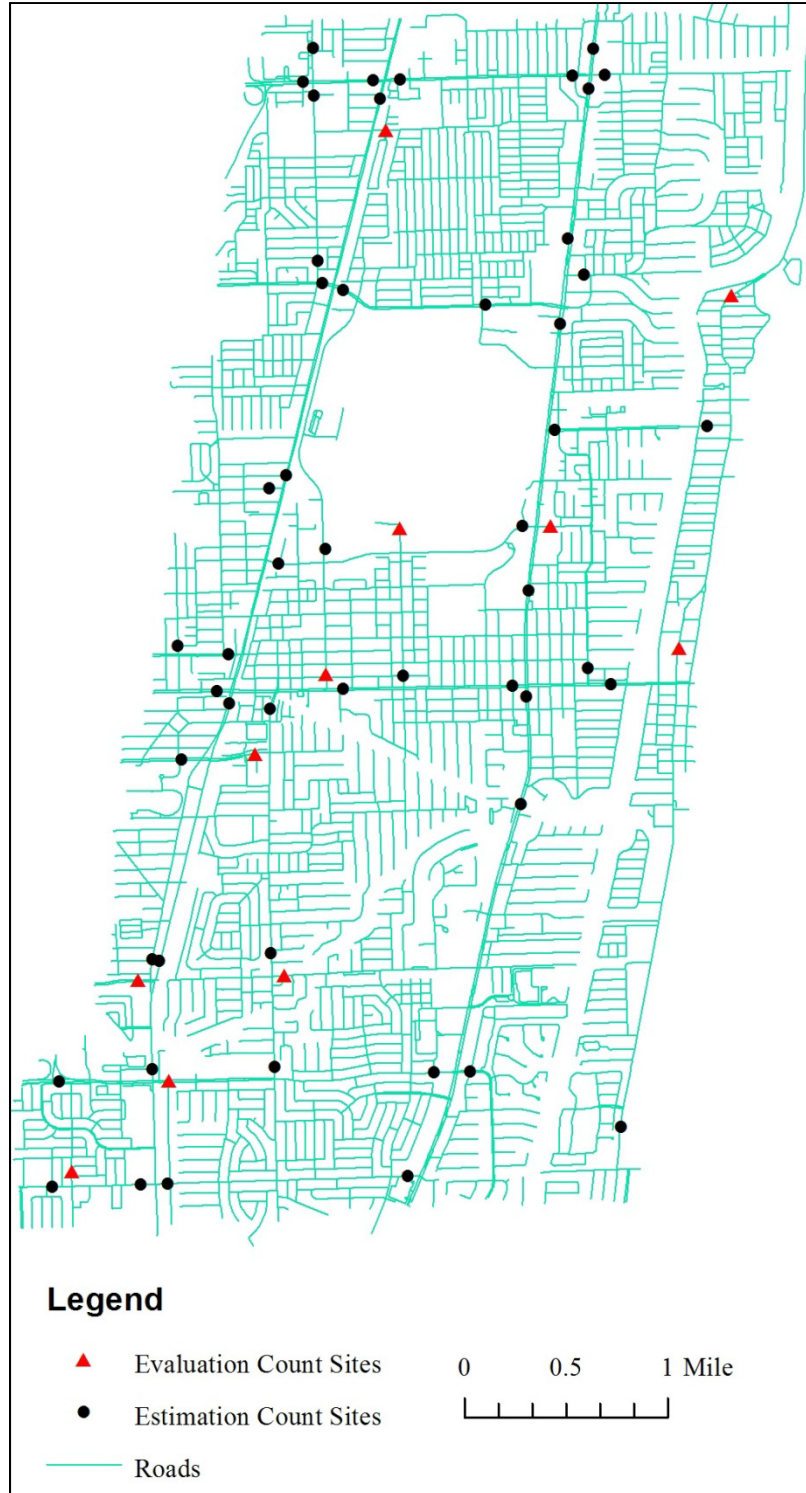


Figure 5-6 Study Area No. 5

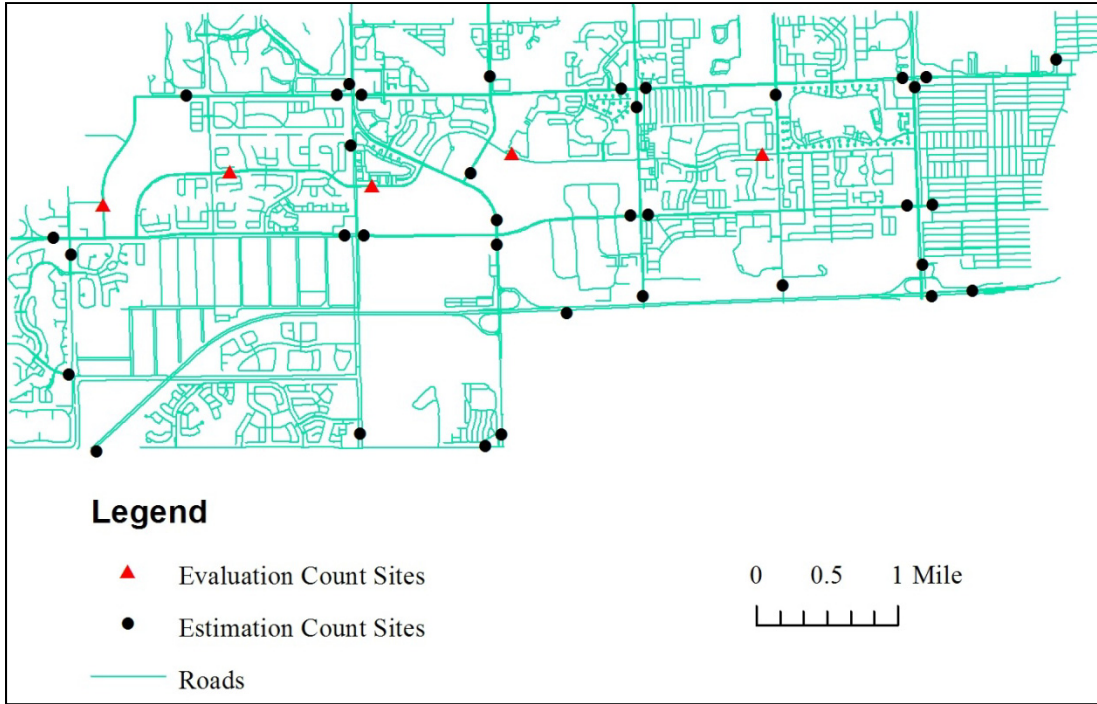


Figure 5-7 Study Area No. 6

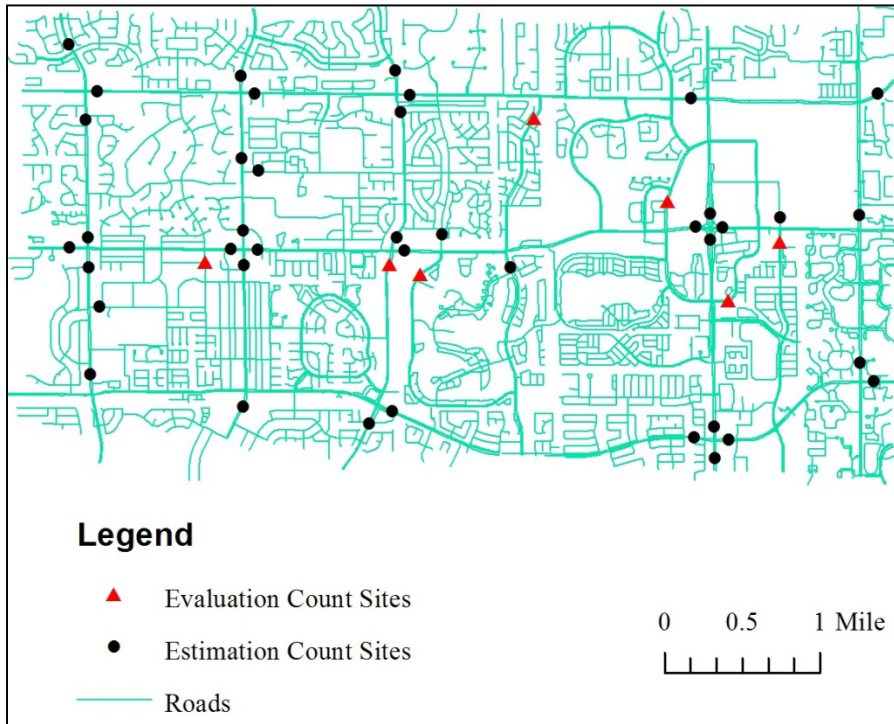


Figure 5-8 Study Area No. 7

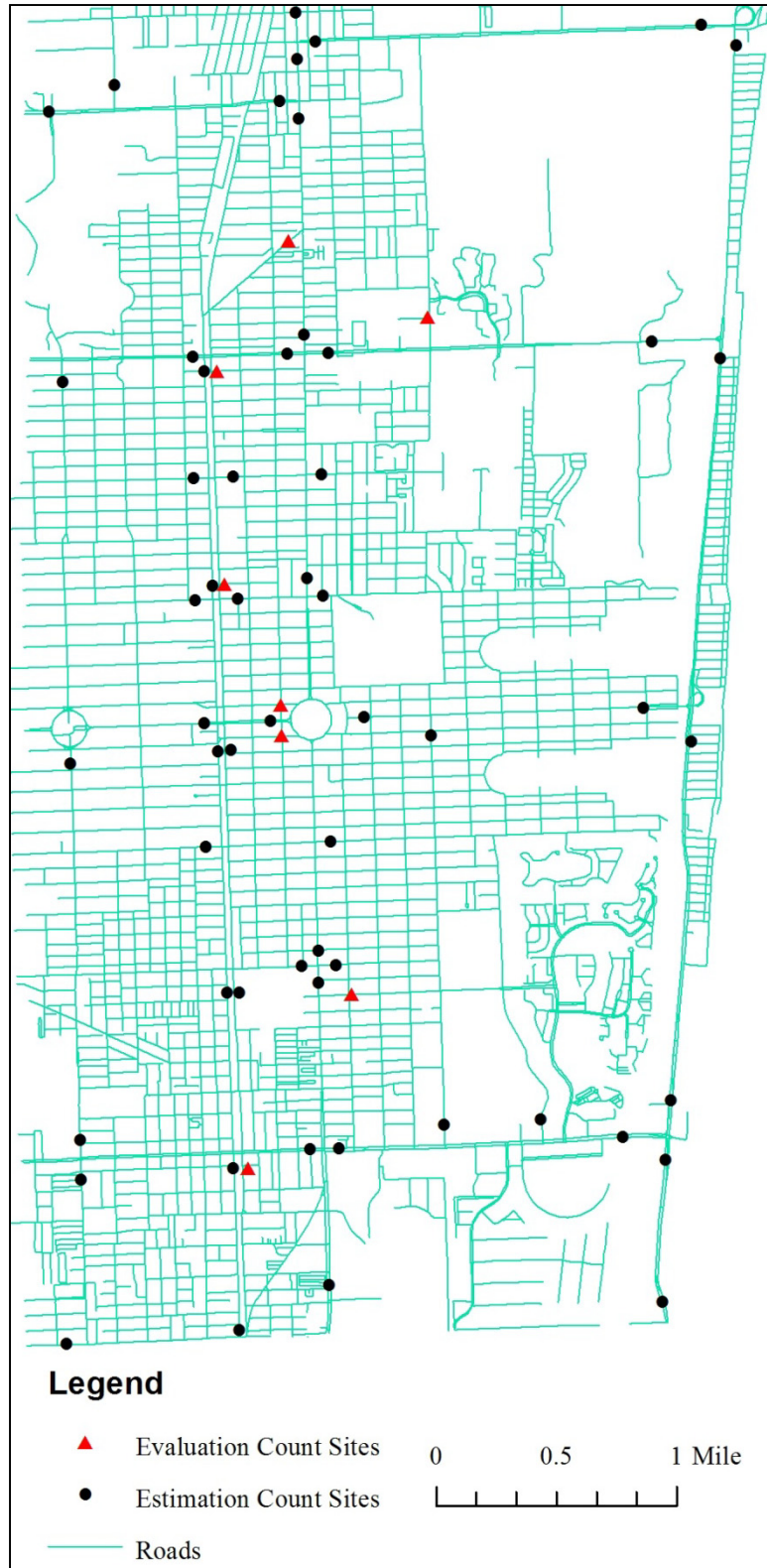


Figure 5-9 Study Area No. 8

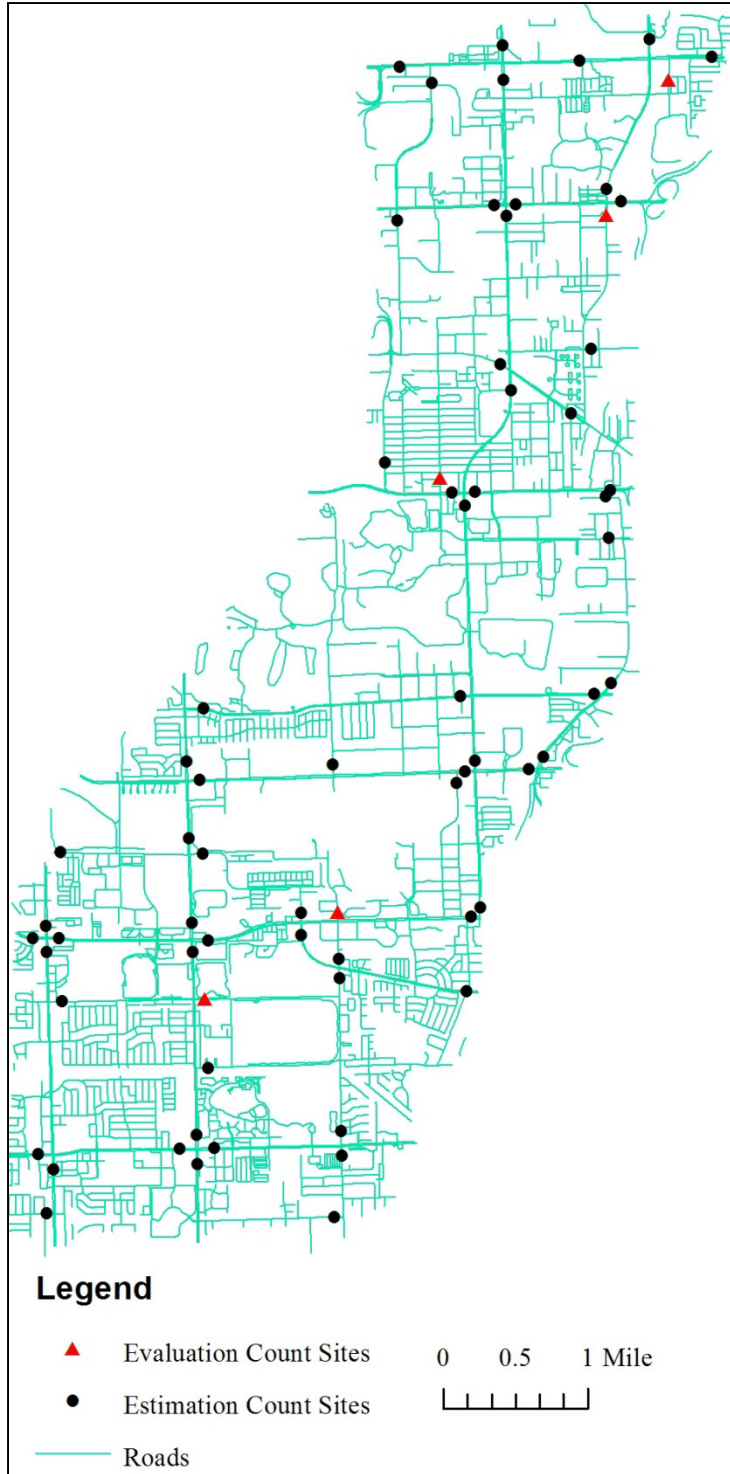


Figure 5-10 Study Area No. 9

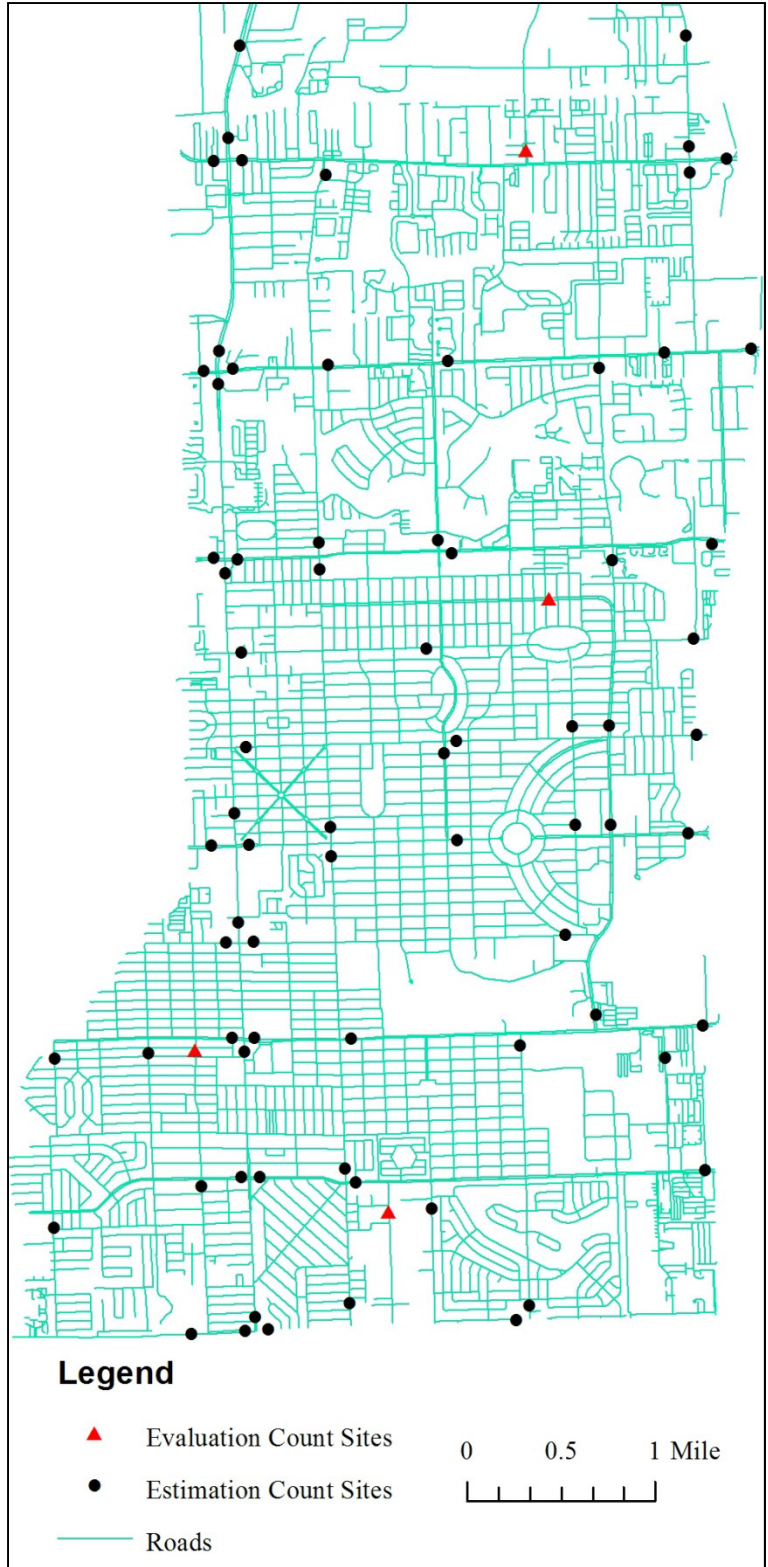


Figure 5-11 Study Area No. 10

5.5. Overall Performance Comparison

The AADT values estimated using the three methods were compared. Figures 5-12, 5-13, and 5-14 compare the ground truth AADTs with the results of the USF method, URS method, and the proposed method, respectively. As expected, the AADT values estimated from the three methods are within a reasonable range (i.e., lower than 30,000 vehicles/day) since all the testing locations were on local roads. Figure 5-12 shows that the USF method overestimates AADT for a greater percentage of evaluation count sites. On the contrary, as shown in Figure 5-13, the URS method underestimates AADT for a greater percentage of evaluation count sites. Figure 5-14 shows that the traffic estimations of the proposed method are more representative of the ground truth data. From the figures, it can be stated that the USF method tend to overestimate while the URS method tend to underestimate the AADT values for local roads.

