11-21-2007

A Static Traffic Assignment Model Combined with an Artificial Neural Network Delay Model

Zhen Ding
Florida International University, zhen.ding@gmail.com

DOI: 10.25148/etd.FI08081515

FLORIDA INTERNATIONAL UNIVERSITY

Miami, Florida

A STATIC TRAFFIC ASSIGNMENT MODEL COMBINED WITH AN ARTIFICIAL NEURAL NETWORK DELAY MODEL

A dissertation submitted in partial fulfillment of the

requirements for the degree of

DOCTOR OF PHILOSOPHY

in

CIVIL ENGINEERING

by

Zhen Ding

2007

To: Interim Dean Amir Mirmiran
College of Engineering and Computing

This dissertation, written by Zhen Ding, and entitled a Static Traffic Assignment Model Combined with an Artificial Neural Network Delay Model, 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.

_____
L. David Shen

_____
Albert Gan

_____
Mohammed Hadi

_____
Lee-Fang Chow

_____
Ronald Giachetti

_____
Fang Zhao, Major Professor

Date of Defense: November 21, 2007

The dissertation of Zhen Ding is approved.

_____
Interim Dean Amir Mirmiran
College of Engineering and Computing

_____
Dean George Walker
University Graduate School

Florida International University, 2007

DEDICATION

I dedicate this dissertation to my parents for their everlasting love and encouragement, and to my dear wife for her unconditional support and unlimited patience. This is not only my achievement but theirs as well.

ACKNOWLEDGMENTS

There are many people who helped me complete this doctoral dissertation. I want to first thank my academic advisor, Dr. Fang Zhao, who welcomed me to the Civil Engineering program at FIU, and patiently guided me through the dissertation process. I am deeply indebted to her constructive academic suggestions, years of mentorship and encouragement, as well as her consistent financial support. She dedicated many weekends and holidays to supervise my dissertation. Her wisdom and effort often motivated me to work harder and to achieve more.

I would also like to thank Dr. Albert Gan for his advice on my study, as well as my dissertation. As a member of the FIU Institute of Transportation Engineers chapter, for which Dr. Gan is the faculty advisor, I have enjoyed and benefited greatly from many of its activities. I am also grateful to Dr. Mohammed Hadi for his challenging questions and insightful suggestions that have been very helpful.

Finally, I would like to thank all the members of my dissertation committee, including Dr. L. David Shen, Dr. Lee-Fang Chow, and Dr. Ronald Giachetti. They have been generous to provide me with support and suggestions.

ABSTRACT OF THE DISSERTATION

A STATIC TRAFFIC ASSIGNMENT MODEL COMBINED WITH AN ARTIFICIAL NEURAL NETWORK DELAY MODEL

by

Zhen Ding

Florida International University, 2007

Miami, Florida

Professor Fang Zhao, Major Professor

As traffic congestion continues to worsen in large urban areas, solutions are urgently sought. However, transportation planning models, which estimate traffic volumes on transportation network links, are often unable to realistically consider travel time delays at intersections. Introducing signal controls in models often result in significant and unstable changes in network attributes, which, in turn, leads to instability of models. Ignoring the effect of delays at intersections makes the model output inaccurate and unable to predict travel time. To represent traffic conditions in a network more accurately, planning models should be capable of arriving at a network solution based on travel costs that are consistent with the intersection delays due to signal controls. This research attempts to achieve this goal by optimizing signal controls and estimating intersection delays accordingly, which are then used in traffic assignment. Simultaneous optimization of traffic routing and signal controls has not been accomplished in real-world applications of traffic assignment. To this end, a delay model dealing with five major types of intersections has been developed using artificial neural networks (ANNs). An ANN architecture consists of interconnecting artificial neurons. The architecture may

either be used to gain an understanding of biological neural networks, or for solving artificial intelligence problems without necessarily creating a model of a real biological system. The ANN delay model has been trained using extensive simulations based on TRANSYT-7F signal optimizations. The delay estimates by the ANN delay model have percentage root-mean-squared errors (%RMSE) that are less than 25.6%, which is satisfactory for planning purposes. Larger prediction errors are typically associated with severely oversaturated conditions.

A combined system has also been developed that includes the artificial neural network (ANN) delay estimating model and a user-equilibrium (UE) traffic assignment model. The combined system employs the Frank-Wolfe method to achieve a convergent solution. Because the ANN delay model provides no derivatives of the delay function, a Mesh Adaptive Direct Search (MADS) method is applied to assist in and expedite the iterative process of the Frank-Wolfe method. The performance of the combined system confirms that the convergence of the solution is achieved, although the global optimum may not be guaranteed.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ACRONYMS AND ABBREVIATIONS

1.	(%RMSE)	Percent Root Mean Squared Error
2.	AADT	Annual Average Daily Traffic
3.	ADT	Average Daily Traffic
4.	ANN	Artificial Neural Network
5.	DS	Direct Research
6.	FDOT	Florida Department of Transportation
7.	FSUTMS	Florida Standard Urban Transportation Modeling Structure
8.	FT	Facility Type
9.	FTI	Florida Traffic Information
10.	GA	Genetic Algorithm
11.	GIS	Geographical Information Systems
12.	HCM	Highway Capacity Manual
13.	MAE	Mean Absolute Error
14.	MFW	Method of Frank-Wolfe Algorithm
15.	MSA	Method of Successive Average
16.	OD	Origin-Destination
17.	PTMS	Portable Traffic Monitoring Sites
18.	RMSE	Root-Mean-Square Error
19.	TRANSYT-7F	Traffic Network Study Tool version 7
20.	v/c	Ratios of Traffic Volume to Capacity

1. INTRODUCTION

1.1 Research Background

In transportation planning, a travel demand model is often applied to forecast future travel demand of various transportation facilities and transportation network performance. As part of a demand model, a traffic assignment model estimates a network flow pattern, i.e., travel volumes using a specific transportation mode on network links for a given origin-destination (OD) matrix. Usual practice apply Wardrop's principle of user equilibrium (Ortuzar and Willumsen, 2001) that specifies that each traveler chooses the shortest (travel time) path subject to every other driver doing the same. The most important elements in traffic assignment are demand (represented by an OD matrix), network link capacities that generally describe the facilities' ability to meet travel demand, and travel cost (often measured by travel time). A solution of traffic assignment needs to overcome the fundamental difficulty that travel times are a function of demand, while demand is affected by travel time. Accurately modeling travel time is also a challenge. This is because the limitation of travel demand models, most of which are macroscopic simulation models, are unable to simulate the real-time traffic operation and have to treat the demand analysis problem at an aggregate level, including modeling demand as a daily demand, or for peak hour or hours and off-peak hours.

Delays at signalized intersections often contribute significantly to total travel time, especially on urban arterials under congested conditions. Because delay time may be directly translated into level of service or loss of productivity, it has significant economic implications. Therefore, minimization of delays is also an important goal of

transportation planning applications. Long-range planning models need to deal with intersection delays in the modeling procedure. In the four-step model procedure, with the exception of trip generation, all other three steps including trip distribution, mode choice, and traffic assignment, rely on accurate estimations of travel time. However, current planning models often consider intersection delays in a limited manner. That is, the stochastic nature of signalized delay is often circumvented and quantified as a type of deterministic travel cost. Ignoring delays at signalized intersections is a frequent practice opted for by many planning models, which inevitably affects the accuracy of traffic models. Therefore, adequately considering intersection delays is essential to improving the performance of planning models.

The complexity of modeling intersection delays for a planning model lies mainly in the variety of roadway geometries, signal plans, and the means of data collection and processing. Generally speaking, three categories of input data are required to estimate intersection delays. They are signal timing plans, traffic flow of each lane group, and geometric conditions. The cycle length, green splits, and traffic flow rate are required for control delay estimation, and the link capacity, lane group, and the segment length are important for queue delay estimation. For corridor analysis, signal progression may pose important influence on control delays. A major problem is that such data are often unavailable for forecasting purposes. For a transportation planning model, such data coverage for all of the intersections in a network may easily overburden not only data collection but the modeling procedure itself. Therefore, intersection delay estimation needs to involve as few variables as possible in a planning model for practical applications.

Traffic assignment and signal optimization, though usually dealt with separately, are two processes that interact with each other. To explicitly consider this interaction, numerous studies have been done on the integration of these two processes, often called a combined control and assignment problem. The combined problem involves two folds of optimization that are respectively aiming at optimizing signal timing and shortest path (traffic assignment). The solution of the combined problem, often called the mutually consistent point, is to reach a network flow pattern that is simultaneously optimal for both shortest path selection of traffic assignment and the signal timing at an intersection. The simple iterative optimization and assignment (IOA) is a frequently applied approximate algorithm used to reach mutually consistent solutions of signal settings and traffic assignment flows by intermittently/alternatively performing signal plan optimization and traffic assignment until convergence. However, theoretically speaking, IOA is not an optimization method for the combined system and usually fail to converge.

Because the delay model is expected to work together with a traffic assignment model, it is not practical to perform signal phasing design for every intersection using the standard traffic analysis procedures, which are both time-consuming and data intensive. A traffic assignment model needs to follow the demand of a planning model – forecasting future traffic conditions of a transportation network, of which the signal timing plan and the intersection geometries are unknown for a future forecast year. Therefore, the delay model has to estimate delays according to simple geometry information and volumes resulting from every traffic assignment iteration, while requiring no other information from the traffic assignment.

Another fundamental issue of this study is the convergence of the combined

system of traffic assignment and the intersection delay model. Traffic assignment is performed according to costs, which are partially determined by signal controls because signal controls determine intersection delays. For the combined problem, the traffic assignment problem, when cast as an optimization problem based on the standard user equilibrium (UE), usually does not converge (Lee and Machemehl, 2005). For the solution to be useful, the methodology must be able to reach a convergent solution and, at the same time, appropriately consider intersection delays.

1.2 Problem Statement

Intersection delays make up a large proportion of the total travel time in urban areas. However, current planning models are unable to properly consider travel time delays at intersections for the following reasons:

(1) The estimation of delay time at an intersection requires detailed intersection configuration and signal timing information, both for the base year and for a forecast year. Such information is often unavailable for use in planning models. While signal plans in the base year are known and may be coded, they are unknown for a future year and cannot be assumed to be the same as the base year (Zhao and Ding, 2006). This makes a base year model unsuitable for forecast applications if intersection delays must be considered.

(2) Estimating intersection delay during a model run using the method of Highway Capacity Manual (TRB 2000) is time-consuming because the number of intersections may be large and many iterations of traffic assignment will be necessary to reach a convergent solution.

(3) There are currently no commercial application models capable of incorporating signal optimization into the traffic assignment process due to non-convergence problems. The first two of these problems have been preliminarily dealt by the author with a research grant from the Florida Department of Transportation (FDOT). The results show that, with some reasonable simplifying assumptions, delays at intersections may be estimated with adequate accuracy. The further development of methods to address the convergence issue is urgently needed. At this point, direct search algorithms, requiring no explicit gradient information, may be applicable to solve for the combined system (Sheffi, 1985). Lacking efficient algorithms and empirical results, the combined control and assignment study often pose essential questions: How different are mutually consistent points from each other as network size increases and as realistic travel cost functions are used? Having recognized the non-convexity of the problem, can one search method effectively improve the quality of local solutions?

The failure of current travel models to consider intersection delays has a number of implications. Firstly, network travel cost cannot be accurately estimated. Consequently, assignment results may be inaccurate, and wrong transportation investment decisions may be made, resulting in possible waste of tax payers' money and the inability of the future transportation system to meet the travel demand. Secondly, travel time is critical to modal split. Inaccurate travel time estimations may result in incorrect estimation of transit demand, which may lead to improper investment in public transit.

To improve the accuracy of highway and transit travel time estimation, delays at intersections need to be considered carefully. A traffic assignment model that is able to accurately reflect intersection delays and produce convergent solutions is needed.

1.3 Research Objectives and Scope

This dissertation is aimed at investigating the feasibility of incorporating intersection delays into a traffic assignment model. To achieve this goal, a methodology will be developed to estimate intersection delays and to consider such delays during traffic assignment. The methodology must be simple, in the sense that it does not require information that is normally unavailable for long-range planning purposes. It must also be feasible

The first focus of this research is on developing an accurate and convenient intersection delay model, which performs based on signal setting, turning volume, and geometric conditions of an intersection. This dissertation aims at developing a combined model of an intersection delay estimating model and a traffic assignment model. The combined model is expected to be able to quickly converge to an optimal solution.

The specific objectives of this dissertation are to:

1) Understand the state-of-the-art in intersection modeling in travel demand models;
2) Establish simplifying standards for intersections with varied geometry, pedestrian activities, and traffic flow patterns in order to alleviate the difficulty in delay modeling and simulation;
3) Develop a delay estimating model that can be combined with a traffic assignment model; the delay model needs to be capable of estimating delays based on

changing control parameters including cycle length resulted from continuous signal optimization among traffic assignment iterations;

4) Search for an optimization algorithm that converges to a repeatable, stable, and bounded solution of both the delay estimating model and traffic assignment;
5) Determine a set of criteria to quantify and evaluate the solution of the combined system of the delay model and traffic assignment.

To limit the scope of the research, the following assumptions are made:

1) Signals at intersections within a network are not coordinated. This assumption is made due to the complexity of signal progression. It is much more complicated to describe a corridor or a subarea with signal progression in a planning model, and it will be time-consuming to optimize a coordinated signal plan for multiple intersections.
2) Only a limited number of intersection types will be considered. Although in practice there are many different types of intersections, developing delay models for all of them will be a significant undertaking. Because the goal of this research is to study the feasibility of a traffic assignment process incorporating a delay model, the traffic assignment and the delay model will work for five frequently-seen generic intersection types.
3) Small networks will be used for testing the methodology. This reduces computational time and allows numerous tests to be conducted. This limitation will not cause the methodology to be invalid or lose scalability. Computational efficiency of a travel demand model is important and will be investigated in the future in separate research.

1.4 Dissertation Organization

This dissertation is organized into five chapters. Chapter 1 introduces the background of this research, puts forward the problem to be solved, and sets the goals and objectives as well as assumptions. Chapter 2 provides a literature review on generic delay models, artificial neural networks (ANN), and the application of delay models in traffic assignment. The final part of literature review focuses on the algorithms searching for the solution of the combined system of traffic assignment and delay model. Chapter 3 firstly establishes the system architecture for combined system, then respectively outlines the procedure to prepare research data, to develop the delay model, and to complete the combined system. Chapter 4 completes comprehensive analysis on the performance of the ANN delay model as well as traffic assignment that has combined the delay model. A regression analysis is also presented to support the advantage of the ANN delay model by comparing the output statistics. As to the solution of the combined system, the converging pattern of traffic assignment iterations is identified in applications of both small and large networks. Finally, Chapter 5 provides conclusions, and identifies the limitations, original contributions, and conclusions of this research. Future research is also recommended.

2. LITERATURE REVIEW

In this chapter, literature related to this research is reviewed. Section 2.1 describes the Webster's delay model, which is the most fundamental of all delay models, is presented. Section 2.2 provides a discussion of the fundamental theories of delay models for a signalized intersection where the traffic is under conditions ranging from under-saturation to over-saturation. The most important applications, such as the 2000 version of the Highway Capacity Manual (HCM) and 2002 version of the Florida Quality/Level of Service Handbook, are also introduced. Section 2.3 further discusses some research efforts to improve the major delay models. Section 2.4 gives a description of the origin, architecture, and advantages of an ANN and its applicability to delay estimation. Section 2.5 focuses on issues related to combined models of signal optimization and traffic assignment, among which there is one core issue of this dissertation – an algorithm that ensures fast and accurate convergence. As used for the various research tasks, some applicable software programs are also briefly described and compared.

2.1 Webster's Delay Model

Many techniques are available for estimating delays at intersection approaches. However, little research has been performed to assess the consistency of estimates of various models (Dion *et al*., 2004). Moreover, the applicability of the delay models needs to be determined due to their different data requirements and algorithms. For a transportation planning model, a balance between simplicity and accuracy is essential when choosing a delay modeling technique.

Delay estimation techniques often have varied accuracy and their own limitations. For example, when the v/c ratio approaches 1, steady-state delay models tend to produce unrealistically large delay estimates, while over-saturation delay models will yield close to zero delays. Among many reasons for such differences, the most important is the v/c ratio. The technical complexity of delay models increase considerably when the volume is near the capacity.

Intersection delays may include two components: queue delay and control delay. Queue delay, or stop delay, is difficult to quantify due to its stochastic nature affected by random arrivals. Sophisticated techniques may work better in estimating queue delays, but are often impractical for planning models due to intense data requirements. It is often difficult to find a well-balanced queue delay model for integration into a planning model.