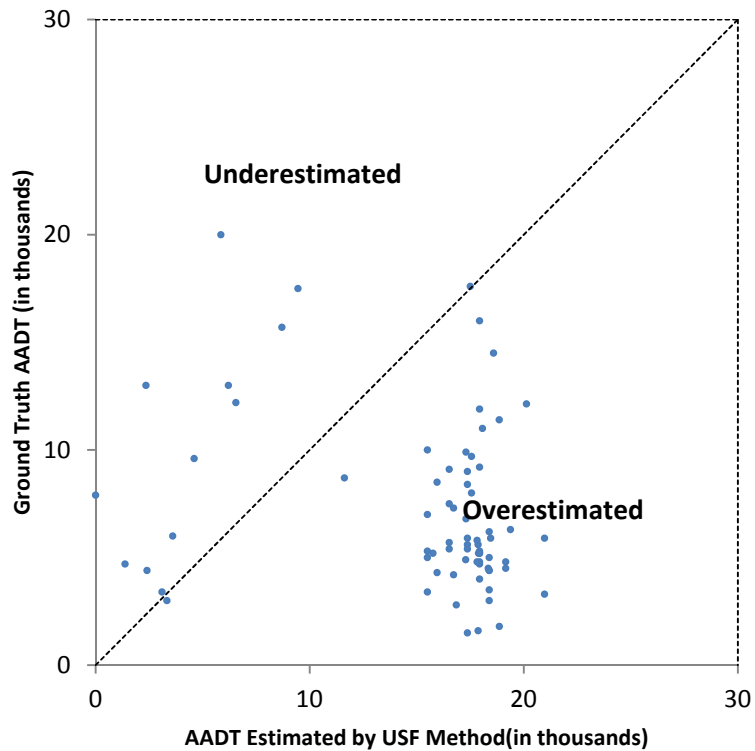


Figure 5-12 Comparison of USF Estimated AADT with Ground Truth AADT

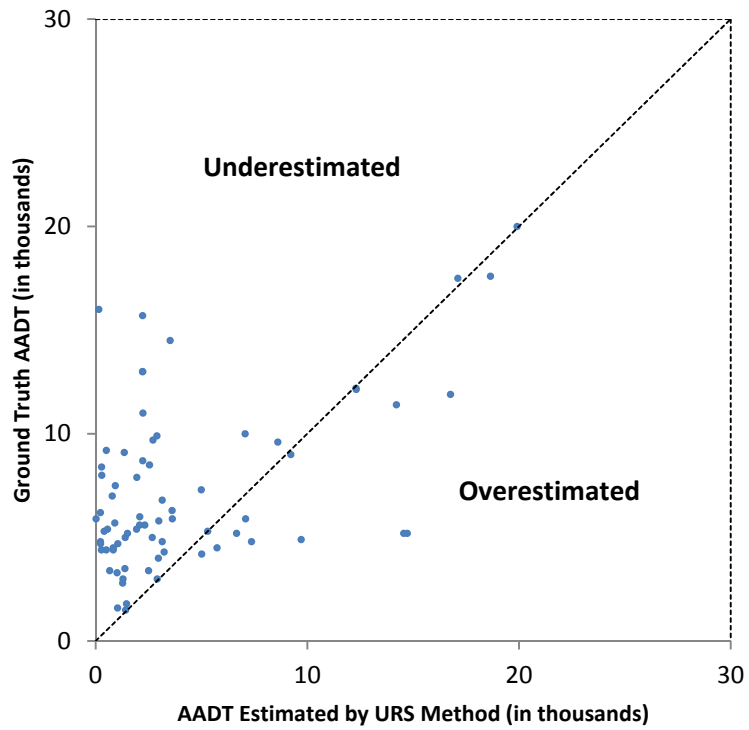


Figure 5-13 Comparison of URS Estimated AADT with Ground Truth AADT

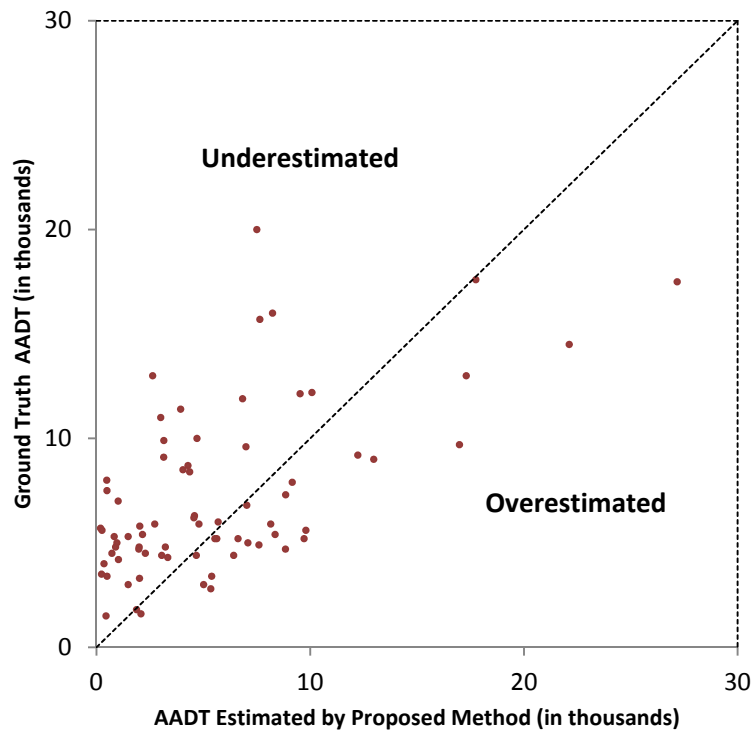


Figure 5-14 Comparison of Proposed Method Estimated with Ground Truth AADT

Table 5-4 compares the accuracy of the three estimation methods using the following three error estimates: Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and Mean Absolute Percentage Error (MAPE). Compared to the USF method, both the proposed method and the URS method have consistently lower estimation errors; the proposed method has an 8% lower MAPE estimation error than the URS method. The results indicate that the proposed method has a better overall performance among the three methods.

Table 5-4 Comparison of Estimation Errors

Errors	USF Method	URS Method	Proposed Method
MAE	10,047	4,124	3,642
RMSE	10,891	5,338	4,484
MAPE	211%	60%	52%

However, it is worth noting that there could be errors in the AADT values adjusted from the raw traffic counts and, hence, the ground truth AADT might not be the “actual” AADT value. Therefore, the results might not accurately reflect the actual difference among the three methods. Nevertheless, to some extent, this evaluation will reflect the advantages of the proposed parcel-level travel demand analysis method since the results are compared to the same ground truth data and the random errors have unbiased influence on the three methods.

5.6. Reasonableness Check

Depending on the availability of traffic count data, most of the traffic count sites used for this evaluation are located on local roads that are directly connected to the state roads. The lower-level local roads such as the community roads were not used in this

evaluation because of the lack of traffic count data. However, the proposed method is expected to perform better even for lower-level local roads as the proposed method's trip generation is based on detailed parcel level data. To verify this assumption, the AADT values estimated using the three methods for the available lower-level local roads were checked and compared. Figure 5-15 gives an example of the comparison. The figure shows the estimation results for the roads in a community of approximately 160 houses.