Control delay is the result of vehicles having to accelerate or decelerate at an intersection because of the traffic control. It is determined from signal setting, volume, and geometric conditions of an intersection. When control delays are incorporated into a planning model, they need to be updated repeatedly within traffic assignment iterations. A major problem is that in a planning model, of which the main purpose is to forecast future traffic conditions of a transportation network, signal timing plans and intersection geometry are unknown for a given forecast year. It is impractical to perform signal phasing and timing design for every intersection using the standard traffic analysis procedures, which are both time-consuming and data intensive. Therefore, it is necessary to facilitate the signal design and optimization procedure through simplifying assumptions (Aashtiani and Iravani, 1999).

The achievable accuracy of a planning model also depends on realistic objectives of an intersection delay model. Nowadays, adaptive signal settings and signal coordination are becoming more and more common in urban areas. As a result, the platoon effects of traffic progression are often significant and cannot be ignored in delay estimation. However, generic delay models are often inadequate in reflecting progression conditions. For example, the delay model of the HCM 2000 merely uses a progression adjustment factor to account for progression while treating a studied intersection as isolated. Beginning with HCM 1994, the delay calculations employ one of the most frequently used delay models based on the work by Webster (1958) as expressed in the following form:

$$d = \frac{C(1-\frac{g}{C})^2}{2(1-\frac{v}{s})} + \frac{(\frac{v}{c})^2}{2v(1-\frac{v}{c})} - 0.65(\frac{c}{v^2})^{1/3}(\frac{v}{c})^{(2+\frac{g}{C})} \qquad \text{Eq 2.1}$$

where

d = control delay per vehicle (s)

c = lane group capacity (veh/h)

C = cycle length (s)

g = effective green time (s)

s = saturation flow rate (veh/h), and

v = demand for subject lane group or approach (veh/h)

This formula has three parts. The first term estimates the average approach delay assuming uniform arrivals, which is consistent with the deterministic queuing models

mentioned earlier. The second term considers the additional delays attributed to the randomness of vehicle arrivals. The third term is an empirical correction factor that reduces the estimated delay by 5~15% to be consistent with simulation results. Equation 2.1 is among the most fundamental and frequently referenced equations of its kind. There have been many efforts to determine various parameters based on local conditions or developing theoretical modifications. As a result, many delay models often have a form similar to that of Webster's formula.

Numerous time-dependent delay formulas have been proposed and incorporated into a number of capacity guides, such as the 1994 and 1997 Highway Capacity Manuals and guides used in Australia and Canada. Details of the delay models applied in the HCM will be discussed in Section 2.3. The delay models in HCM 2000 currently used in the United States, Australia, and Canada all originated from the Webster's formula (Dion *et al*., 2004).

Webster's formula makes the simplifying assumption that the arrival function is uniform (i.e., arrivals are at a constant rate, v (veh/s)). With the uniform delay formula, random arrivals are not considered. At isolated intersections, vehicle arrivals are more likely to be randomly distributed. The assumption of uniform arrival implies that the queue of vehicles at an intersection operating under under-saturated conditions is always cleared before the next red signal. Generally considered to be the earliest model of its kind, Webster (1958) proposed a stochastic model that assumes that arrivals are Poisson distributed with an average rate v (veh/h). The "overflow delay" is ascribed to individual cycle failures, even with the v/c ratio for the entire analysis period is always less than 1.00.

Following Webster's work, some other stochastic models have been proposed, including, for instance, the models by McNeil (1968) and Heidemann (1994). These models all share several basic assumptions. First, the number of arrivals within a fixed time interval follows a known distribution, usually a Poisson distribution. This distribution does not change over time, which implies that these models should not be applied to estimate delays of coordinated intersections, where arrivals are platooned as a result of upstream traffic signals. Second, while it is recognized that temporary over-saturation may occur due to random arrivals, it is assumed that the system remains under-saturated throughout an analysis period. A primary consequence of such steady-state stochastic delay modeling is that the estimated delays tend to infinity as traffic demand approaches saturation (v/c ratio = 1.0). This is considered by many a weakness of this type of model (Roess *et al*., 1998). The concept of a time-dependent delay model was originally proposed and enhanced by Kimber and Hollis (1979). A proper delay estimation model theoretically should perform better for different demand levels. For low v/c ratios, the model is expected to produce delay estimates similar to those produced by deterministic queuing delay models assuming constant uniform arrivals. As demand increases, a growing proportion of delay is attributed to the random vehicle arrivals and the failure of all queued vehicles to clear in certain cycles. As the v/c ratio approaches 1.0, the model shall not approach infinity, but instead shall generate estimates tangent to the deterministic over-saturation model as Eq. 2.1 does.

2.2 Generic Intersection Delay Models and Applications

This section introduces the theory and history of generic delay models, which have been comprehensively studied for their characteristics and weaknesses. Applications such as the HCM 2000 and the Florida Department of Transportation Florida Quality/Level of Service Handbook (2002) are also described.

2.2.1 Intersection Delay Models

Almost every real-world model of delays at a signalized intersection begins with the Webster delay model (Eq. 2.1). Hurdle (1984) and Dion *et al*. (2004) provide excellent reviews of major delay models. They also studied the basic principles and simplifying assumptions that are not well-tailored to the real world. Although some improvements on methodologies and assumptions have been made, the theoretical core of delay models has remained basically unaltered. Hurdle's summary, which was based on a comparison of steady-state models and deterministic models, is still essentially instructive to this day.

Most signal intersection delay models fall into two categories, steady-state models and deterministic queuing models. The former are usually considered useful only for predicting delays at intersections with light loads, while the latter do well only in the analysis of heavily loaded intersections where volume overwhelms capacity ($v/c > 1$). These models ignore the random arrivals effect on the queue length when intersections are slightly saturated. Because their assumptions are based on different v/c values, these two types of models are incompatible. However, when the load is heavy but v/c is still less than one, some good models are expected to produce excellent estimates. In TRANSYT, developed by Transport and Road Research Laboratory, an algorithm based

on a compromise between these two types of models is employed. The algorithm, while not a solid and realistic model, is able to illustrate some intuitive ideas. The TRANSYT algorithm may be represented by an approximated formula (Robertson and Gower, 1977):

$$D = 15\frac{T}{c}((v-c)+((v-c)^2+240\frac{v}{T})^{1/2}) \qquad \text{Eq 2.2}$$

where

D = total delay for an intersection approach (veh/s),

c = capacity of an intersection approach (veh/h),

v = demand for subject lane group or approach (veh/h), and

T = duration of analysis period.

A derivation of the TRANSYT random delay equation was presented by Kimber and Hollis (1979). The basic idea is to achieve a smooth transition between the steady-state and over-saturation models in the v/c range around 1. However, the smooth transition between the two types of models is not the result of any detailed analysis. Instead, it is based on an intuitive understanding of what happens. As pointed out by Hurdle (1984), to improve the delay estimates, more refined queue behavior models are required. Unfortunately, such models tend to be too complicated and demanding where data input is concerned.

As a continued effort to study steady-state versus deterministic models, Dion *et al*. (2004) compared the delay estimates at under-saturated and over-saturated pre-timed signalized intersections. Deterministic queuing models are classic applications for predicting delays for signalized intersections. These models view traffic on each intersection approach as a uniform stream of arriving vehicles seeking service from a

control device that provides a high service rate. However, when the ratio of v/c is much lower than 1, the random effect is too evident to be ignored. This may be partly why such models have been applied mainly at intersections with far more arrivals per cycle than those that can be served during a green interval (v/c > 1). In such cases, the random effect may be negligible, and model performance is fairly adequate. Equations for calculating the average uniform vehicle delays during a cycle are presented below (Dion *et al*., 2004). Note that Eq. 2.3 is, in fact, identical to the formula in the HCM.

$$d_1 = \frac{C \times (1 - \frac{g}{C})^2}{2\left(1 - \min(1, X) \times \frac{g}{C}\right)} \qquad \text{Eq 2.3}$$

where

d_1 = uniform delay (s)

c = lane group capacity (veh/h)

C = cycle length(s)

g = effective green time (s)

s = saturation flow rate (veh/h)

v = demand for subject lane group or approach (veh/h)

$X = v/c$ ratio or degree of saturation for lane group

Steady-state stochastic delay models are one type of stochastic delay model that attempt to account for the randomness in vehicle arrivals. One fundamental, and most often referenced example, is Webster's model (Eq. 2.1, Webster, 1958). These models all assume that the number of arrivals in a given time interval follows a known distribution, typically a Poisson distribution, and that this distribution does not change over time. It is

also assumed that the system remains under-saturated over the analysis period. Although temporary over-saturation may occur due to the randomness of arrivals, the system is assumed to have been running long enough to settle into a steady state.

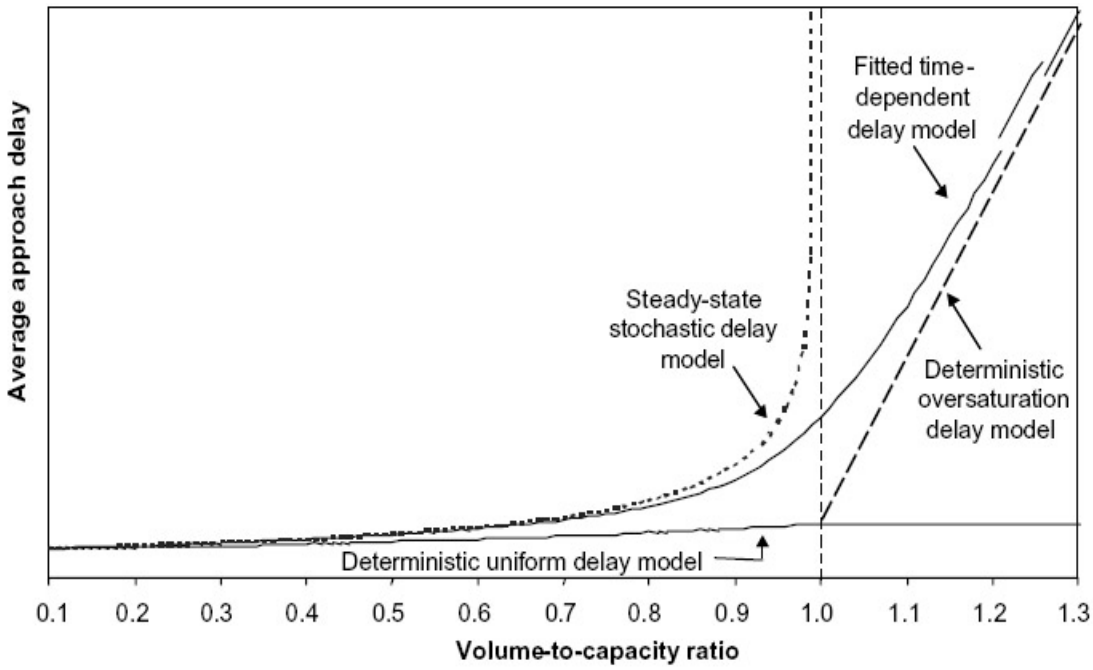

Figure 2.1 Steady-State Stochastic Models versus Deterministic Over-saturation Models (Dion *et al*., 2004)

To improve the performance of steady-state stochastic delay models and the deterministic queuing models, the concept of a general time-dependent delay model was introduced by Kimber and Hollis (1979) using the coordinate transformation technique. This technique transforms the equation of a steady-state stochastic delay model so that it becomes asymptotic to a deterministic over-saturation model. Although according to Hurdle (1984), there is no rigorous theoretical basis for this approach, empirical evidence confirms that the results are reasonable. Therefore, the delay models in the capacity guides of the U.S., Australia, and Canada, which are similar to each other, are all based

on the coordinate transformation technique. All of these models assume steady-state traffic conditions. Under stochastic equilibrium conditions, the arrival and departure flow rates remain stationary for an indefinite period of time. The number of arrivals is also assumed to follow a Poisson distribution, which remains constant over time, and the headways between departures have a known distribution with a constant mean value.

In addition to the majority of stochastic and deterministic models, a microscopic traffic simulation model is used to track individual vehicle movements in simulated networks, which allows such models to consider virtually any traffic conditions, ranging from under-saturation to severe over-saturation. The models determine the delay incurred to an individual vehicle traveling in a network with different characteristics by comparing simulated and ideal travel times. Dion *et al*. (2004) also employ the INTEGRATION microscopic traffic simulation software to arrive at delay estimates. The simulation model integrates dynamic traffic simulation and traffic assignment. Delay is estimated for each individual vehicle by calculating, for each traveled link, the difference between the vehicles simulated travel time and the travel time that the vehicle would have experienced on the link at free-flow speed. The average delay estimates from the INTEGRATION simulation model are in general agreement with the estimates from the various models such as the 1981 Australian Capacity Guide, the 1995 Canadian Capacity Guide, and the 1997 HCM delay models. Dion *et al*. (2004) pointed out a strong consistency in the delays estimated by the time-dependent stochastic delay models and by the INTEGRATION microscopic traffic simulation model.

Consistent with the conclusions by Hurdle (1984), Dion *et al*. found the same trend in the results from stochastic and deterministic models. All of the analytical delay

models generated similar results when the v/c ratios were low. Deterministic queuing always made the lowest estimates because this type model considers only uniform arrivals. Therefore, they are unable to consider the potential additional delays that arise from the random over-saturation delays caused by a platoon of arriving vehicles.

To summarize how to simply and effectively consider intersection delays, almost every real-world model of delays at a signalized intersection begins with the Webster delay model (Eq. 2.1).

2.2.2 Applications of Generic Delay Models

In the U.S., the HCM is the most comprehensively used reference of delay models (Troutbeck and Blogg, 1998), although the HCM's methodology comes with limitations that have been widely criticized. The intersection delay methodology of the HCM ignores the potential impact of downstream congestion on intersection operation as well as turn-pocket overflows on through volume and intersection operation. That is, the intersection is analyzed as an isolated facility. Therefore, the delay calculations merely reflect the average control delay experienced by all vehicles that arrive in the studied period, including delays incurred beyond the studied period when the lane group is over-saturated. Control delay includes movements at slower speeds and stops on intersection approaches as vehicles move forward in queue position or slow down upstream of an intersection.

For a given lane group, the average control delay per vehicle is calculated by

$$d = d_1 (PF) + d_2 + d_3 \qquad \text{Eq 2.4}$$

where

d = control delay per vehicle (s/veh)

d_1 = uniform control delay assuming uniform arrivals (s/veh)

PF = uniform delay progression adjustment factor accounting for effects of signal progression

d_2 = incremental delay to account for effect of random arrivals and over-saturation queues

d_3 = initial queue delay accounting for delay to all vehicles in analysis period due to initial queue at start of analysis period (s/veh)

In Eq. 2.4, d_1 and d_2 are defined as follows:

$$d_1 = \frac{C \times (1 - \frac{g}{C})^2}{2\left(1 - \min(1, X) \times \frac{g}{C}\right)} \qquad \text{Eq 2.5}$$

where

C = cycle length(s)

g = effective green time

X = v/c ratio or degree of saturation for lane group

and

$$d_2 = 900T\,[(X-1) + \sqrt{(X-1)^2 + 8kl\frac{X}{cT}}\,] \qquad \text{Eq. 2.6}$$

where

T = duration of analysis period

k = incremental delay factor dependent on controller settings

l = upstream filtering/metering adjustment factor

c = lane group capacity (veh/h)

X = lane group *v/c* ratio or degree of saturation

Both calculations of d_1 and d_2 assume no initial queue at the beginning of the analysis period of duration T:

$$d_3 = \frac{1800 Q_b (1+u)t}{cT} \qquad \text{Eq 2.7}$$

where

c = lane group capacity (veh/h)

Q_b = initial queue at the start of period T (veh)

T = duration of analysis period

t = duration of unmet demand in T (h)

u = delay parameter

These delay terms are estimated from variables or parameters that are related to operations upstream of the subject intersection. They include six vehicle arrival types (HCM, 2000), green time ratio (g/C), percentage of vehicles arriving during green time, degree of saturation (v/c), lane capacity, length of analysis period, and size of queue at the start of each cycle. Conditions of the downstream segments and intersections are usually ignored. As the HCM 2000 indicates, "The potential impact of downstream intersection on the upstream intersections is not taken into account." When a downstream intersection influences an upstream one, additional parameters/variables need to be considered other than those in the HCM 2000. The other major limitation is that random overflow at the connected link is not considered.

The HCM also provides procedures for calculating delays at two-way stop controls and all-way stop controls. To simplify the calculations, it is assumed that left turning lanes are always present on the major street.