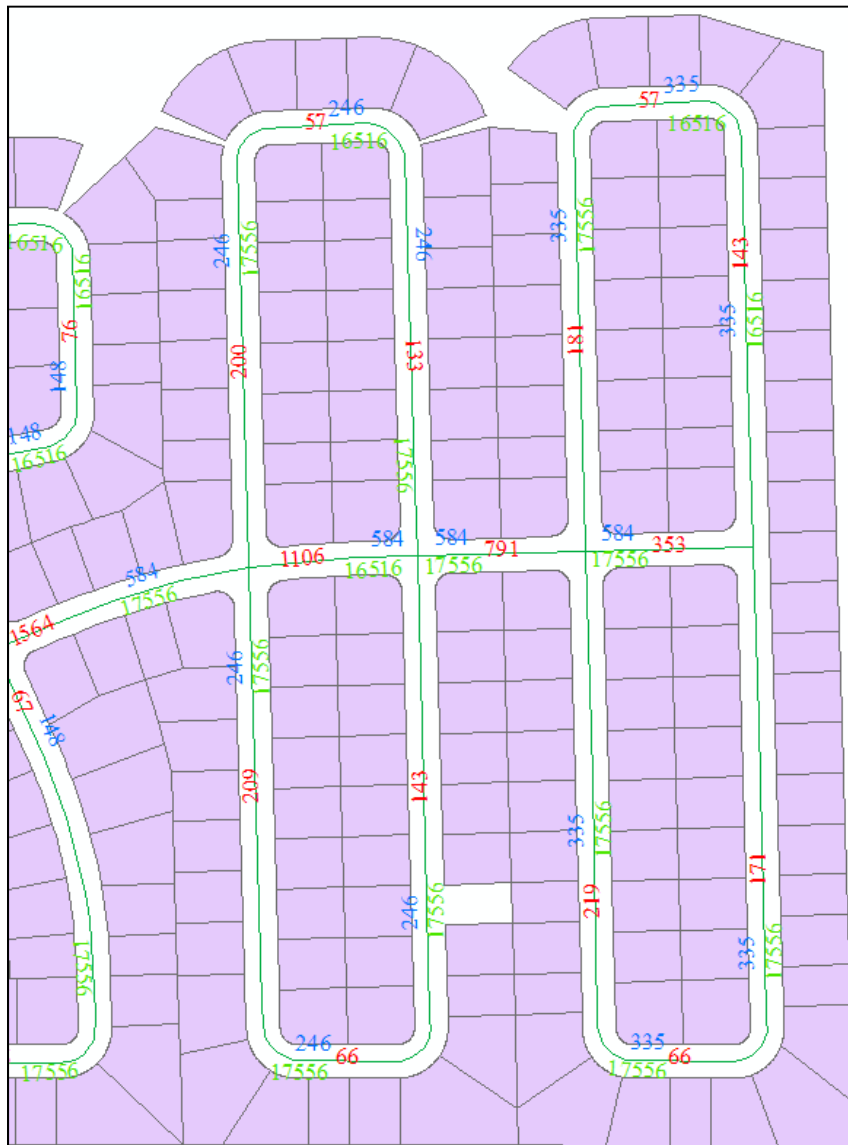


Figure 5-15 Example of AADT Estimation for Roads in a Community

In Figure 5-15, the AADTs estimated by the proposed method, the USF method, and the URS method are displayed in red, green, and blue, respectively. Since there are no traffic count data available for lower-level community roads, the estimated AADT values are compared based on the number of houses and their layout. The AADT values estimated by the USF method were obviously very high and the estimations from the URS method tend to be low for higher-level community roads. In addition, the USF method unrealistically estimated similar AADT values for all the road segments in this community, and to an extent, the URS method performed better with estimating different AADT values. The proposed method provided most accurate and reasonable estimations that are consistent with the layout of the houses.

5.7. Summary

In this chapter, the proposed method was evaluated by comparing it with the USF method and the URS method at different levels. First, the performance of the three estimation methods was compared for a single study area, and the results indicated that the proposed method performs best. The proposed model was found to give more accurate AADT estimations for the central region of the study area compared to the boundaries. Second, ten study areas were selected from Broward County in Florida to compare the sensitivity of the three methods to the change in the study locations and the area types. The ten study areas were chosen based on the availability of sufficient traffic count data. The standard deviations of the estimation errors for these study areas were compared. Compared to the USF and the URS methods, the results showed that the proposed method provides more reliable and stable results when the location of the

study areas and area types are changed. The combined results from the ten study areas also proved that the parcel-level travel demand model method has the best overall performance. Third, the AADT values for lower-level local roads were estimated and it was found that the proposed method performs better for lower-level local roads with no traffic count data. In summary, the evaluation results showed that the parcel-level travel demand method is an accurate AADT estimation method for local roads.

CHAPTER 6

SUMMARY, CONCLUSIONS AND RECOMMENDATIONS

6.1. Summary of Research Approach and Results

A comprehensive literature review was conducted to investigate and assess the existing AADT estimation methods. While all the researched AADT estimation methods have their advantages, it was also found that the methods have limitations in estimating traffic on local roads. The traditional factor approach is reliable, but is not practical to cover all local roads with portable count sites. The most widely researched regression modeling method cannot provide accurate estimations due to the limitation that it cannot capture passer-by trips. The image processing method cannot retrieve and estimate volume accurately because of the sparse and infrequent travel pattern on local roads. Most of the machine learning methods usually improve the traditional factor approach, but they still need to deploy portable count stations to collect short-term traffic count data.

The parcel-level travel demand model method, a new approach to accurately and efficiently estimate AADT for local roads, was researched. The model consists of the following four steps: network modeling, parcel-level trip generation, parcel-level trip distribution, and parcel-level trip assignment. Unlike the traditional travel demand forecasting model, the parcel-level travel demand model was simplified and optimized to estimate AADT values on local roads. In the parcel-level trip generation step, the DOR parcel data and the ITE trip generation rates and equations were used to estimate the trips generated by the parcels. The parcel trips were then distributed to the traffic count sites

on the major roads based on the fact that most travelers choose to access the major roads as soon as possible to minimize travel time. Since traffic congestion rarely occurs on local roads, the simple all-or-nothing trip assignment method was used to assign trips to the local roads. This assignment method has minimized the model running time without compromising with the model's performance.

Cube and ArcGIS were used as the development tools to implement the proposed model. ModelBuilder, a tool provided by ArcGIS, was used to preprocess the data and to calculate the final estimated AADT values. Cube was used to build the network from the highway shape file and automatically add the centroids and centroid connectors to the roadway network. Cube Voyager scripts were developed to implement the four major model steps.

The proposed parcel-level travel demand model method was applied to Broward County in Florida, and the results were compared with the USF method that uses the regression approach, and the AADT estimation method proposed by the URS Corporation. The results of the three methods were compared to the ground truth AADTs estimated from the traffic count data. Among the three methods, the proposed method had the lowest estimation error. The evaluation results showed that the parcel-level travel demand method is a more accurate AADT estimation method with lower estimation errors.

6.2. Conclusions

The major objective of this research was to develop an improved method of estimating AADT for local roads by applying travel demand modeling techniques. A

parcel-level travel demand analysis model was proposed, implemented, and evaluated. Compared to the USF method and the URS method, the proposed method had more accurate AADT estimations for local roads. It can be concluded that the proposed method is a new and practical approach that can provide better AADT estimations for local roads. A summary of the important findings are discussed below:

The DOR parcel data and the *ITE Trip Generation Report* are a valuable data resource to perform the parcel-level trip generation, a critical step for the entire model. Several issues were encountered while matching the DOR parcel data with the land use categories available in the *ITE Trip Generation Report*. This is because the *ITE Trip Generation Report* has more detailed land use types than the DOR parcel data. Additional data may therefore be required to take full advantage of the relation between the two sets of data. As a result of these differences, care should be taken while using the ITE trip generation rates or equations from the *ITE Trip Generation Report*.

Traffic count data for major roads are important to accurately estimate AADT values on the local roads. Traditionally, traffic count data are often collected only for the roads with count sites. In this research, the parcel-level trip distribution model used the traffic count data (or more specifically, AADT estimated from the traffic count data) as the basis to distribute trips generated by the parcels. Traffic volume through the traffic count sites was assumed to be from or to the parcels, and the traffic count sites with high traffic volume was assumed to attract more trips from the parcels. For trip distribution step, several distribution ranges were researched. Smaller distribution ranges that distribute trips from a parcel to the traffic count sites close to and surrounding that parcel were found to generate better results. Further, enough traffic count data

should be available to evenly cover the entire study area. Uneven coverage of the traffic count sites may introduce bias into the trip distribution.

The traditional travel demand forecasting model and the parcel-level travel demand model are quite similar, yet, they have a few major differences. While the traditional model performs the travel demand analysis on an abridged roadway network with only major roads, the proposed model simulates the trips on an unabridged roadway network including the local roads. The traditional mode choice step was omitted in the parcel-level model, since transit trips were not considered on local roads. While the traditional zone-level travel demand forecasting model can be pushed to the limits and applied to the most detailed parcel level, challenges do exist. One of the challenges is the huge number of parcels which have to be preprocessed to improve the model efficiency. The proposed model might not be capable of handling the parcels in a broad study area. One of the solutions could be to divide the study area into multiple subareas to run the model separately. While dividing a study area, care should be taken such that the boundaries intersect as fewer roads as possible. When the boundaries intersect more roads, the model will be less accurate for the boundary area compared to the central region.

Besides the challenge of the huge number of parcels, the proposed model also needs enough traffic count data to evenly cover the major roads in a study area. Insufficient or unevenly covered traffic count data will affect distribution and assignment of trips, which in turn will affect the accuracy of the estimations. Further, to evaluate the model, traffic count data for as many local roads as possible are required.

In spite of the above discussed challenges, it can be concluded that adopting the parcel-level travel demand modeling method to explore the detailed DOR parcel data and the traffic count data is a practical approach to estimate AADTs on local roads.