In Florida, an important application of the HCM methodologies is the Quality/Level of Service Handbook of the Florida Department of Transportation (FDOT), referred to as FDOT Q/LOS herein, and its software, which is nationally recognized as the leading planning application of the HCM for the evaluation of automobile/truck LOS. According to Figure 2.2, both control delays and LOS criteria apply the HCM procedures. While operational analyses, such as intersection signal timing, are sometimes conducted at the planning level, the handbook does not provide the necessary tools for actual design or operation of facilities or services where more appropriate resource documents or analysis methods are available.

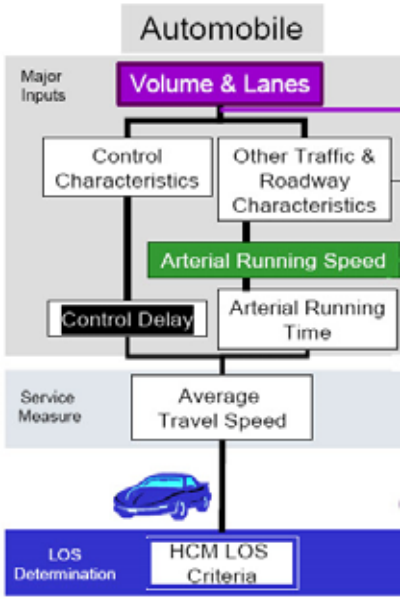

Figure 2.2 Control Delay in the Q/LOS Procedure of Florida (Quality/Level of Service Handbook, 2002)

The handbook's planning level analyses make extensive use of simplifying assumptions to primary Q/LOS evaluation techniques and default values to operational models. For example, a major simplifying assumption, which is essential to the development of the Generalized Tables in the FDOT Q/LOS, is the selection of a single effective green ratio (g/C) for all of the intersections of an arterial.

FDOT has determined that, for generalized planning analyses, the "weighted effective green ratio" yields the closest results to actual conditions. The weighted effective g/C of an arterial is the average of the critical intersection's through g/C and the average of the other intersections' through g/C. Another significant planning assumption is that mainline non-through movements are adequately accommodated. Typically, the through movement is the straight movement. However, occasionally the "through" movement is a right or left turning movement, with the straight ahead movement being considered a non-through movement. Most analyses of through movements in the HCM are relatively straightforward. Complications arise with the treatment of turning/merging movements, especially for signalized intersections and arterials. By handling non-through arterial movements (i.e., turns from the arterial and side street movements) in a general way, Q/LOS analyses are greatly simplified.

FDOT recommends the use of two submodels of FDOT Q/LOS, HIGHPLAN and ARTPLAN respectively, for highways and arterials. The assumed free flow speed is five mph higher than the posted speed. For arterial planning, traffic volume is included as a variable in the current 2002 version of the FDOT Q/LOS Handbook and the accompanying software. Specifically, FDOT include traffic volume as a variable in calculating running speeds and to better reflect running speeds of through vehicles, as

opposed to the total mix of through and turning vehicles. The Quality/Level of Service Handbook and its software are designed for the evaluation of roadway users' quality/level of service (Q/LOS) at planning and preliminary engineering levels. Q/LOS analyses are based on three types of input variables: roadway, traffic, and control. For an urban arterial, ten variables having a significant impact on volume calculation in LOS analysis are:

- Number of through lanes
- Left turn lanes
- Paved shoulder/bicycle lane/outside lane width
- Sidewalk
- Average annual daily traffic (AADT)
- Planning analysis hour factor (K)
- Directional distribution factor (D)
- Bus frequency
- Signalized intersection spacing
- Effective green ratio (g/C)

Most of these variables are required and are used in the standard HCM 2000 procedures. The software, as well as the handbook, is based on the HCM 2000 techniques. ARTPLAN is primarily applicable for urban signalized roadways.

2.3 Research Efforts to Improve Generic Delay Models

Many efforts have been made to overcome the limitations of the widely applied HCM delay model. For instance, under over-saturated traffic conditions, Benekohal and Kim (2005) found counterintuitive results because the progression adjustment factor (PF) is not applied to signalized delay models when there is an initial queue, as recommended in the HCM. On some occasions, delays under an initial queue condition end up being shorter than delays with a zero initial queue. Under over-saturated conditions, when there

is an initial queue, the HCM 2000 delay model yields the same uniform delay values for all arrival types, which does not seem reasonable because platooning affects delay. Benekohal and Kim propose a new uniform delay model considering platoon impact for over-saturated traffic conditions when progression is poor. This approach directly quantifies the platooning effects in delay, eliminating the need to apply a progression adjustment factor. Like the HCM 2000, the proposed model is applicable with or without an initial queue:

$$d_1 = 0.5sg\,[Q_1C + Q2(C\text{-}t_1) - q_oC^2 - sg^2] \qquad \text{Eq 2.8}$$

where

q_{av} = average arrival rate (veh/s)

q_{pl} = platoon arrival rate (veh/s)

q_n = non-platoon arrival rate (veh/s)

t_1 = platoon duration time (s)

q_o = overflow rate (q_{av} minus c) (veh/s)

Q_1 = number of arrivals when queue increase rate changes for the first time (= $q_p\ t_1$)

Q_2 = number of arrivals at the end of cycle (= $q_{av}C$)

Compared to inputs in the HCM, this arrival based model also requires platoon duration time (t_1), platoon flow rate (q_{pl}), and non-platoon flow rate (q_n) for calculating platoon and non-platoon arrival rates and compute the delay. The additional input may be difficult to collect from the planning perspective. However, the authors declared that this arrival-based approach was more accurate than the HCM approach.

Another major limitation of the HCM methodologies is that its delay model only deals with isolated intersections. At present, most delay models deal with congestion

delays without giving consideration to the impact of downstream congestion and traffic disturbances that may include waiting queues at downstream signalized approaches (Ahmed and Abu-Lebdeh, 2005). Closely spaced intersections are frequently seen in urban areas of the U.S. Other, more distantly spaced intersections with heavy traffic flow may also cause potential bottlenecks where downstream congestion may still cause unacceptable delays at upstream intersections.

The control delay from the HCM 2000 is a combination of three delays with a progression adjustment factor (PF) as shown below (Eq. 2.4):

$$d = d_1(PF) + d_2 + d_3$$

These three delays may be computed based on the following information: offsets, green phase at downstream intersection, distance between intersections, link traveling speed of vehicles, queue lengths, queue spillovers, speed of shockwaves, and so on. A new delay term may be needed to capture the influence of traffic operations at a downstream intersection and/or link on the neighboring upstream intersection. To estimate the length of delay due to a downstream disturbance, Ahmed and Abu-Lebdeh (2005) introduced a fourth delay term (d_4). This term will be determined and quantified by the geometry and traffic operational characteristics of both upstream and downstream intersections. Traffic disturbances at a downstream intersection may cause an interruption in flow on the link between two intersections. Consequently, a number of shockwaves are generated. Shockwave analysis is applied to evaluate the significance of a downstream disturbance for an upstream intersection. The average speed of traffic will be a function of space that is not occupied by traffic.

$$d_4 = \frac{n}{2}[2d_{(4)_1} + (n-1)h_v(\frac{1}{v_1} - \frac{1}{\lambda_1})] \quad \text{Eq 2.9}$$

where

n = total number of vehicles queued at the upstream intersection

h_v = effective space headway (m)

v_1 = speed of mid-block stopping wave (m/s)

λ_1 = speed of mid-block starting wave (m/s)

$d_{(4)_1}$ = portion of d_4 incurred by the first vehicle at an upstream intersection

$$d_{(4)_1} = (\frac{L_1}{v_2} + \frac{L_2}{v_1} + off - \frac{L_2}{v_a} - \frac{L_2}{\lambda_1}) \quad \text{Eq 2.10}$$

where

Off = offset (s)

L_1 = queue length measured from the downstream intersection stop line to the tail of the queue (m)

L_2 = remaining space on link (not occupied by vehicles) (m)

v_1 = speed of mid-block starting wave (m/s)

v_2 = speed of starting wave at downstream intersection (m/s).

v_a = average link speed (m/s)

Because the queue length at the downstream approach directly impacts the magnitude of d_4, the model needs to include parameters such as offsets, incoming volume from the upstream intersection, and other traffic control variables. Due to the many variables involved, including green/red phase, offsets, and average link speed, data requirements at this level of detail may overburden the transportation planning model.

Another direction of research is queuing theory. Troutbeck and Blogg (1998) compare queue accumulation and decay for a high-definition approach given random arrival and departures. The approximation of queue length and delay has been commonly called "coordinate transformation technique" following the publication by Kimber and Hollis (1979). Kimber and Hollis' theory is fairly similar to what is described by Hurdle (1984) regarding control delay, which is a mathematical representation of the steady-state queue length versus an over-saturation (deterministic) curve. As shown in Figure 2.3, the transformed equation by Kimber and Hollis (1979) produces a modified curve that transitions from steady-state models to deterministic ones. Troutbeck and Blogg compare the "coordinate transformation technique" with a solution to time-dependant and equilibrium queues by Newell (1982), whose methodology is based on the diffusion equation with the additional estimate of the variance.

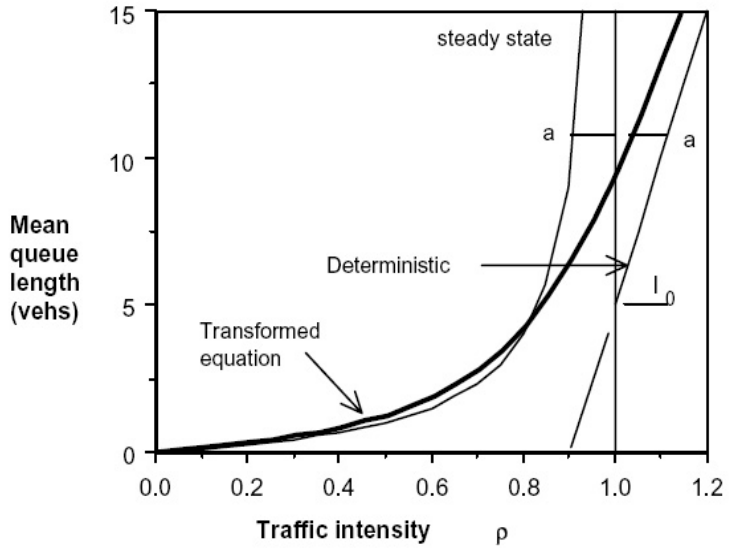

Figure 2.3 The Modified Curve of Transform Technique (Troutbeck and Blogg, 1998)

As Hurdle (1984) points out, Kimber and Hollis' approach simply uses mathematical expressions that fit the curve shown in Figure 2.3. Kimber and Hollis admit that in the limiting cases (the two ends of the curve) their results are correct and that in the intermediate regions the behavior of their functions is sensible. Kimber and Hollis' method provides little understanding of the system, particularly when a system reaches a critical point or as the demand approaches the capacity.

2.4 Artificial Neural Network

Other than mathematical formulas, some other non-linear search algorithms might be worth careful consideration in estimating delays at signalized intersections. The computations by artificial neural networks (ANNs) have emerged in the past few decades as a powerful paradigm that has found applications in almost all engineering branches. Neural networks were inspired by the mechanisms by which real biological neurons work in the human brain. The decision making process of the brain is simulated by an artificial network of neurons manipulating data among the many nonlinear nodes operating in parallel. Hornik *et al*. (1989) state that the multitasking ability of the human brain to simultaneously consider a large number of pieces of information and constraints is actually due to the powerful neural architecture of connections or parallel distributed processing. A trained network can predict output response to a high degree of accuracy much faster than sophisticated conventional models.

A neural network needs to learn from an enormous number of samples so that a particular input leads to a specific target output. During intense training, the network is constantly adjusted, based on a comparison of the network output and the target (original

records), until the network output matches the target. Typically many such input/target pairs are used in this supervised learning for a network (Demuth *et al*., 2006).

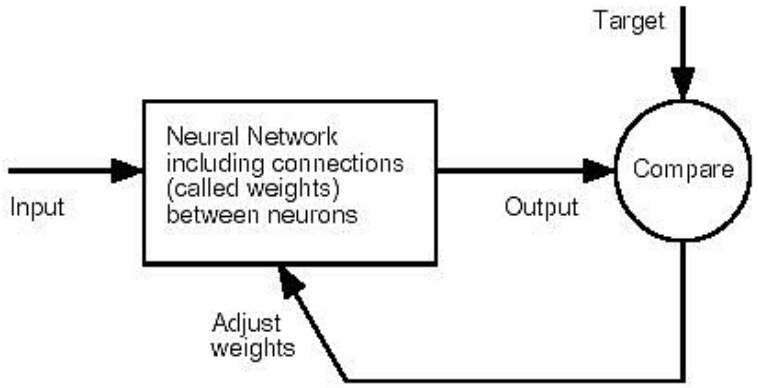

Figure 2.4 General Process of Supervised Learning of an ANN (Demuth *et al*., 2006)

The main type of ANN used in this study is referred to as a multilayered, feed-forward neural network. The following are essential:

1. A feed-forward propagation rule,

2. A network topology (i.e., the number of nodes, layers, and their connectivity), and

3. A learning rule. The error back-propagation algorithm (also known as the generalized delta rule) is the most commonly used learning rule (Demuth *et al*., 2006).

The feed-forward neural networks that use the error back-propagation learning rule are generally referred to as back-propagation neural networks. A typical back-propagation neural network architecture used in this paper is sketched in Figure 2.5. The *g* and *f* are transfer functions for the neurons in the hidden layer and in the output layer, respectively:

$$g = \sum_{p=1}^{n} (W_p X_p) + b_p \qquad \text{Eq 2.11}$$

$$f = \sum_{k=1}^{m} (W_k g_k) + b_k \qquad \text{Eq 2.12}$$

where

w = connection weight between neurons

b = bias term of corresponding nodes

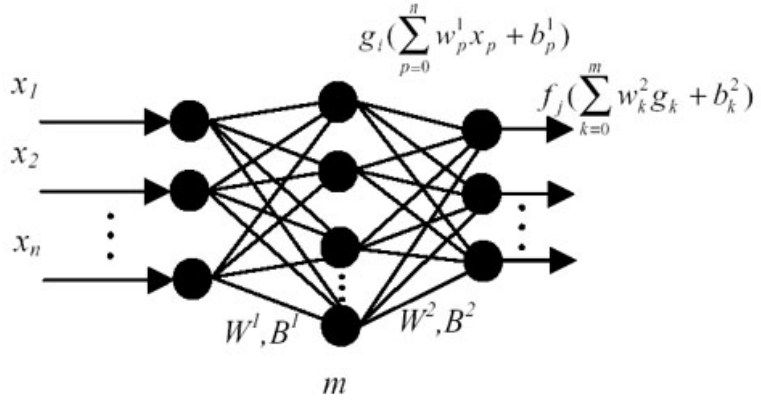

Figure 2.5 A Typical Structure of a Multilayered Neural Network (Demuth *et al*., 2006)

The multilayered back-propagation ANN usually has one input layer, one output layer, and constructed processing elements (artificial neurons) termed hidden layers. The hidden layers are sandwiched between the input and output layers. The neurons in these hidden layers allow the network to represent and compute more complicated associations between input and output patterns. The network operation consists of a highly nonlinear functional mapping of the neurons in the hidden layers between the input and output variables. Each artificial neuron or processing element receives several input signals X_j originating from previous nodes and then processes each signal considering its connection weight W_{ij}. For example, the relationship between the input signals and the level of internal activity of the processing element is given by the weighted sum of its inputs as follows:

$$N_i = \sum^{n} (W_{ij} X_j) - b_i \qquad \text{Eq 2.13}$$

$j=1$

where

N_i = net input signal (level of internal activity) in node i,

W_{ij} = connection weight between artificial neurons i and j,

X_j = value of signal coming from previous node j,

b_i = bias term of node i (corresponds to an activation threshold),

n = number of input signals from previous nodes.

When the weighted sum of the input signals exceeds the activation threshold *bi*, the artificial neuron outputs a signal y_i dictated by a transfer function *f(x)*. The output signal is then expressed as a function of the input signal N_i by:

$$y_i = f(N_i) \qquad \text{Eq 2.14}$$

where

$f(x) = 1 / (1 + e^{-x})$, may be a sigmoid function which accepts input over the range $(-\infty, +\infty)$ and uniquely maps the output y_i into the range [0,1].

The neural network modifies the connection weights between the layers and the node biases in ensuing iterations to allow a type of learning for the network. The weights and node biases are shifted until the error between the desired output and the actual output is minimized. Learning (or training) is the process whose objective is to adjust the link weights and node biases so that when presented with a set of inputs, ANN produces the desired outputs.

In recent years, artificial neural networks (ANNs) have been frequently employed in classification, optimization, and prediction. ANNs are suitable in such circumstances to predict the behavior where cause and effect relationships are little known. ANNs also

have the advantage of a well-defined process that requires no algorithmic conversion of an input into an output.

2.5 The Combined Model of Intersection Delay and Traffic Assignment

This section summarizes the research efforts in incorporating control delays into the traffic assignment process. Control delay estimating models need to be reasonably simplified before being employed to improve the accuracy of traffic assignment.

Having incorporated the delay model, traffic assignment still follows a generic methodology and a set of assumptions. Hence, users are always making wise and informed decisions, and the network's traveling cost cannot be reduced further. However, signal phasing design and traffic assignment procedures are mutually dependent on one another. Equilibrium is reached only when the necessary conditions of both aspects are met. The studies by Gartner and Al-Malik (1996), and Lee and Machemehl (1999) reveal many attempts to optimize the combined signal control and assignment problem. An iterative procedure may be applied on a network with more realistic intersections than the two-phase intersections discussed earlier. The simple iterative process, when unable to reach convergence, often continues to an endless oscillation (Lee and Machemehl., 2005). To dampen the oscillation, the method of successive averages (MSA), known also as a simplified transformation of the Frank-Wolfe algorithm, (Sheffi, 1985) may be useful. The MSA is based on a predetermined move size along the descent direction, and the procedure may be demonstrated as follows:

1. Initialization. Perform an equilibrium assignment based on a set of initial travel

costs t_0. This generates a set of network flows x_a. Set $n := 1$.

2. Update. Set $t_a = t_a(x_a)$,
3. Direction finding. Perform an equilibrium assignment based on current set of travel costs t_a, which yields an auxiliary network flow pattern y_a.
4. Move. Obtain the new flow pattern setting, set $a = (1/n)$.
   a) $x^{n+1}_a = x^n_a + (1/n)(y^n_a - x^n_a)$
5. Convergence criterion. Examine the similarity of network flows of successive iterations. If convergence is attained, stop. If not, set $n := n+1$ and go to step 1.

The major difficulty of the Frank-Wolfe algorithm is due to non-convexity (Lee and Machemehl, 1999) and circumvented in this way. Another alternative is using a direct search algorithm requiring no gradient information.

2.5.1 Simultaneous Optimization of Signal Settings and Traffic Assignment

Signal timing design for an isolated intersection has been covered in the HCM and in many standard textbooks such as that by Roess *et al*. (1998). Many commercial signal optimizers are available such as TRANSYT-7F, and Synchro.

For real-world applications, researchers often need to find an appropriate accuracy extent for strategic planning purposes. To consider a regional model with large zones and a relatively coarse network with delay functions for links, Hill (1998) also implemented delay functions based on selected analytic models for priority, roundabout, and signal controlled intersections. Zhou and Vaughan (1999) performed intersection modeling by treating complicated intersection situations using the macro capabilities of EMME/2 other than the normal assignment methods. EMME/2 has network calculation

modules to calculate the capacity and effective green time of turning movements. Their general approach to the new strategic highway assignment module involves calculating the effective green time and capacity for every movement in a network, which are fed into a turn penalty function to estimate the movement delays. The equilibrium assignment adds the movement delay to link delay to assign traffic that are used, in turn, to calculate the effective green time and link capacity in the following iteration. However, the model requires more input variables than traditional travel demand models. Furthermore, much more effort needs to be made to locate the input data, which include shared lane existence, signal control availability, opposed flow information, green time, and cycle time. A turn penalty function is applied to calculate delays of each movement at an intersection. This type of turn penalty function is in fact developed from a more general function form that embraces the delay functions seen in the Highway Capacity Manual and the Canadian and Australian methods and that appears the same as that mentioned earlier (Eq. 2.15).

$$D = D_u(x) + D_o(x) \qquad \text{Eq. 2.15}$$

where

D = total delay of a turning movement (s)

D_u = uniform delay (s)

D_o = overflow delay (s)

The turn penalty function shown in Eq. 2.15 is expected to estimate realistically the delay when the degree of saturation, x, is closer to 1.0. The model by Zhou and Vaughan (1999) is able to effectively represent various conditions at signalized intersections. Its iterative approach with a new turn penalty function usually proves to

achieve relatively quick convergence. However, little is discussed on the signal control optimizations during the iterative procedures.

Other efforts include a study by Ceylan and Bell (2004) using the genetic algorithm (GA) approach to solve traffic signal control and traffic assignment problems to optimize signal timings with stochastic user equilibrium link flows for an entire network. Levinson and Kumar (1994) also developed a delay model based on Hurdle's study and estimated the cycle time and green time using the methodologies suggested by Roess *et al*. (1998). The output of the intersection model is the average delay of a turning movement. The delay model is actually an application of Webster's formula. One important finding by Levinson and Kumar is that loading from highly aggregate zones to a single point will over-saturate the network at that point and seriously disrupt signal timings. The limitation of their method is that signal timing plans are often not fully optimized. A more intuitive method by Gartner and Al-Malik (1996) is promising for theoretical applications. This method employs a solution procedure that enables the simultaneous optimization of the two problems: signal setting and link volume estimation. That is, the signal settings produce link costs that determine a flow pattern such that these settings are optimal for it. Signal settings are determined by a network optimization procedure, for example, MAXBAND or TRANSYT-7F, on the basis of traffic volume data previously collected under the existing signal settings. The key to an efficient control strategy is to measure the effect of new signal timings because drivers adjust to them, thus resulting in new user-optimized traffic flow patterns. Gartner and Al-Malik's model simultaneously evaluates the route choice behavior of the motorist and determines the

corresponding optimal signal settings, both of which are essential to rerouting traffic for the purpose of reducing congestion and avoiding bottlenecks.

The model is among the first to introduce a way of expressing signal controls as flow variables in a deterministic manner. However, this experimental procedure considers only an individual signalized intersection with the following simplifications:

1. Only two conflicting streams,
2. Two-phase operation,
3. Fixed cycle length given, and
4. One isolated intersection (offset is not considered).

Therefore, it is still a distance from real-world applications. The traffic assignment aspect of Gartner and Al-Malik's application follows the generic methodology and assumptions. In other words, the users are always making wise and informed decisions and the network's traveling cost cannot be reduced further. Having developed a flow-dependent signal control model, signal setting and traffic assignment procedures are ready to be combined into one inclusive model. The equilibrium is reached only when the necessary conditions of both aspects are met. Compared to the study performed by Gartner and Al-Malik (1996), Lee and Machemehl (1999) made a further attempt to optimize the combined signal control and assignment problem. An iterative procedure was applied on a network with more realistic intersections than the two-phase intersections discussed earlier. Because Wardrop's two principles define, respectively, the user equilibrium (UE) and the system-optimized (SO) assignments, Lee and Machemehl suggest an iterative procedure to solve the combined problem of signal

optimization and traffic assignment, which are treated as two sub-problems as shown in Figure 2.6.

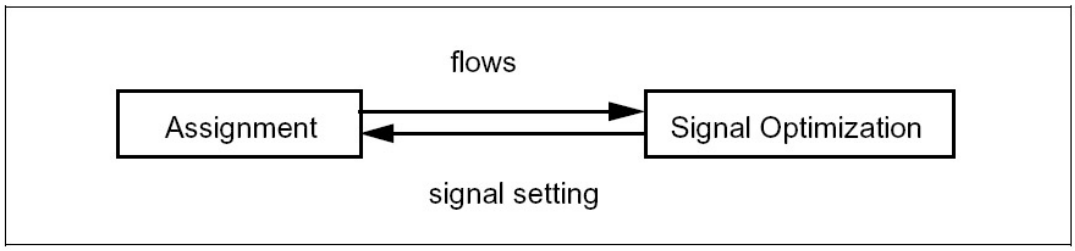

Figure 2.6 Iterative Optimization and Assignment Procedure

The assignment uses link performance functions resulting from signal optimization. Signal optimization is performed with flow patterns provided by the assignment sub-problem. This is so-called the Iterative Optimization and Assignment Procedure, or simply Iterative Approach (Cantarella and Sforza, 1987). The procedure continues until it converges to one solution, which is termed mutually consistent because the flow is at UE and the signal setting is optimal at the same time. Similar to the study by Gartner and Al-Malik (1996), Lee and Machemehl (1999) utilize Webster's delay function in traffic assignment. Because the equilibrium network-traffic signal optimization problem is not necessarily convex, it may have multiple local solutions. Therefore, it is possible that some local and mutually consistent solutions will show poor performance compared to the others. The driver route choice rule is minimum time path selection so that drivers follow deterministic user equilibrium. The objective function of the model is to minimize the total travel time of the equilibrium network, that is,

$$\min z(x) = \sum_a \int_0^{X_a} t_a(\omega) d\omega \qquad \text{Eq 2.16}$$

subject to

$$\sum_k f_k^{rs} = q_{rs} \qquad \forall r,s$$

$$f_k^{rs} \geq 0 \qquad \forall k,r,s$$

$$x_a = \sum_r \sum_s \sum_k f_k^{rs} \delta_{a,k}^{rs} \qquad \forall a$$

where

$z(x)$ = the travel time required by all network flows on the network.

x_a = the flow on link a,

f_k^{rs} = the flow on path k of origin r and destination s.

There are two difficulties in solving the above objective function (Lee and Machemehl, 1999). First, due to the non-convexity, z may have various local minima. As a result, any gradient-based search will find only a local minimum. Second, z requires knowledge of the OD pattern, which is difficult to develop for large, sophisticated networks. The iterative approach has typically been a practical alternative.

To solve z, Lee and Machemehl use two approaches, namely local search and iterative approach, to compare the optimal solutions. It is found that when the network is small, there may be only several distinct local solutions, which may be obtained easily by local searches. Although the mutually consistent solution is intrinsically suboptimal, it is close to the local solutions for a small network when demand level is low. As demand increases, the difference will grow. For a large network, there may be enormous local or quasi-local solutions. Therefore, any local search may easily result in worse solutions if the initial solution is not in a good domain neighborhood. Lee and Machemehl used a simplified method based on a gradient approximation suggested by Sheffi and Powell

(1983). Because the iterative approach includes a signal optimization procedure, it finds a good solution showing a short total travel time, which may not be mutually consistent until convergence, regardless whether the initial solution is in a good neighborhood or not. Then the search drifts to find a mutually consistent point. For a large network with high demand, there may be many mutually consistent points, such that it is likely to find one close to the signal optimal point.

Simplified gradient estimation local searches show promising performance as well as computational efficiency. However, for large networks with high demand, the iterative approach tends to find better solutions and is more valuable in real-world applications. Another promising alternative is the direct search method known as unconstrained optimization techniques that do not explicitly use derivatives. The phrase "direct search" describes sequential examination of trial solutions involving comparison of each trial solution with the "best" solution obtained up until that time, together with a strategy for determining (as a function of earlier results) what the next trial solution will be. The procedure employs straightforward search strategies that employ no techniques of classical analysis except where there is a demonstrable advantage in doing so (Kolday *et al*., 2003). Many users from the scientific and engineering communities preferred to avoid the calculation of gradients, which was for a long time the single biggest source of error in applying optimization software. At present, two things have become increasingly clear about the direct search method (Kolday *et al*., 2003):

1. Direct search methods remain an effective option, and sometimes the only option, for several varieties of difficult optimization problems.

2. For a large number of direct search methods, it is possible to provide rigorous guarantees of convergence.

A preliminary study has been performed using the direct search method to find at least a local optimum. The local optimum ensures the equilibrium between traffic assignment and signal controls. In other words, the signal timings optimized are not affected by the negligible change of the assigned volume, and so it is with intersection delays. The issues regarding the applied optimization search algorithm in the model will provide solid proof of convergence and the relevance to real-world applications.

2.5.2 Convergence Solutions and Search Algorithms

The combined system aims at solving Eq. 2.16. The calculation is developed from a link travel cost calculation. The link flow on a single link may be calculated as

$$q_a = \sum_m \sum_n \sum_k P_k^{mn} \delta_{ak}^{mn} \qquad \text{Eq 2.17}$$

where

δ_{ak}^{mn} = 1 if link a is on path k and 0 otherwise

P_k^{mn} = flow on route k connecting OD pair (m, n)

f^{mn} = trip demand rate between origin m and destination n

If $t_a(q_a, q_b)$ denotes the average travel time on link a (q_b denotes the conflicting flow on link b), the user equilibrium objective function is

$$z(Q) = \frac{1}{2} \sum_a [\int_0^{q_a} t_a(w, q_b) dw + \int_0^{q_a} t_a(w, 0) dw] \qquad \text{Eq 2.18}$$

and the corresponding system optimization function is

$$z(Q) = \sum_a q_a t_a(q_a, q_b) \qquad \text{Eq 2.19}$$

Sheffi (1985) establishes two conditions that are required for the user-equilibrium problem to have a unique solution. The link travel time is a strictly increasing function to the flow of that link; and a link's own flow exerts more influence on its travel time than the flows on any other link do.

1. $\frac{\partial t_a(q_a, q_b)}{\partial q_a} > 0$

2. $\frac{\partial t_a(q_a, q_b)}{\partial q_a} > \frac{\partial t_a(q_a, q_b)}{\partial q_b}$

The objective functions Eq. 2.18 and Eq. 2.19, satisfying neither condition, are, hence, non-convex. Although two simple network examples are presented with good performance, when confronting a complex network, it has to be determined whether to search for the best among the multiple solutions or to modify the network to converge to a single solution. Sheffi (1985, p.117) also formulated fundamentals on the uniqueness of the UE flow that provide a solution regarding traffic assignment.

However, in this study, the model aims at actual application. Therefore, the regular theoretical assumption that the link performance functions are independent of each other has to be relaxed, as it is not always valid. For example, left turning movements in signalized intersections have a discernible influence on the green time allocated to the other movements and, thus, the delays. On the other hand, the delays that left turning traffic receives are often not dominated by the left turning volume. In the real world, the link interactions tend to be asymmetric. That is to say, the marginal effect of one link's flow, x_a, on the travel time of the other link, b, is not equal to the effect of x_b

on the travel time of link *a*. It has been repeatedly proven that there is no known mathematical program able to find the equilibrium flow pattern for a standard (fixed-demand) UE model (Sheffi, 1985; Lee and Machemehl, 2005. Researchers have been striving to apply direct solution algorithms to tackle the problem (Sheffi, 1985). However, it is now known that a necessary and sufficient condition for the monotonicity of the link travel time function is that the Jacobian matrix (Eq. 2.20) must be positive definite (Smith, 1979). The Jacobian matrix is composed of the partial derivatives of the total link travel time function with respect to all link flows. The necessary and sufficient condition may not be valid in the real world, and so a unique equilibrium solution may not be available. There could be multiple equilibriums for a UE traffic assignment considering link interactions on delays.

$$J = \begin{bmatrix} \frac{\partial t_1(x)}{\partial x_1} & \frac{\partial t_2(x)}{\partial x_1} & \cdots\cdots & \frac{\partial t_a(x)}{\partial x_1} & \cdots\cdots \\ \frac{\partial t_1(x)}{\partial x_2} & \frac{\partial t_2(x)}{\partial x_2} & \cdots\cdots & \frac{\partial t_a(x)}{\partial x_2} & \cdots\cdots \\ \cdots\cdots & \cdots\cdots & \cdots\cdots & & \cdots\cdots \\ \frac{\partial t_1(x)}{\partial x_a} & \frac{\partial t_2(x)}{\partial x_a} & \cdots\cdots & \frac{\partial t_a(x)}{\partial x_a} & \cdots\cdots \\ \cdots\cdots & \cdots\cdots & \cdots\cdots & \cdots\cdots & \cdots\cdots \end{bmatrix} \qquad \text{Eq 2.20}$$

Heydecker (1983) suggested two issues that should be settled in order for a traffic assignment to produce desirable results. One is that the assignment should have one single, stable solution. The other is that the procedure should always be able to converge to the solution. Heydecker's standards may be applied in the evaluation of the combined system that integrates a traffic assignment model and a delay estimating model.