6.3. Recommendations for Future Research

Even though this dissertation has achieved the proposed research objective, the following areas require further research:

- More demographic and land use data are required to improve the accuracy of the parcel-level trip generation step. In the current parcel-level trip generation, if a parcel land use type encompasses multiple ITE land use types, the model used the average value of the estimated parcel trips based on each ITE land use type. However, this assumption might not represent the actual land use proportions. Even though this type of parcel land use takes a very small portion, the results of the parcel-level trip generation can be improved using more detailed demographic and land use data and estimating parcel trips based on the actual existence of each ITE land use category.
- In the parcel-level trip distribution step, the trips generated by a parcel were distributed to the traffic count sites which can be reached within a specific free flow travel time range. The trip distribution results might be more accurate if the trips can be distributed to the traffic count sites on the boundary roads of the TAZ within which the parcel is located. This approach was attempted, but it was found to be difficult to define the distribution range based on space. Further research is needed to implement this approach.

- As the maximum number of zones supported by Cube is 32,000, the proposed model cannot cover an area with more than 32,000 parcels. One solution for this limitation is to divide the study area into subareas and run the model for each subarea separately. As the accuracy of the estimated trips for the parcels close to the boundaries is affected by the division, cordon lines have to be established by following the higher-level roadways such as freeways and the natural barriers such as canals. This approach will result in intersecting fewer local roads. The procedure and methodology to appropriately divide a broad area and to automatically implement the model for an area with more than 32,000 parcels need further research.
- Cube was used to build the network file from roadway shape file and create centroid connectors. If Cube can provide the programming interface to automate this process, the entire model would be more efficient. Further inquiry and research are required to implement this functionality.

REFERENCES

- AASHTO (1992). *AASHTO Guidelines for Traffic Data Programs*. AASHTO Standing Committee on Planning, Highway Subcommittee on Traffic Engineering, Joint Task Force on Traffic Monitoring Standards.
- Anthony, M. and P. L. Bartlett (1999). *Neural Network Learning: Theoretical Foundations*. Cambridge University Press, Cambridge, UK.
- Berk, A., Richard (2004). *Regression Analysis*. SAGE Publications, Los Angeles, California.
- Cambridge Systematics, Inc. (1994). *Use of Data from Continuous Monitoring Sites, Volume I: Recommendations and Volume II: Documentation*. Final Report Prepared for FHWA.
- Castro-Neto, M., Y. Jeong, M. K. Jeong, and L. D. Han (2009). "AADT Prediction Using Support Vector Regression with Data-dependent Parameters," *Expert Systems with Applications*, Vol. 36, No. 2, pp. 2979-2986.
- Cheng, C. (1992). *Optimum Sampling for Traffic Volume Estimation*. Ph.D. Dissertation, University of Minnesota, Minnesota.
- Cherkassky, V. and Y. Ma (2004). "Practical selection of SVM parameters and noise estimation for SVM regression," *Neural Networks*, Vol. 17, No. 1, pp. 113–126.
- Cook, D. R. and S. Weisberg (1999). *Applied Regression Including Computing and Graphics*. John Wiley, New York.
- Deacon, J. A., J. G. Pigman, and A. Mohenzadeh (1987). *Traffic Volume Estimates and Growth Trends*. UKTRP-8732, Kentucky Transportation Research Program, University of Kentucky, Kentucky.
- Dougherty, M. (1995). "A Review of Neural Networks Applied to Transport," *Transportation Research Part C: Emerging Technologies*, Vol. 3, No. 4, pp. 247-260.
- Drucker, H., C. J. C. Burges, L. Kaufman, A. Smola, and V. Vapnik (1996). "Support Vector Regression Machines," *NIPS's 1996*, Vol. 9, pp. 155-161.
- FHWA (2001). *Traffic Monitoring Guide*. Federal Highway Administration, U.S. Department of Transportation.
- Gurney, K. (2009). *An Introduction to Neural Networks*, UCL Press, London, UK

Himanen, V., P. Nijkamp, and A. Reggiani (1998). *Neural Networks in Transport Applications*. Ashgate Publishing, Farnham, Surrey, UK.

ITE Trip Generation Report (8th ed.). (2008). Institute of Transportation Engineers, Washington, DC.

Jiang, Z. (2005). *Incorporating Image-Based Data in AADT Estimation: Methodology and Numerical Investigation of Increased Accuracy*. Ph.D. Dissertation. Department of Civil Engineering, Ohio State University, Columbus, OH.

Jiang, Z., M. R. McCord, and P. K. Goel (2006). "Improved AADT Estimation by Combining Information in Image- and Ground-Based Traffic Data," *Journal of Transportation Engineering*, Vol. 132, No. 7, pp. 523-530.

Jiang, Z., M. R. McCord, and P. K. Goel (2007). "Empirical Validation of Improved AADT Estimates from Image-Based Information on Coverage Count Segments," *Compendium of Papers, the 86th Annual Meeting of the Transportation Research Board*, National Research Council, Washington, DC.

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

Larose, D. T. (2005). *Discovering Knowledge in Data: An Introduction to Data Mining*. Wiley-Interscience, Hoboken, NJ.

Li, J., C. Xu, and J. D. Fricker (2008). "Comparison of Annual Average Daily Traffic Estimates: Traditional Factor, Statistical, Artificial Neural Network, and Fuzzy Basis Neural Network Approach," *Compendium of Papers, the 87th Annual Meeting of the Transportation Research Board*, National Research Council, Washington, DC.

Li, J. and J. D. Fricker (2008). "Applying K-Nearest Neighbor Algorithm for Statewide Annual Average Daily Traffic Estimates," *Compendium of Papers, the 87th Annual Meeting of the Transportation Research Board*, National Research Council, Washington, DC.

Lu, J., T. Pan, and P. Liu (2007). *Assignment of Estimated Average Annual Daily Traffic Volumes on All Roads in Florida*. Final Report Prepared for Florida Department of Transportation, Tallahassee, FL.

McCord, M. R., C. J. Merry, X. D. Sun, and F. Jafar (1995a). "Resolution Effects on Vehicle Counts and Classification through Remote Sensing," *Journal of the Transportation Research Forum*, Vol. 35, No. 1, pp. 41-52.

McCord, M. R., C. J. Merry, and J. D. Bossler (1995b). *The Feasibility of Traffic Data Collection using Satellite Imagery*. Final Report Prepared for Federal Highway Administration, Washington, DC, and Ohio Department of Transportation, Columbus, OH, ODOT Agreement No. 6952.

McCord, M. R., Y. Yang, Z. Jiang, B. Coifman, and P. Goel (2003). "Estimating AADT from Satellite Imagery and Air Photographs: Empirical Results," *Transportation Research Record: Journal of the Transportation Research Board*, No. 1855, pp. 136-142.

Michalski, R. S., I. Bratko, and M. Kubat (1998). *Machine Learning and Data Mining Methods and Applications*. John Wiley & Sons, Inc., Hoboken, New Jersey.

Mitchell, T. (1997). *Machine Learning*. McGraw Hill, Columbus, OH.

Mohamad, D., K.C. Sinha, T. Kuczek, and C.F. Scholer (1998). "An Annual Average Daily Traffic Prediction Model for County Roads," *Transportation Research Record 1617*, Transportation Research Board, National Research Council, Washington, D.C., pp. 69-77.

Oppenheim, N. (1995). *Urban Travel Demand Modeling: From Individual Choices to General Equilibrium*. Wiley-Interscience, Hoboken, NJ.

Park, N. (2004). *Estimation of Average Annual Daily Traffic (AADT) Using Geographically Weighted Regression (GWR) Method and Geographic Information System (GIS)*. Ph.D. Dissertation, Florida International University, Miami, FL.