A simple iterative process unable to reach convergence often continues to an endless oscillation (Lee and Machemehl, 2005). To dampen the oscillation, the method of

successive averages (MSA), known also as a simplified transformation of the Frank-Wolfe algorithm (Sheffi, 1985; Ortuzar and Willumsen, 2001), may be useful. The method is based on a predetermined move size along the descent direction, and the procedure may be demonstrated as follows:

1. Initialization. Perform an equilibrium assignment based on a set of initial travel costs t_0. This generates a set of network flows x_a. Set $n := 1$.
2. Update. Set $t_a = t_a(x_a)$,
3. Direction finding. Perform an equilibrium assignment based on current set of travel costs t_a, which yields an auxiliary network flow pattern y_a.
4. Move. Obtain the new flow pattern setting, set $a = (1/n)$.
   $x^{n+1}_a = x^n_a + (1/n)(y^n_a - x^n_a)$
5. Convergence criterion. Examine the similarity of network flows of successive iterations. If convergence is attained, stop. If not, set $n := n+1$ and go to step 1.

The original Frank-Wolfe algorithm, different from the MSA in step 4, optimizes move size factor a using mathematical programming methods so that convergence of UE might be more efficiently reached. The goal of the Frank-Wolfe algorithm is to find a downhill direction and proceed to step 5. However, because at step 5 an UE assignment applies an ANN delay model, it is difficult or impractical to solve for a using traditional mathematical programming that uses the gradient or higher derivatives of the objective function.

Among direct search methods requiring no gradient or derivatives, the mesh adaptive direct search (MADS) algorithm is one that might be applied to seek the optimized a in the step 4. Capable of minimizing the potentially non-smooth function, the

MADS allows local exploration in a dense set of directions in the space of optimization variables (Audet and Dennis, 2006). As a direct search algorithm that computes a sequence of points that get closer and closer to the optimal point, at each step MADS searches a set of points, called a mesh, around the current point — the point computed at the previous step of the algorithm. The mesh is formed by adding the current point to a scalar multiple of a set of vectors. If MADS finds a point in the mesh that improves the objective function at the current point, the new point becomes the current point at the next step (MathWorks, 2004). MADS, when employed in the combined system, finds a scalar, α, which solves the program in the form of Eq. 2.21.

$$\min z(x^n + \alpha(y^n - x^n)) = \sum_a \int_0^{x+\alpha(y-x)} t_a(x) d\omega \qquad \text{Eq 2.21}$$

subject to

$$0 \le \alpha \le 1$$

2.5.3 The Applicable Software

According to Traffic Analysis Tools Primer (Alexiadis *et al*., 2004), majority of analytical/deterministic tools employ the procedures of the HCM. These tools conveniently predict capacity, density, speed, delay, and queuing on a variety of transportation facilities and are validated with field data, small-scale experiments, or laboratory test beds. Analytical/deterministic tools are suitable for analyzing the performance of isolated or small transportation facilities. However, they are limited in their capability to study networkwide system effects.

For many applications, the HCM is the most comprehensively applied and acknowledged traffic analysis technique in the U.S. The HCM procedures are ideal for

handling the performance of isolated facilities with moderate congestion problems. They are quick and reliable for predicting if a facility will be operating beyond its capacity, and they have been well experimented through enormous field validation efforts. However, the HCM procedures are generally inadequate in their ability to assess system effects. Majority of the HCM methods and models assume that the performance of an intersection or road segment is not adversely influenced by conditions of adjacent streets. Long queues at one facility that interferes with another nearby location violate the assumption. If the HCM procedures do not meet the needs of the analysis, it requires the users to determine whether microscopic, mesoscopic, or macroscopic simulation is necessary. If it is not indispensable to microscopically track individual vehicle movements, the analysts may enjoy advantage of the simpler data entry and control optimization features available in regular mesoscopic or macroscopic simulation models.

For comprehensive traffic analysis functions including signal timing optimization and signal coordination, Synchro and TRANSYT-7F have been widely applied. TRANSYT-7F has been popular since the 1980s and many extensions have been produced for various customized applications. Some research (Wong *et al*., 2001) has indicated that TRANSYT-7F is usable to model intersection delays while considering coordination effects.

Synchro has a friendly user interface for most traffic analysis of signals and is more practical than TRANSYT-7F. Synchro uses two methods for calculating delays: One is based on Webster's formula, and the other, newer one is called the Percentile Delay Method. It is assumed that each of these scenarios will be representative of 20% of the possible cycles of signal phases. For each scenario, traffic for each approach is

adjusted to that percentile. Delays are calculated using the adjusted volumes, and green times are calculated. If the intersection is near saturation or above saturation, additional time will be added to account for vehicles carried over between cycles. However, Synchro does not provide a macro running mode that is capable of processing hundreds of simulations and optimizations automatically. In this respect, TRANSYT-7F is superior because of its convenient macro function.

CUBE, a travel demand model software package by Citilabs, is capable of considering intersection delays and is widely applied in Florida. The control delay estimates are by default based on the HCM's delay model.

2.6 Summary

Although the intersection delay estimation technique of the HCM has been widely applied, it is merely based on curve fitting rather than a sound mathematical model of signal systems. Therefore, when signal systems operate under oversaturated conditions (v/c = 1), many traffic conditions are still not well modeled. However, reasonable results are possible under the condition that the users are aware of model limitations. None of the deterministic or steady-state models could produce fully consistent or accurate results. Although not always correct, it has been generally agreed on that most steady-state delay models and deterministic models considered here generate relatively consistent delay estimates when employed for under-saturated signalized intersections with v/c ratios below 0.6 (Dion *et al*., 2004). To develop a new generation of models that reasonably consider variations of travel demand over time, more information on traffic patterns is

essential. At present, it may be unrealistic to expect the availability of such information. However, for microscopic operational analysis, such data may be obtained.

It is a common practice for a traffic assignment model to assume that an intersection is isolated if the estimates are made based on the HCM delay model (Gartner and Al-Malik, 1996; Lee and Machemehl, 1999). The prediction curve based on the TRANSYT traffic model developed by Robertson (1977) has been widely accepted as an effective tool for evaluating queues and delays on links in a network. The traditional delay models are mostly too awkward to be incorporated by a planning model due to either data requirements or disappointing functionality. The ANN method is promising because it is highly capable of handling nonlinear fitting. Moreover, there are theoretically enough simulation scenarios to train the ANN model by adjusting the internal weights. A global optimum of the combined system of traffic assignment and delay model is impossible to reach using traditional mathematical programming. However, direct search methods have the potential to guarantee convergence solutions. Therefore, although IOA may be useful now and then, IOA strengthened by a direct search algorithm is recommended. It is still an open field as far as finding an advanced search algorithm for the purpose of the combined system goes.

3. RESEARCH METHODOLOGY

3.1 System Architecture

The research mainly studies a combined system with an architecture illustrated in Figure 3.1. The box in the upper left corner is the process during which a dataset of traffic volumes and delays is created. The dataset formulates a direct relationship or approximating function between volumes and corresponding delay. Based on this dataset, the intersection delay model will be calibrated and will predict the intersection delay for given volumes at an intersection. Finally, this model will be applied during the traffic assignment iteration process of a planning model. Resulting from traffic assignment, the assigned volume will be provided to the delay model, which will, in turn, estimate the intersection delays. The delay estimates will then be used to update the travel costs in the next traffic assignment.

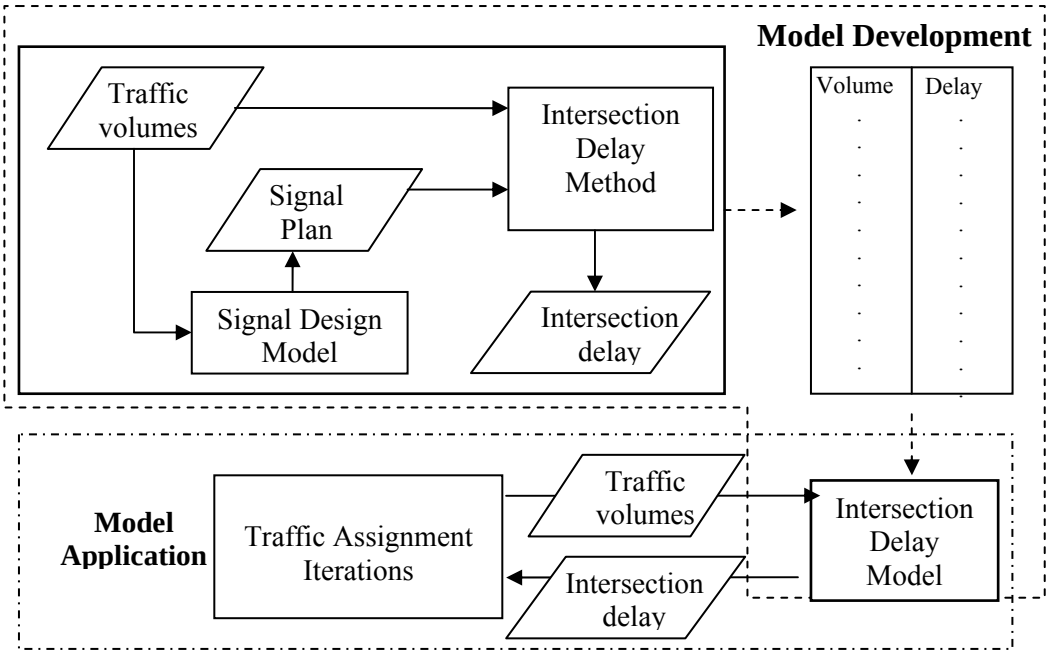

Figure 3.1 Conceptual Process of the Proposed Methodology (Zhao and Ding, 2006)

The methodology has mainly two advantages. First, the delay model is able to estimate delays for an intersection with an implicitly optimized signal plan. The second advantage is that the delay model takes its input directly from a planning model and estimates delays that can be easily used to update the travel cost in the planning model, therefore forming a tight integration of the two processes. The input data to the delay model are movement volumes and facility information including facility type, link capacity, and number of lanes at an intersection.

In Section 3.2, data preparation for calibrating the delay model and for testing the combined delay-assignment model is described. The delay model is built using the Artificial Neural Network (ANN) technique. The development of the ANN delay model is presented in Section 3.3. Finally, in Section 3.4, a combined system that integrates the delay model and traffic assignment model is discussed.

3.2 Data Preparation

3.2.1 Study Networks / Intersections

A virtual street network consisting of 20 signalized intersections is constructed. The generic geometric conditions, including speed limits, are maintained. However, the traffic load on the network is a large set of random OD matrix. There are several factors, such as pedestrians and on-street parking, that are either unavailable or uneconomical for explicit consideration by the delay model during traffic assignment. As such, they may be more conveniently applied in other circumstances.

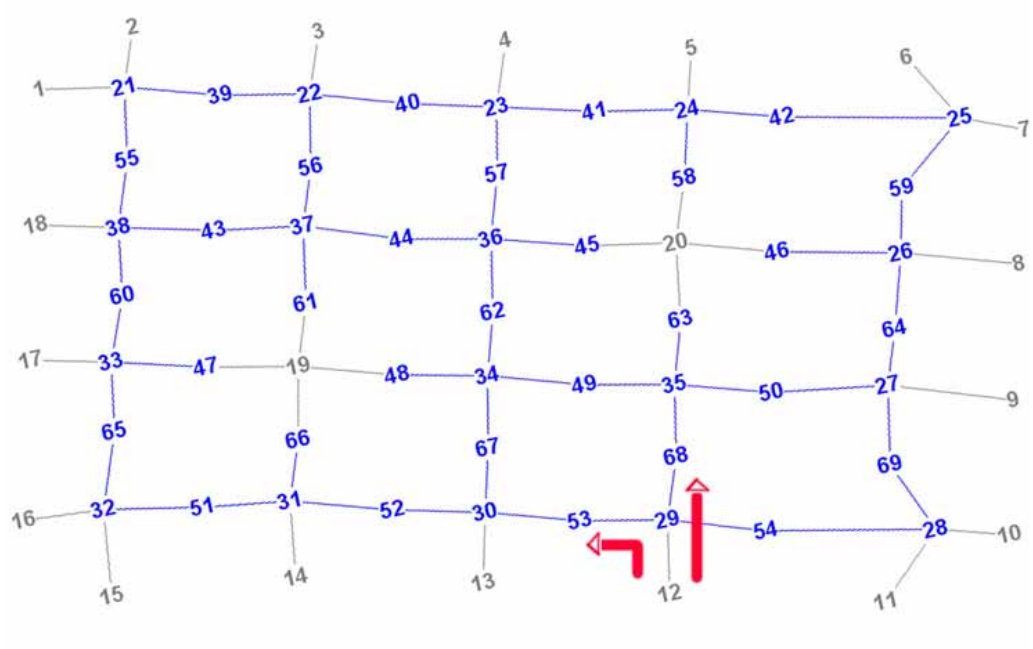

Figure 3.2 The Large Network for Concept Demonstration

Considering that this study aims at a promising method to solve for the combined system, one fundamental simplification is to reduce the infinite number of intersection configurations into manageable categories so that the later simulation of various intersection conditions and the training of the ANN delay model are convenient. The intersections are divided into five types that are most-frequently seen in the Gainesville urban area as shown in Table 3.1. The geometry conditions are simplified using uniform lane width, link length, number of lanes, and speed limits determined by the facility type and area type. The frequent on-street parking and pedestrians are not considered. The delay model avoids incorporating certain local conditions that may skew the delay estimates when applied in similar circumstances. Therefore, the considered network with assumed parameters is highly generic and fairly different from the original in terms of geometric conditions.

Table 3.1 Intersection Types of Different Facility Types and Lanes

Intersection Type Code	Description of Intersection Type
2322	Divided three-lane arterial with divided two-lane arterial
2222	Divided two-lane arterial with divided two-lane arterial
2241	Divided two-lane arterial with one-lane local road
3141	Undivided one-lane arterial with one-lane local road
4141	One-lane local road with one-lane local road

To facilitate and expedite the experimental operations of the combined system, a simple small network is also constructed (Figure 3.3). The small network may save a great deal of running time while still measuring the performance of the combined system. The convergence problem is expected to be dealt with first on the small network.

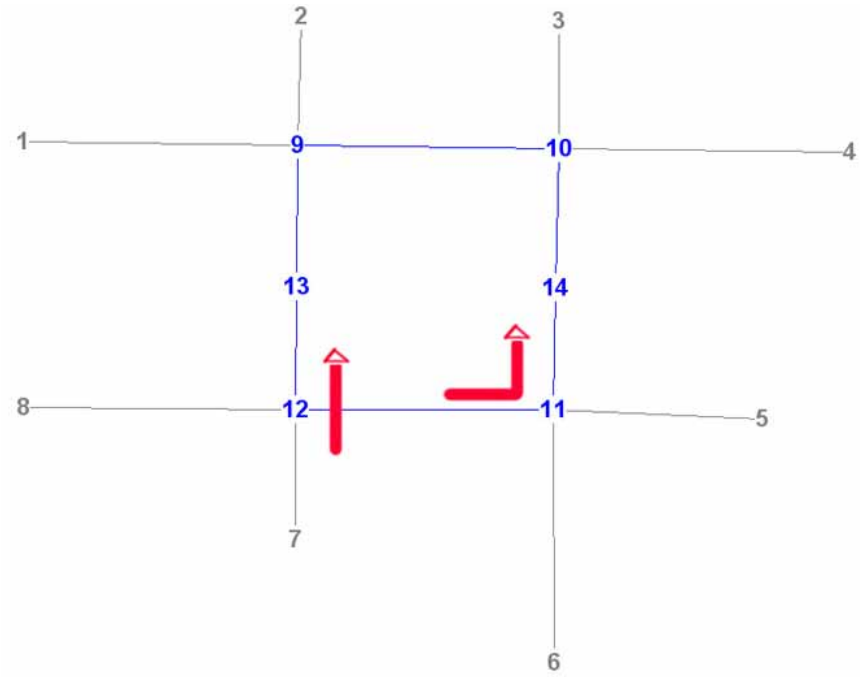

Figure 3.3 The Small Network for Concept Demonstration

3.2.2 Simulation Scenarios

The delay model is developed based on an ANN architecture, which requires sufficient scenario data for training an ANN. Due to the lack of comprehensive field data for the typical intersections operations, a large number of simulations are performed

using TRANSYT-7F, which is a signal timing optimizer as well as a traffic simulator with a batch mode option. Serving as the data source for the ANN delay model, these scenarios contain most of the possible traffic conditions for a studied intersection. The corresponding timing plans are optimized by TRANSYT-7F based on the inputs of geometry and volumes. Finally, the movement delays may also be calculated based on the timing plans.

The ANN delay model needs to "learn" from varied volume conditions to be able to predict delays accurately. To obtain control delays at intersections of different geometric conditions, datasets are developed to represent most traffic conditions at an intersection. Because a generic four-leg intersection has 12 movements, for which there are an infinite number of possible volume conditions, it is impractical to enumerate all possible volume conditions of all of the movements at an intersection. One simplification used in most signal optimizers such as Synchro and TRANSYT-7F is to combine right-turn traffic with through volume if they share the same lane. This reduces the 12 movements to eight movements. In other words, a simulation scenario of an intersection has eight samples of movements (four left-turning and four through movements) and, therefore, eight samples of delay estimates.

During the simulations, TRANSYT-7F firstly optimizes the signal plans based on the intersection volumes and then produces delay estimates accordingly. Together with the geometry information of the studied intersection, the volumes and the corresponding delays form the data required for the ANN delay model training.

3.3 The Development of an ANN Delay Model

To estimate intersection delays with adequate accuracy based on inputs directly available from traffic assignment, the Artificial Neural Network (ANN) technique is applied to develop the delay model. MATLAB programming is used to develop the ANN model by establishing relationships between traffic conditions and intersection delays. To be specific, the inputs are the movement volumes and facility information including facility type, area type, and number of lanes at an intersection, while the output is the movement's intersection delay.

For all types of identified major intersections, the ANN model has two internal architectures that deal with the left-turning traffic and the through traffic, respectively. The model estimates the control delay for each movement using simulated volumes of all approaches at an intersection. The performance of the ANN delay model may be easily evaluated through comparison of the model estimates with TRANSYT-7F simulations.

3.3.1 Architecture of the ANN Delay Model

Learning rules, which determine the architecture of ANN models, are important to model performance. Among the commonly used learning rules, back-propagation trains a multilayer feed-forward network with differentiable transfer functions to perform function approximation, pattern association, pattern classification, as well as a number of optimization strategies (Demuth *et al*., 2006). The term *back-propagation* refers to the process by which derivatives of network errors with respect to network weights and biases may be computed. The architecture of a multilayer network is not completely constrained by the problem to be solved. It has been suggested that a two-layer (sigmoid/linear) network may represent any functional relationship between inputs and

outputs, provided that enough neurons are used (Demuth *et al*., 2006). In this study, the ANN models have two layers of neurons. As Figure 3.4 shows, one layer using a sigmoid transfer function handles input vector **p**, which is weighted by vector **w.** The second layer is the output layer that produces result *A*, which follows a linear relation as Figure 3.5 indicates. Thus, the network models an approximate mathematical relation:

$$A = f(wp + b) \qquad \text{Eq 3.1}$$

where

A = ANN output

w = weight assigned to inputs

p = inputs

b = adjusting bias

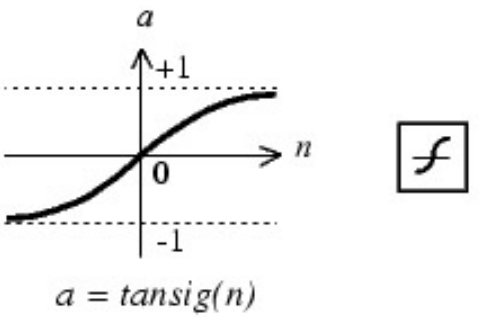

Figure 3.4 A Sigmoid Transfer Function of an ANN Layer (Demuth *et al*., 2006)

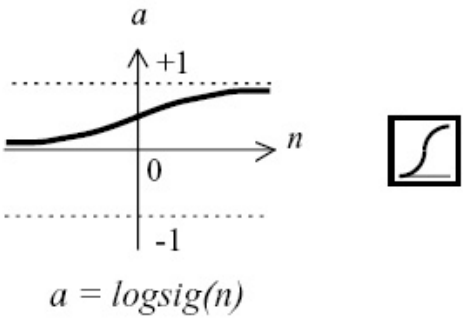

Figure 3.5 A Logarith-Based Transfer Function of an ANN Output Layer (Demuth *et al*., 2006)

The number of neurons in the sigmoid layer is required to exceed that of the inputs (Demuth *et al*., 2006). By trial-and-error, one sigmoid layer with 50 neurons is determined as the best for the ANN models. The number of inputs to the network is determined by the problem at hand, and the number of neurons in the output layer is based on the number of outputs required. However, the number of layers between network inputs and the output layer, as well as the sizes of the layers, is to be determined by the analyst. Although in principle, a network with just one hidden layer can be taught to approximate any continuous functional mapping, the experiments in this study have shown that mappings from one real space to another are often better learned by networks with two hidden layers. However, in the present study, it is found that one hidden layer is enough for neural network generalization. According to Demuth *et al*. (2006), this often happens in the feed-forward neural network. Hence, more layers seem redundant for training purposes.

The inputs that the ANN model receives, include movement volumes, facility type, link capacity, and number of lanes, which are typically available from a planning model. Because TRANSYT-7F considers through and right-turning volumes in the same lane group, 12 movements at an intersection are reduced to eight movement volumes, two facility types, two numbers of lanes, and two link capacities that are fed to the ANN for the output of a delay estimate for an approach movement at a four-leg intersection:

D_c : movement delay of the studied approach (s)

$v_{11\sim41}$: through volume of four approaches (vph)

$v_{12\sim42}$: left-turn volume of four approaches (vph)

$c_{1\sim2}$: link capacity of two links (vphpl)

$f_{1\sim2}$: facility type of two links

$l_{1\sim2}$: number of lanes of two links

Figure 3.6 illustrates the spatial relationship of movement volumes, link capacities, number of lanes, facility type, and the corresponding delay estimate. Facility type implies the intersection categories. And link capacity, together with movement volumes, has significant implications on density of traffic on the link. In the preliminary stage of developing the ANN model, all variables that are available from traffic assignment and seemingly related are incorporated so that the ANN model may fully apply all potential information for its training, although a certain variable may be redundant with the other one if simultaneously serving as the inputs of the ANN model.

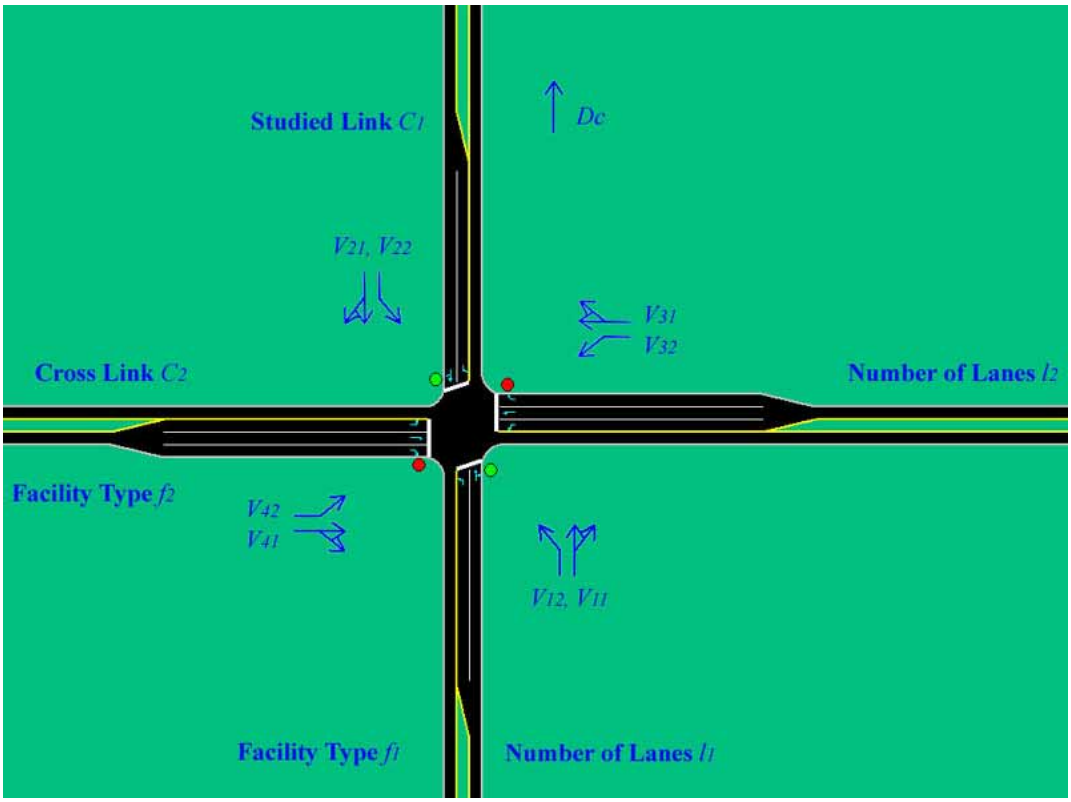

Figure 3.6 Spatial Relationships of the Input Variables for the Delay Model

Using the above information, a delay model should reflect the relationship between delay and the variables that describe volumes, capacity, and facility types of the two crossing streets, as described in a general form in Eq. 3.2:

$$D_c = F(v_{11\sim41}, v_{12\sim42}, c_{1\sim2,} f_{1\sim2}, l_{1\sim2}) \quad \text{Eq. 3.2}$$

The functional relationship, *F*, may be expressed using different modeling techniques. A set of multiple linear regression models is developed to serve as a benchmark to evaluate the proposed ANN delay model. The regression models take the form below:

$$D_c = b_0 + b_1.v_{11} + b_2.v_{12} + b_3.v_{21} + b_4.v_{22} + b_5.\ v_{31} + b_6.v_{32} + b_7.v_{41} + b_8.v_{42} + b_9.c_1 + b_{10}.c_2 + b_{11}.f_1 + b_{12}.f_2 + b_{13}.l_1 + b_{14}.l_2 \quad \text{Eq. 3.3}$$

where

$b_{0\sim14}$ = regression coefficients

Due to the correlation among these variables, tests of multicollinearity are conducted. Muliticollinearity, in practical terms, means that the predictor variables are linearly related with each other, which may cause serious numerical and statistical difficulties in fitting a regression model (Mason *et al.*, 1975). Variance inflation factors (VIF) are a direct measure of multicollinearity. A predictor with VIF larger than 10 usually needs to be removed from the MLR models. In Table 3.2, VIF of c_1, c_2, l_1, and l_2 are very large, which implies significant linear relationship between link capacity and number of lanes. Therefore, the variable representing the two link capacities (c_1 and c_2) are discarded. The remaining 12 predictor variables demonstrate no more high VIF (Table 3.3), and therefore are taken as the inputs of the ANN delay model. A comparison

of the performance of the MLR models and the ANN model will be discussed in Chapter 4.

Table 3.2 VIF of the Preliminary MLR Delay Model ($b_{0\sim14}$: linear coefficients)

VIF	Through Movement	Left Turn Movement
b_1	2.65	1.60
b_2	2.65	1.60
b_3	1.60	2.65
b_4	1.60	2.65
b_5	2.63	1.60
b_6	2.63	1.60
b_7	1.60	2.63
b_8	1.60	2.63
b_9	410.52	410.54
b_{10}	377.06	377.08
b_{11}	20.24	20.24
b_{12}	288.27	288.28
b_{13}	23.82	23.82
b_{14}	252.43	252.44

Table 3.3 VIF of the Ultimate MLR Delay Model

VIF	Through Movement	Left Turn Movement
b_1	2.55	1.55
b_2	2.55	1.55
b_3	1.55	2.55
b_4	1.55	2.55
b_5	2.55	1.55
b_6	2.55	1.55
b_7	1.55	2.55
b_8	1.55	2.55
b_{11}	7.05	7.05
b_{12}	4.67	4.67
b_{13}	6.91	6.91
b_{14}	4.55	4.55

As to the back-propagation learning rule used to train the ANN models, the Levenberg-Marquardt algorithm is usually the fastest of several training algorithms implementing the back-propagation learning rule. It provides a memory reduction feature when the training data set is large. Several other training algorithms are also considered.

Requiring no line search, the scaled conjugate gradient algorithm is a good general purpose training algorithm. The Bayesian regularization algorithm is a modification of the Levenberg-Marquardt training algorithm used to produce networks that generalize better. It reduces the difficulty of determining the optimum network architecture.

The performances of ANN models using the above three training algorithms were examined for all intersection types. It was found that the scaled conjugate gradient usually produced a relatively better fit compared to the other two, based on three evaluation criteria: the regression R-squared value, the root-mean-square error (RMSE), and the percent root-mean-square error (%RMSE). The RMSE, also known as the standard error, represents the average error of model predictions. The %RMSE is a statistic indicating the percentage of the average expected error in the actual value, and it has been adopted by the FDOT as a criterion in calibrating travel models. The formulas for computing the RMSE and the %RMSE are given below:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}\left(x_i^c - x_i^v\right)^2}{n-1}} \qquad \text{Eq 3.4}$$

$$\%RMSE = \frac{RMSE}{\sum_{i=1}^{n} x_i^c \Big/ n} \qquad \text{Eq 3.5}$$

where

x^c = delay estimates from TRANSYT-7F simulation (seconds per vehicle)

x^v = delay predictions by ANN delay model (seconds per vehicle)

n = number of ANN inputs (simulated scenarios)

For all types of intersections, the ANN delay models that apply the scaled conjugate gradient algorithm usually demonstrate relatively better performance than the other two. Table 3.4 shows the performance statistics of the ANN delay model for intersection type 2241. Because of the relatively lower %RMSE, a multilayer feed-forward ANN architecture with a scaled conjugate gradient algorithm is applied in developing the ANN delay models.

Table 3.4 Performance Statistics of Three Training Algorithms for 2241 Type of Intersection

Training Algorithm	Regression R-squared	RMSE	%RMSE
Scaled Conjugate Gradient	0.855	8.34	18.77
Bayesian Regularization	0.767	10.47	22.23
Levenberg-Marquardt	0.807	9.55	20.27

3.4 The Combined System

The well-trained ANN delay model is designed to interact with a standard static traffic assignment model. A typical assignment model builds paths based upon link costs (travel time in this study) and assigns trips to those paths for each origins and destinations. After all origins and destinations have been processed, link costs are updated based upon the level of congestion on each link. The entire path and assignment process is repeated until termination criteria are reached. The volumes from each assignment are combined to form a weighted assignment. Different criteria are applied to determine if enough iterations have been performed. The input format of the ANN delay model must be compatible with the output format of the traffic assignment procedure. Likewise, the traffic assignment also has to perform based on the ANN delay model's output. The delay model and the traffic assignment call for one execution of each other during every iteration of the combined system, as shown in Figure 3.7. The delay estimates, serving as

a data source shared by both the delay model and the traffic assignment, require a compatible format for the data exchange interface connecting the delay estimates and the traffic assignment.

When traffic assignment incorporates control delays, it is always an indispensable issue to pursue the convergence, which is one of the major challenges in this research.

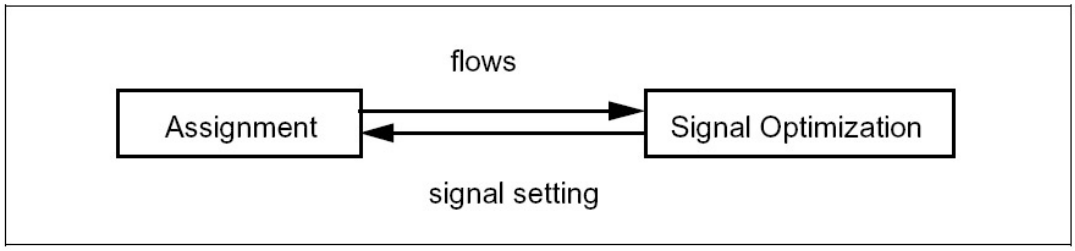

Figure 3.7 Iterative Signal Plan Optimization and Assignment Procedure

The combined system has an obvious dilemma when attempting to arrive at an optimized solution. The ANN delay model implicitly optimizes the signal settings based on the received assignment volume. The resulting intersection delays then cause the traffic assignment to re-calculate the updated route costs and re-assign the traffic onto the network. Thus, in the following iteration, once again will the ANN delay model have to re-adjust the underlying signal settings, as well as the traffic assignment, to re-optimize the assigned volumes. As Figure 3.7 shows, such a simple iterative process unable to reach convergence more often than not continues to an endless oscillation (Lee and Machemehl, 2005). Therefore, an optimization search algorithm is required to prove that the combined model may always reach a solution regarding equilibrium for both the delay model and the traffic assignment.

At the final stage of the study, the performance of the combined system needs to be evaluated based on a two-fold consideration. First, to validate the convergence of the iterations, the following criteria may be referred to (Horowitz, 1989):

- Premature termination of iterations leads to significant error.
- A solution is replicable for a given problem.
- Different starting points should reach the same solution.