Roess, R. P., E. S. Prassas, and W. R. McShane (2004). *Traffic Engineering 3rd Edition*. Prentice Hall, Upper Saddle River, NJ.

SafetyAnalyst. <http://www.safetyanalyst.org>. Access date: March, 2011.

Sharma, S. C., P. Lingras, F. Xu, and G. X. Liu (1999). "Neural Networks as Alternative to Traditional Factor Approach of Annual Average Daily Traffic Estimation from Traffic Counts," *Transportation Research Record 1660*, Transportation Research Board, National Research Council, Washington, D.C., pp. 24-31.

Sharma, S. C., P. Lingras, G. X. Liu, and F. Xu (2000). "Application of Neural Networks to Estimate AADT on Low Volume Roads," *Transportation Research Record 1719*, Transportation Research Board, National Research Council, Washington, D.C., pp. 103-111.

Sharma, S. C., P. Lingras, F. Xu, and P. Kilburn (2001). "Application of Neural Networks to Estimate AADT on Low-volume Roads," *Journal of Transportation Engineering-ASCE*, Vol. 127, No. 5, pp. 426-432.

Sheffi, Y. (1985). *Urban Transportation Networks: Equilibrium Analysis with Mathematical Programming Methods*. Prentice-Hall, Inc., Englewood Cliffs, NJ.

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

Shon, D. (1989). *Traffic Volume Forecasting for Rural Alabama State Highways*. Thesis, Auburn University, Alabama.

Smola, A. J. and B. Schölkopf (1998). "A Tutorial on Support Vector Regression," NeuroCOLT, *Technical Report NC-TR-98-030*, Royal Holloway College, University of London, UK.

Sossiau, A. B., A. B. Hassam, M. M. Carter, and G. V. Wickstrom (1978). *Quick-response Urban Travel Estimation Techniques and Transferable Parameters. User's Guide*. NCHRP Report 187, TRB, National Research Council, Washington D.C.

Wikibooks, http://en.wikibooks.org/wiki/Artificial_Neural_Networks/Print_Version. Access date: March, 2011.

Xia, Q., F. Zhao, L.D Shen, and D. Ospina (1999). *Estimation of Annual Average Daily Traffic for Non-State Roads in a Florida County*. Research Report, Florida Department of Transportation, Tallahassee, FL.

Vapnik, V. (1995). *the Nature of Statistical Learning Theory*. Springer-Verlag, New York, NY.

Zhao, F. and S. Chung (1999). "Estimation of Annual Average Daily Traffic in a Florida County Using GIS and Regression," *Transportation Research Record 1660*, Transportation Research Board, National Research Council, Washington, D.C., pp. 32-40.

Zhong, M. and B. L. Hanson (2009). "GIS based Travel Demand Modeling for Estimating Traffic on Low-class Roads," *Transportation Planning and Technology*, Vol. 32, No. 5, pp. 423-439.

APPENDIX A

MATCHING RESULTS FOR DOR PARCEL DATA AND ITE TRIP

GENERATION REPORT

This appendix presents the matching of land use types of DOR parcel data and *ITE Trip Generation Report*, the selected independent variables, and the selection of the estimation method (average rate or regression equation) for each land use types.

Table A-1 Land Use Type Matching, Independent Variables, and Rate/Equation

Parcel Code	Parcel Land Use	ITE Code	ITE Land Use	Independent Variable Used	Average Rate / Equation
000	Vacant Residential		N/A		
001	Single Family	210	Single-Family Detached Housing	Dwelling Unit	Average Rate
002	Mobile Homes	240	Mobile Home Park	Dwelling Unit	Average Rate
003	Multi-family	220	Apartment	Dwelling Unit	Average Rate
004	Condominiums	230	Residential Condominium/ Townhouse	Dwelling Unit	Average Rate
005	Cooperatives	265	Timeshare	Dwelling Units	Average Rate
006	Retirement Homes	255	Continuing Care Retirement Community	Occupied Units	Average Rate
007	Boarding Homes (Institutional)	254	Assisted Living	Occupied Beds	Average Rate
008	Multi-family less than 10 units	220	Apartment	Dwelling Units	Average Rate
009	Undefined reserved for DOR		N/A		
010	Vacant Commercial		N/A		
011	Stores One-Story	850	Supermarket	Gross Floor Area	Average Rate
012	Mixed Use, i.e., Store and Office	710	General Office Building	Gross Floor Area	Average Rate

Parcel Code	Parcel Land Use	ITE Code	ITE Land Use	Independent Variable Used	Average Rate / Equation
013	Department Stores	875	Department Store	Gross Floor Area	Average Rate
014	Department Stores	875	Department Store	Gross Floor Area	Average Rate
015	Regional Shopping Malls	820	Shopping Center	Gross Leasable Area	Average Rate
016	Community Shopping Centers	820	Shopping Center	Gross Leasable Area	Average Rate
017	One-Story Non-Professional Offices	710	General Office Building	Gross Floor Area	Average Rate
018	Multi-Story Non-Professional Offices	710	General Office Building	Gross Floor Area	Average Rate
019	Professional Service Buildings	710	General Office Building	Gross Floor Area	Average Rate
020	Airports, Marinas, Bus Terminals, and Piers	010	Waterport/Marine Terminal	Acres	Average Rate
		090	Park-and-ride Lot with Bus Service	Acres	Average Rate
		420	Marina	Acres	Average Rate
021	Restaurants, Cafeterias	931	Quality Restaurant	Gross Floor Area	Average Rate
		932	High-Turnover(Sit-Down) Restaurant	Gross Floor Area	Average Rate
		933	Fast-Food Restaurant without Drive-Through Window	Gross Floor Area	Average Rate
		934	Fast-Food Restaurant with Drive-Through Window	Gross Floor Area	Average Rate
		937	Coffee/Donut Shop with Drive-Through Window	Gross Floor Area	Average Rate
022	Drive-in Restaurants	932	High-Turnover(Sit-Down) Restaurant	Gross Floor Area	Average Rate
023	Financial Institutions	912	Drive-in Bank	Gross Floor Area	Average Rate
024	Insurance Company Offices	710	General Office Building	Gross Floor Area	Average Rate

Parcel Code	Parcel Land Use	ITE Code	ITE Land Use	Independent Variable Used	Average Rate / Equation
025	Repair Service Shops	814	Specialty Retail Center	Gross Leasable Area	Average Rate
026	Service Stations	853	Convenience Market with Gasoline Pumps	Gross Floor Area	Average Rate
027	Automotive Repair, Service, and Sales	843	Automobile Parts Sale	Gross Floor Area	Average Rate
028	Parking Lots, Mobile Home Sales	814	Specialty Retail Center	Gross Leasable Area	Average Rate
029	Wholesale, Manufacturing, and Produce Outlets	823	Factory Outlet Center	Gross Floor Area	Average Rate
030	Florist, Greenhouses	814	Specialty Retail Center	Gross Leasable Area	Average Rate
031	Drive-in Theaters, Open Stadiums	443	Movie Theater without Matinee	Gross Floor Area	Average Rate
032	Enclosed Theaters, Auditoriums	443	Movie Theater without Matinee	Gross Floor Area	Average Rate
033	Night Clubs, Bars, and Cocktail Lounges	435	Multipurpose Recreational Facility	Acres	Average Rate
034	Bowling Alleys, Skating Rings, Enclosed Arenas	435	Multipurpose Recreational Facility	Acres	Average Rate
035	Tourist Attractions	415	Beach Park	Acres	Average Rate
036	Camps		N/A		
037	Race Horse, Auto, and Dog Tracks	435	Multipurpose Recreational Facility	Acres	Average Rate
038	Golf Courses	430	Golf Course	Acres	Average Rate
039	Hotels, Motels	310	Hotel	Rooms	Average Rate
		320	Motel	Rooms	Average Rate
040	Vacant Industrial		N/A		
041	Light Manufacturing	110	General Light Industrial	Acres	Average Rate
042	Heavy Manufacturing	120	General Heavy Industrial	Acres	Average Rate
043	Lumber Yards, Sawmills, Planning Mills,	812	Building Materials and Lumber Store	Gross Floor Area	Average Rate