Note that Horowitz's criteria (1989) are essentially more specific, yet otherwise identical to what Heydecker (1983) recommended. Second, the validated convergence must be arithmetically demonstrated through a general criterion of the system's convergence tests that accepts no or only a negligible difference in the network flow patterns for two consecutive iterations. This criterion may be explained by Eq. 3.6 (Sheffi, 1985), which indicates that the signal settings may not be further optimized to reduce the difference of the network flows of two consecutive iterations.

$$\frac{1}{A}\sum_{a}\frac{\left|x_a^{n+1}-x_a^{n}\right|}{x_a^{n+1}} \le k_1 \qquad \text{Eq. 3.6}$$

where

A = number of links in the network

k = a predetermined constant.

Alternatively, another criterion that is based on the change in flows may be used. For instance, the iteration may terminate if

$$\frac{\sqrt{\sum_{a}(x_a^{n+1}-x_a^{n})^2}}{\sum x_a^{n}} \le k_2 \qquad \text{Eq. 3.7}$$

These two equations have different considerations. Eq 3.6 calculated the network flow change averaged onto the whole network, whilst Eq 3.7 solely considers the overall change of network flows in consecutive iterations.

4. ANALYSIS OF RESEARCH RESULTS

4.1 ANN Delay Model Performance Analysis

4.1.1 Data Preparation

The data for training the ANN delay model are extracted from the simulations performed via TRANSYT-7F. In addition to the volumes and delays, the facility type and the capacity of the four cross roads are also needed as input to the ANN models. The capacities are assumed based on the approach's facility type, area type, and lane number defined in the user's manual of FSUTMS (FDOT, 1997) and are given in Table 4.1. Take intersection type 2322 as an example. The first two digits respectively represent a link's facility type (2) and number of lanes (3), and the last two digits give the same information for the crossing link (facility type 2 with 2 lanes).

Table 4.1 Network Link Capacity for Roads of Different Facility Types and Lanes

Intersection Type	Base Capacity (vphpl)
2322	755
2222	750
2241	750/530
3141	592/530
4141	530

To create the training datasets, a total of 150 combinations of through volumes are created. To ensure that the volumes reflect the real traffic conditions, the historical traffic counts from the Florida Traffic Information (FTI) 2004 CD were examined. The peak-hour traffic counts were extracted from the 88 Portable Traffic Monitoring Sites (PTMS) for intersections of facility type 2 in all directions in the Gainesville urban area (Figure 5.1). The peak-hour traffic counts per lane were found to approximately follow a normal distribution curve, with a mean μ = 462.72 and a standard deviation σ = 135.83, as shown

in Figure 5.2. Therefore, the normal distribution was for intersections of facility type 2 when creating the simulation scenarios. Because no PTMSs are found for undivided arterials (facility type 3) and local collectors (facility type 4), it is assumed that the peak-hour traffic for these facility types also follow a similar normal distribution. The μ and σ are slightly adjusted so that [μ-3σ, μ+3σ] properly contains the assumed base capacities. This range represents the 99% confidence interval of a normal distribution. In other words, the volumes selected will fall within this range with a 99% probability. Table 4.2 gives the normal distribution parameters for different intersection types.

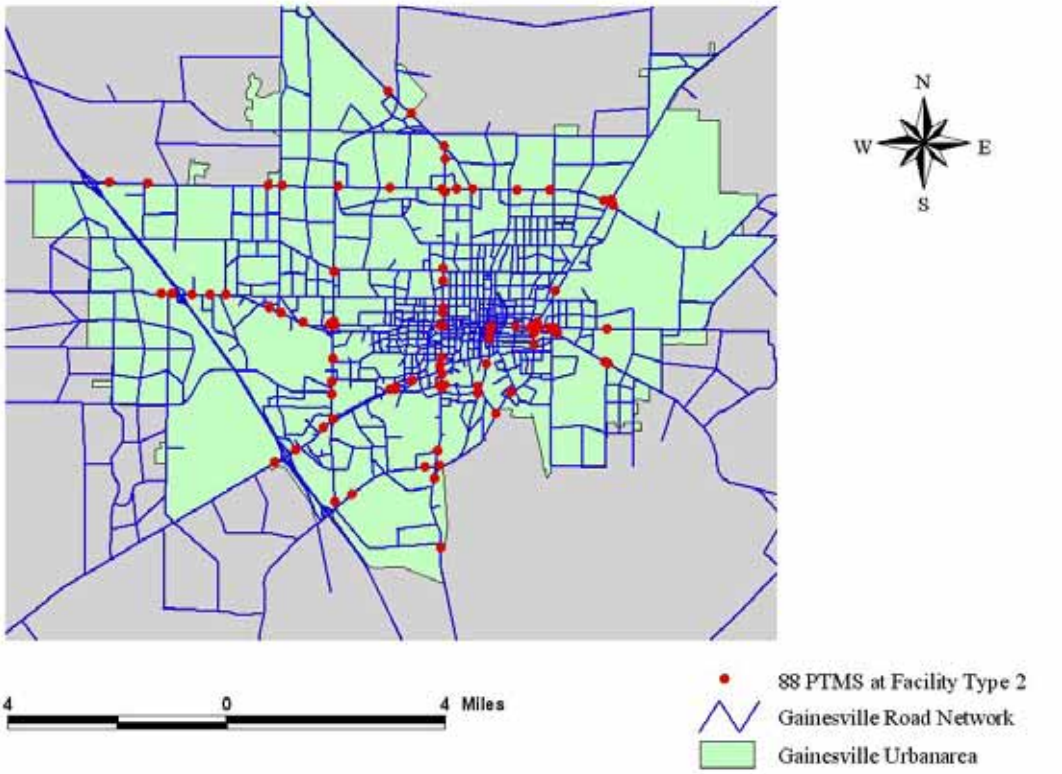

Figure 4.1 Locations of 88 PTMS for Divided Arterials in the Gainesville Urban Area

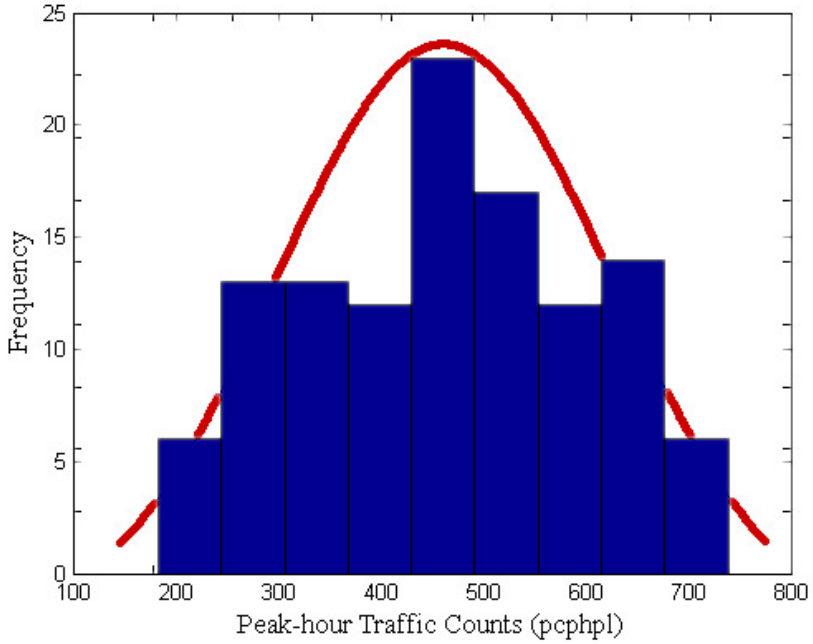

Figure 4.2 Distribution of Peak-hour Traffic Counts of 88 PTMS for Divided Arterials

Table 4.2 Normal Distribution Parameters for Volumes by Intersection Type

Intersection Type	μ	σ
2322	462.72	135.83
2222	462.72	135.83
2241	462.72	135.83
3141	400.00	125.00
4141	400.00	125.00

The 150 combinations of through traffic volumes were created by sampling from normally distributed volumes within the range of [μ-3σ, μ+3σ] using random numbers. For all types of intersections in this study, the heaviest through volume simulated was approximately 1.4 times that of the corresponding approach's base capacity.

The simulation scenarios also require left-turning volumes. Assuming seven turning ratios for generating different combinations of turning traffic, which are 10%, 15%, 20%, 25%, 30%, 35%, and 40%, the number of combinations of four turning ratios is 7^4 = 2,401. However, because an ANN model does not need to be trained with all

possible situations to make predictions, not all of the 2,401 turning ratio combinations have to be considered in training. It was found that the ANN model did not seem to be sensitive to small variations of turning ratios under certain through traffic conditions. For instance, an intersection with a turning ratio combination of [0.1 0.1 0.25 0.3] has no significantly different traffic situation compared to one with a turning ratio combination of [0.1 0.1 0.2 0.3] when the four through volumes are fixed. Therefore, only 12 turning ratio combinations are randomly selected from the 2,401 to combine with the through volumes and create the ANN training scenarios. The performance of ANN delay models has proven that this proportion is adequate for the analysis. As a result, there are 1800 (= 150 × 12) scenarios simulated for each intersection. Extracted from these 1800 scenarios are respectively 7200 (= 1800 × 4 legs) samples for either through or left-turn movements of each intersection type. During the training of the ANN delay model, the input data are divided into three groups: training data, validation data, and testing data. All of the samples are employed to train the ANN delay model except that approximately 15% of the 7200 samples are randomly allocated to test the ANN delay model's accuracy at a later stage. The ANN training employs "supervised learning," that is, the training process is simultaneously supervised by the scaled conjugate gradient training algorithm that applied validating data to the trained ANN to correct potential overfitting. After the training of the ANN model, the testing data are used to evaluate the ANN model's performance.

4.1.2 Evaluation of the ANN Delay Model

The accuracy of the ANN delay model' delay estimates is an essential condition to ensure the overall performance of the combined system. Note that there are in fact two

ANN submodels in one architecture. One is to estimate delays for through and right-turning movements, and the other is for left-turning movements. Unless necessary, the entire model including its submodels will be referred to as the ANN model. It may be observed that the ANN delay model usually demonstrate three characteristics. The delay estimates for the larger intersections (facility type 2) are less accurate than those for the smaller intersections (of facility type 3 or 4). Regardless of the intersection types, the ANN model always perform better estimating delays for the through movement than for left-turning movements. That is, the left-turning delays are more difficult to estimate than the through movement delays. The "normal" traffic loadings result into better delay estimates than "rare" traffic conditions. For example, the conditions from both "tails" of the normal distribution tend to lead more erroneous delay estimates than those from middle of the "bell".

To quantify the ANN model's capability precisely, around 15% of the 7,200 input samples for ANN are randomly selected as testing data without duplicating the training or validation data. There are two intuitive methods to evaluate the ANN model training. One is to plot the error of predictions. The other is to fit a linear regression analysis between the predictions and the estimates. Descriptive statistics, such as regression R-squared, RMSE, and %RMSE, are calculated (Eqs 3.4, Eq 3.5). The %RMSE of delay estimates are controlled below 26%. The independent variables of the regression models are the eight volumes, two numbers of lanes, and two facility types (Table 3.3) and the dependent variable is the delay. The regression coefficients, F-values, P-values, and performance statistics of every regression model are provided in the appendix. The same training and testing data sets are used to develop and evaluate the MLR models

introduced in Chapter 3. The regression R-squared, RMSE, and %RMSE of the ANN delay model and the MLR models are compared in Table 4.3, which shows that the ANN delay model is superior to the regression models with frequent lower RMSE and %RMSE. Although the regression models often present R-squared values that are close to or even better than those of the ANN delay model, R-square values are not taken as the decisive standard judging the goodness of a delay model. R-squared values merely identify how well a linear fit is and help the understanding of the performance of the delay models. A high R-square value may come with a high %RMSE if the samples spread widely but symmetrically along the fitted line. That is the reason why lower %RMSE and RMSE are the major factors judging the delay models. Figure 4.3 to 4.7 illustrate the linear fit and the prediction errors for through and left-turning movements of all five intersection types. The linear fit, take Figure 4.3 for example, is between the ANN delay outputs and TRANSYT-7F simulated delay estimates (targets). The R-squared value is 0.578, which represent best linear relationship between the ANN outputs and the simulated delays. Figure 4.3 exhibits the ANN delay outputs and the simulated delays in different color so that the accuracy of the ANN predictions, delays in this case, may be easily identified. In the figures, a code is used to indicate the type of an intersection, as well as the movement studied, as follows:

Table 4.3 Comparison of Performance Statistics of Two Categories of Models

Intersection Category	Statistics	ANN Models	Regression Models	ANN Models	Regression Models
		Through	Through	Left-turning	Left-turning
2322	R-Squared	0.729	0.705	0.609	0.573
	RMSE	7.13	11.36	17.17	19.78
	%RMSE	21.36	22.95	25.45	28.58
2222	R-Squared	0.754	0.769	0.685	0.647
	RMSE	7.25	12.39	13.92	17.58

	%RMSE	16.27	20.73	22.13	24.97
2241	R-Squared	0.791	0.818	0.662	0.628
	RMSE	5.49	8.17	9.78	13.67
	%RMSE	17.28	18.51	18.73	21.73
3141	R-Squared	0.734	0.771	0.638	0.657
	RMSE	4.69	6.14	9.71	10.75
	%RMSE	14.89	15.84	22.09	22.75
4141	R-Squared	0.655	0.717	0.613	0.576
	RMSE	4.67	5.67	5.51	6.46
	%RMSE	19.23	20.87	16.24	18.06
Overall	R-Squared	0.787	0.746	0.749	0.629
	RMSE	5.1	11.245	12.62	18.56
	%RMSE	16.76	25.23	23.80	26.61

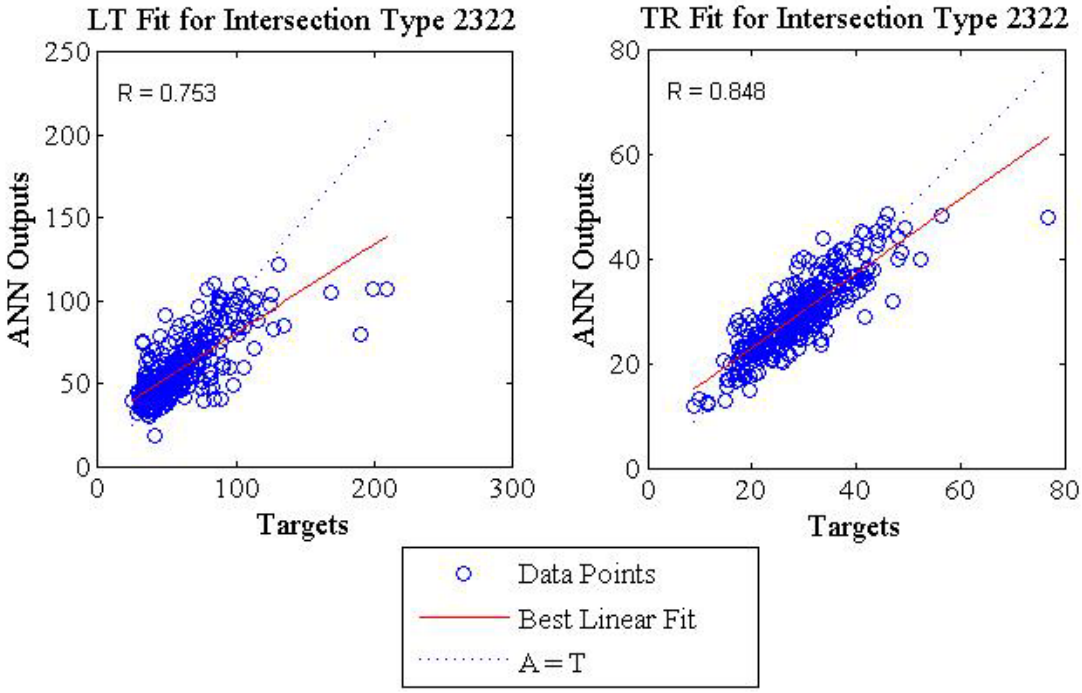

Figure 4.3 Linear Fit of ANN Delay Estimates and Targets for Intersection Type 2322