Parcel Code	Parcel Land Use	ITE Code	ITE Land Use	Independent Variable Used	Average Rate / Equation
044	Fruit, Vegetables, and Meat Packing	110	General Light Industrial	Acres	Average Rate
045	Canneries, Distilleries, and Wineries	110	General Light Industrial	Acres	Average Rate
046	Other Food Processing	110	General Light Industrial	Acres	Average Rate
047	Mineral Processing	120	General Heavy Industrial	Acres	Average Rate
048	Warehouses, and Distribution Centers	150	Warehousing	Acres	Average Rate
049	Industrial Storage (Fuel, Equip, and Material)	110	General Light Industrial	Acres	Average Rate
050	Improved Agriculture		N/A		
051	Cropland Soil Class 1		N/A		
052	Cropland Soil Class 2		N/A		
053	Cropland Soil Class 3		N/A		
054	Timberland		N/A		
055	Timberland		N/A		
056	Timberland		N/A		
057	Timberland		N/A		
058	Timberland		N/A		
059	Timberland		N/A		
060	Grazing Land Soil Class 1		N/A		
061	Grazing Land Soil Class 2		N/A		
062	Grazing Land Soil Class 3		N/A		
063	Grazing Land Soil Class 4		N/A		
064	Grazing Land Soil Class 5		N/A		
065	Grazing Land Soil Class 6		N/A		
066	Orchard, Groves, Citrus		N/A		

Parcel Code	Parcel Land Use	ITE Code	ITE Land Use	Independent Variable Used	Average Rate / Equation
067	Poultry, Bees, Tropical Fish, Rabbits, etc.		N/A		
068	Dairies, Feed Lots		N/A		
069	Ornamentals, Misc. Agriculture		N/A		
070	Vacant Institutional		N/A		
071	Churches	560	Church	Gross Floor Area	Average Rate
072	Private Schools	520	Elementary School	Gross Floor Area	Equation
		522	Middle School/Junior High School	Gross Floor Area	Average Rate
		530	High School	Gross Floor Area	Average Rate
073	Private Hospitals	610	Hospital	Gross Floor Area	Equation
074	Homes for Aged	251	Senior Adult Housing - Detached	Dwelling Units	Average Rate
		252	Senior Adult Housing - Attached	Occupied Dwelling Units	Average Rate
075	Orphanages		N/A		
076	Mortuaries, Cemeteries	566	Cemetery	Acres	Average Rate
077	Clubs, Lodges, and Union Halls	435	Multipurpose Recreational Facility	Acres	Average Rate
078	Sanitariums, Convalescent, and Best Homes	253	Congregate Care Facility	Dwelling Units	Average Rate
079	Cultural Organizations	590	Library	Gross Floor Area	Average Rate and Equation
080	Undefined		N/A		
081	Military		N/A		
082	Forest, Park, and Recreational Areas	411	City Park	Acres	Average Rate
		412	County Park	Acres	Average Rate

Parcel Code	Parcel Land Use	ITE Code	ITE Land Use	Independent Variable Used	Average Rate / Equation
		413	State Park	Acres	Average Rate
		415	Beach Park	Acres	Average Rate
		417	Regional Park	Acres	Average Rate
083	Public Schools	520	Elementary School	Gross Floor Area	Equation
		522	Middle School/Junior High School	Gross Floor Area	Average Rate
		530	High School	Gross Floor Area	Average Rate
084	Colleges	540	Junior/Community College	Gross Floor Area	Average Rate
085	Public Hospitals	610	Hospital	Gross Floor Area	Average Rate
086	Other Counties		N/A		
087	Other State		N/A		
088	Other Federal		N/A		
089	Other Municipal		N/A		
090	Gov. Owned Leased by Non-Gov. Lessee		N/A		
091	Utilities	170	Utilities	Gross Floor Area	Average Rate
092	Mining, Petroleum, and Gas Lands		N/A		
093	Subsurface Rights		N/A		
094	Rights-of-Way Streets, Roads, and Canals		N/A		
095	Rivers, Lakes, and Submerged Lands		N/A		
096	Sewage Disposal, Borrow Pits, and Wetlands		N/A		
097	Outdoor Recreational		N/A		
098	Centrally Assessed		N/A		
099	Acreage not Zoned for Agricultural		N/A		

VITA

TAO WANG

EDUCATION

- 1997 – 2001 B.S., Computer Science
Northeastern University, Shenyang, China
- 2003 – 2006 M.S., Computer Software and Theory
Northeastern University, Shenyang, China
- 2007 – 2012 Doctoral Candidate in Civil Engineering
Florida International University, Miami, Florida

EMPLOYMENT

- 2001 – 2003 Software Engineer
Fujitsu Nanda Software Technology Company, Nanjing, China
- 2007– 2012 Graduate Research Assistant
Lehman Center for Transportation Research
Florida International University, Miami, Florida

AFFILIATIONS

- Student member, Chinese Overseas Transportation Association, 2008– Present
- Student member, Chi Epsilon Honor Society, 2008 – Present
- Student member, Institute of Transportation Engineers, 2007 – Present

PUBLICATIONS

- Gan, A., K. Haleem, P. Alluri, J. Y. Liu, T. Wang, M. Ma, and C. Diaz, “Preparing Florida for Deployment of *SafetyAnalyst* for All Roads,” Final Report, Prepared for Florida DOT, May 2012.
- Gan, A., K. Y. Liu, T. Wang, and M. Hadi, “Design and Implementation of an Efficient Master Geodatabase System for Maintaining Transportation Modeling Networks,” Proceedings of the 91st Annual Meeting of the Transportation Research Board, National Research Council, Washington, DC., January 2012.
- Wang, T., A. Gan, and M. Ma, “FSUTMS Standard Reports,” Program Documentation, Prepared for Florida DOT, June 2010.
- Gan, A., K. Y. Liu, and T. Wang, “Development of a Data Framework for the Florida Standard Urban Transportation Modeling Structure (FSUTMS),” Final Report, Prepared for Florida DOT, June 2010.
- Wang, T., “Road Ranger Contractor Inspection Program (RRCIP) User’s Guide,” Prepared for Florida DOT District 6, January 2008.

- Xu, C. and T. Wang, “Road Ranger Inspection Program (RRIP) User’s Guide,” Prepared for Florida DOT District 6, August 2007.
- Wang, T., “Research of Reinforcement Learning in Computer Chinese Chess,” M.S. Thesis, January 2006.
- Wang, J., T. Wang, Y.H. Luo, and X.H. Xu, “Implementation of Adaptive Genetic Algorithm of Evaluation Function in Chinese Chess Computer Game System,” *Journal of Northeastern University (Natural Science)*, Vol. 26, No. 10, pp. 949 – 952, October 2005.

HONORS AND AWARDS

- Best Essay Gold Coast ITE Chapter Spring 2009 Book Scholarship Award (Second Place), 2009
- International ITE Best Student Chapter Award while serving as ITE FIU Student Chapter Membership Chairperson, 2008
- Outstanding Leadership Award while serving as ITE FIU Student Chapter Membership Chairperson, 2008
- Third-Place Award of the 10th Computer Olympiad (Computer Chinese Chess Group), 2005
- Senior Level Software Engineer Certified by the Ministry of Industry and Information Technology of China