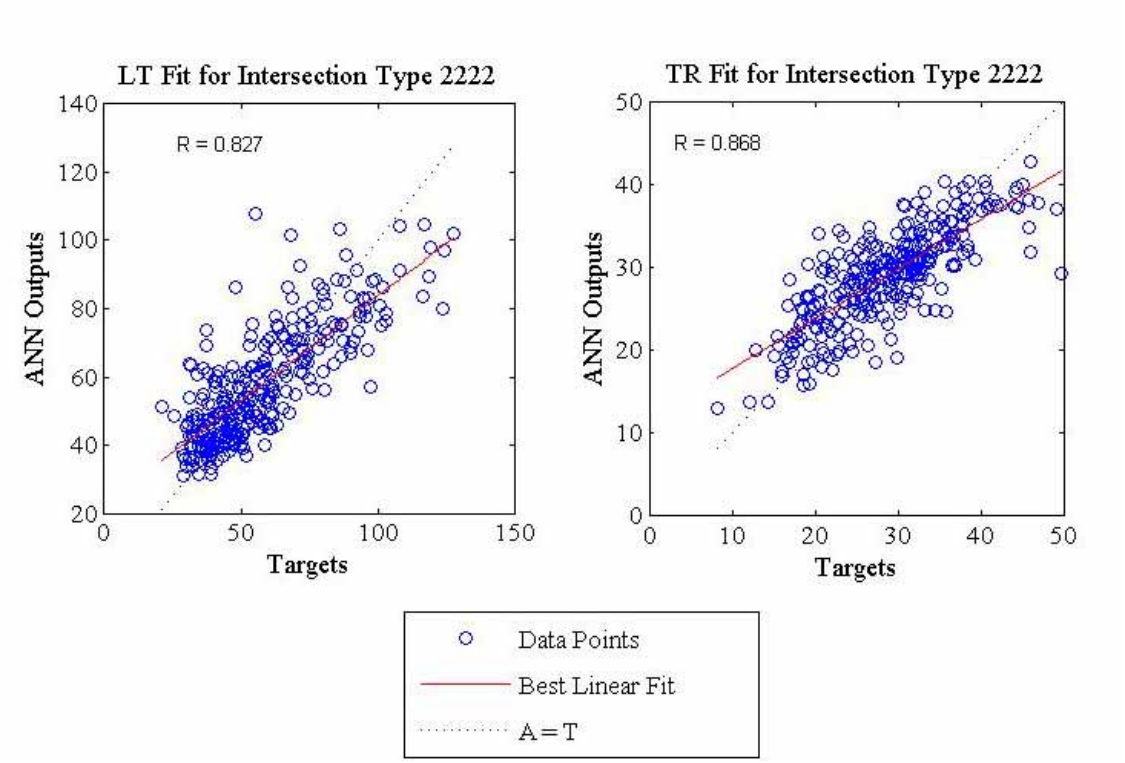

Figure 4.4 Linear Fit of ANN Delay Estimates and Targets for Intersection Type 2222

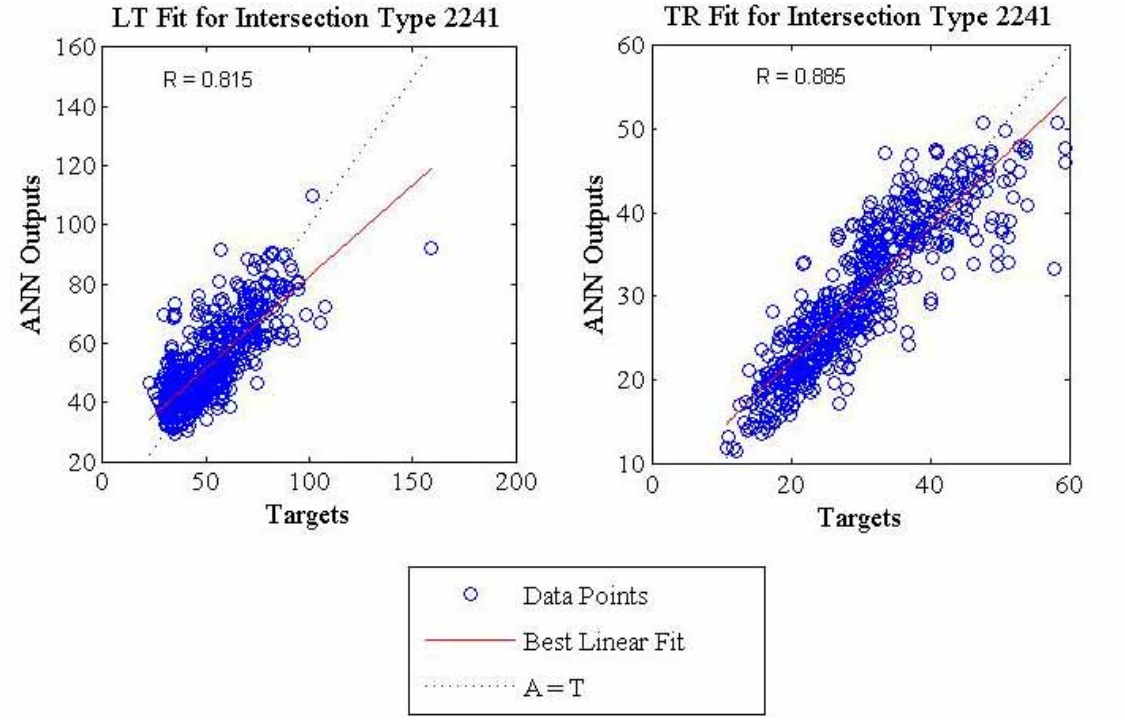

Figure 4.5 Linear Fit of ANN Delay Estimates and Targets for Intersection Type 2241

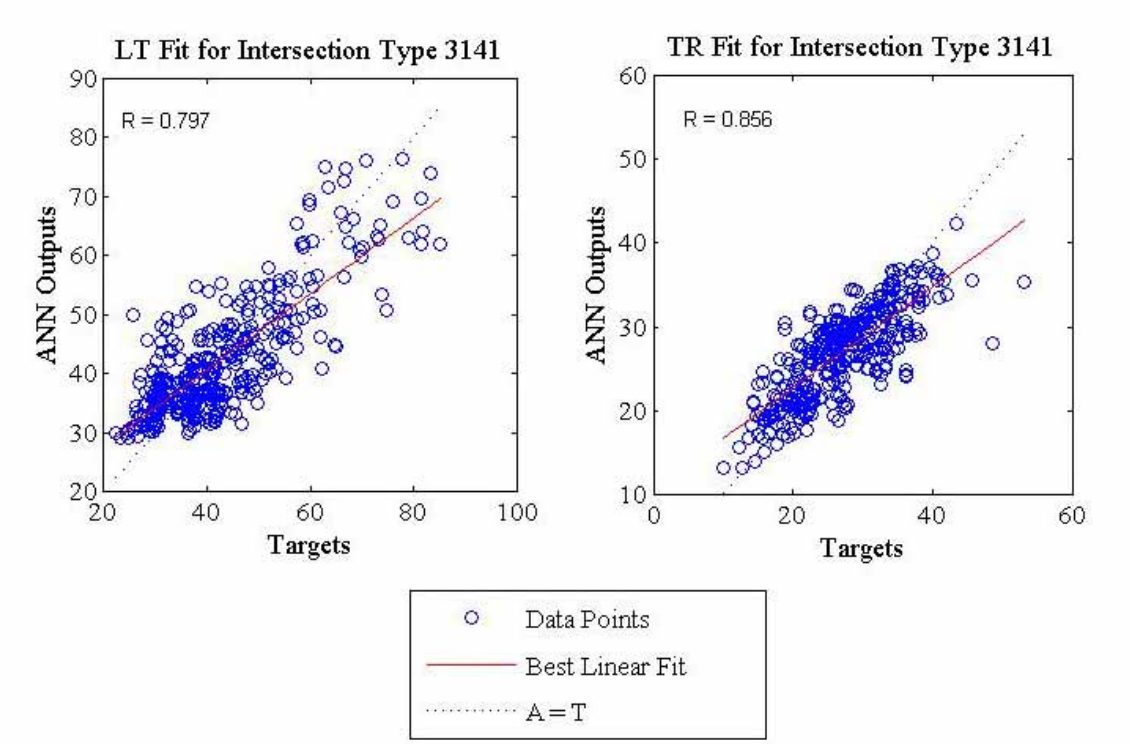

Figure 4.6 Linear Fit of ANN Delay Estimates and Targets for Intersection Type 3141

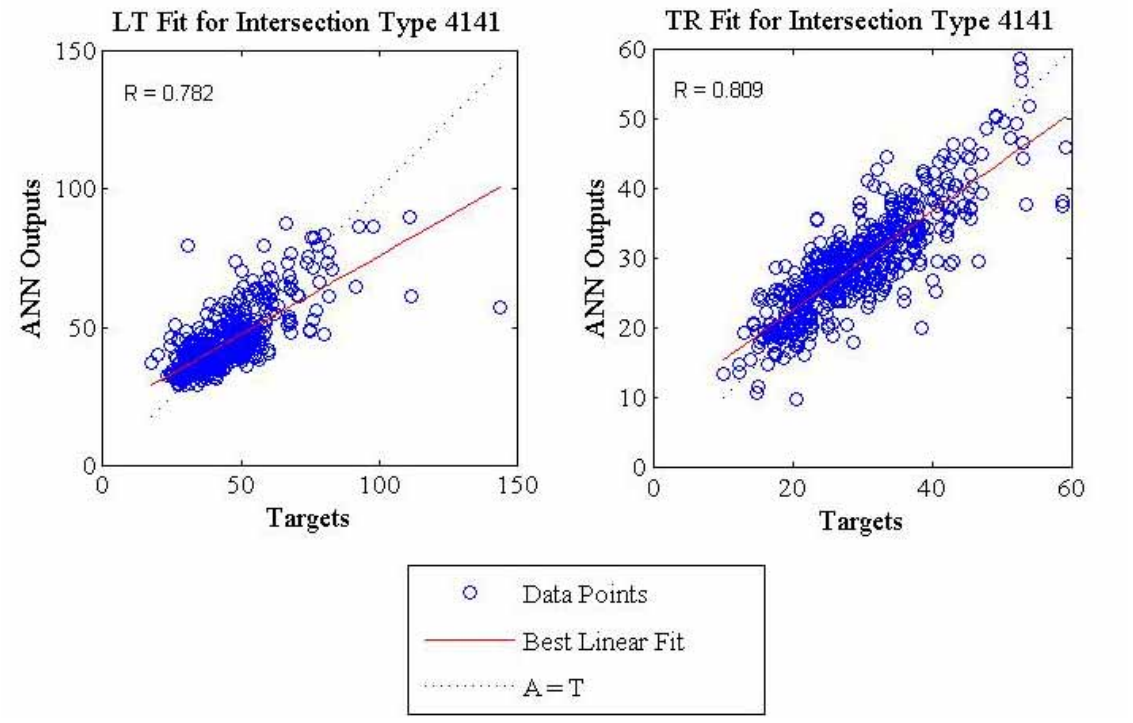

Figure 4.7 Linear Fit of ANN Delay Estimates and Targets for Intersection Type 4141

Another major concern about the delay model is the computing time. It should be short enough to be practical for application. Table 4.4 shows the computing costs for different sizes of experimental networks when using the delay model. The mean of computing time is averaged based on 300 applications of the delay model.

Table 4.4 Average Computing Time of Various Networks

	Number of Intersections	
Network Size	4	20
Average Computing Time (sec)	1.91	4.53

Further analysis is conducted on the characteristics of the prediction errors of the ANN delay model. Figure 4.8 shows the distributions of prediction errors by equal intervals and the trend of the mean absolute errors (MAE) for each intersection type. MAE is calculated to examine the size of forecast errors (Eq. 4). MAE assumes that the severity of a prediction error increases in a linear manner (e.g., a 2% error is twice as serious as a 1% error). For simplicity, only the through and right-turn scenarios are analyzed.

$$MAE = \frac{\sum_{i=1}^{n} \left| x_i^c - x_i^v \right|}{n} \qquad \text{Eq 4.1}$$

where

x^c = delay estimates from TRANSYT-7F simulation (seconds per vehicle),

x^v = delay predictions by ANN delay model (seconds per vehicle), and

n = number of testing inputs

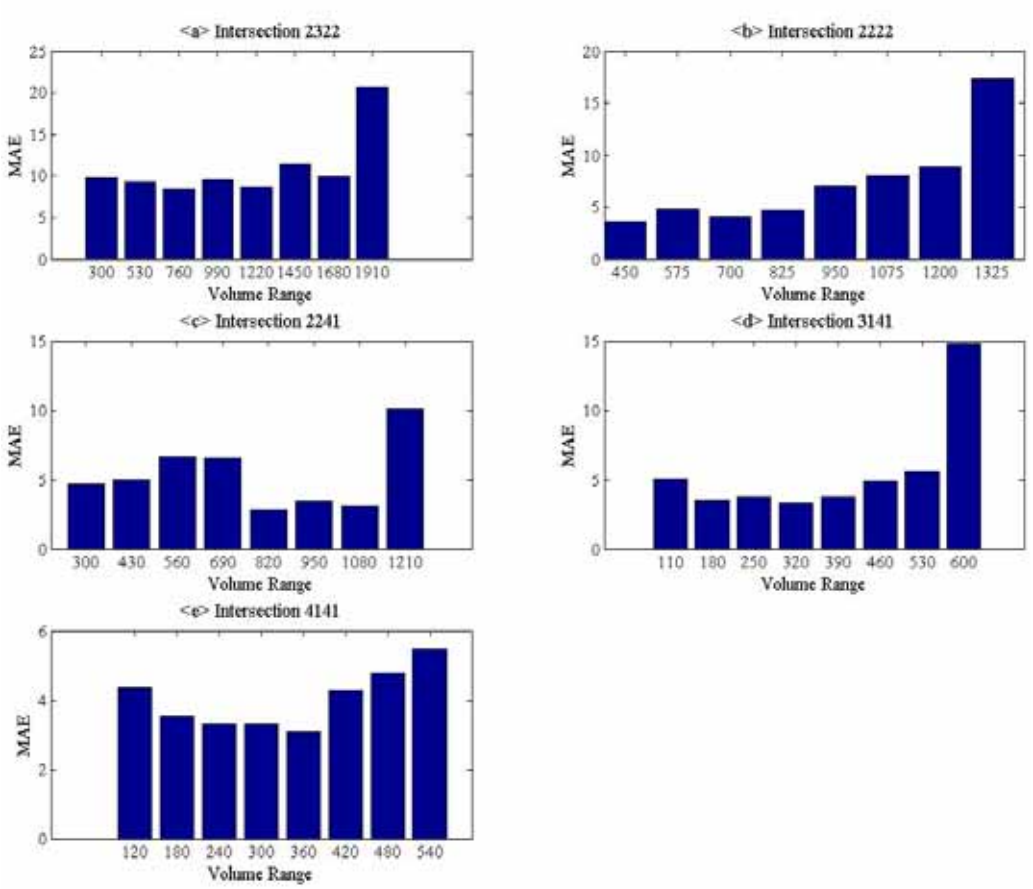

Figure 4.8 Distributions of MAEs for Different Volume Ranges

It may be observed from Figure 4.8 that larger MAEs are encountered more frequently when the volumes reach or exceed the capacity of an intersection. The model performance is far less reliable if the intersection is seriously oversaturated. This problem is less severe for intersection type 4141, which has an acceptable MAE error even under heavy traffic conditions. One possible cause is that this study has employed single period analysis in signal timing optimization in TRANSYT-7F, which may not ensure realistic delay estimation under severely oversaturated conditions.

4.2 The Traffic Assignment Model

In this research, the standard UE static traffic assignment is applied using Cube, a travel demand modeling software package by Citilabs and currently adopted in the State of Florida. The study network is constructed in Cube. An initial turn delay table and an OD matrix with all movement volumes less than 1.3 times the approach capacity are also generated and used as the initial input for traffic assignment. A total of 24,048 trips are assigned to the study network. A typical assignment model first builds the shortest paths for each origin-destination (OD) zone pair based on link travel costs. Trips are then assigned to these paths based on the shortest paths. After all OD pairs have been processed, link travel costs are updated based on the volumes on each link. Then the processes of path building and assignment process repeat. The volumes from each assignment are combined with those obtained from the previous iteration to form a weighted assignment. This process continues until some criterion for termination is met. Different criteria may be used to determine when enough iterations have been performed. In this study, the link travel cost is updated in every assignment, not only by the variation of the link congestion level, but by the changes of the intersection delays as well. For a specific link on the studied network, the total link travel cost consists of two parts:

$$T = T_c + T_l \qquad \text{Eq 4.2}$$

where

T = total link travel cost,

T_c = movement delays of a upstream intersection prior to a destination node, and

T_l = link travel time for the link connecting the upstream intersection and the destination node.

Within one iteration of traffic assignment, T_c is provided by the ANN delay model for every turning movement of an intersection. Based on identification number of intersections, the ANN delay model is able to determine if T_c is resulted from a delay estimate for through traffic, right-turn movements, or left-turns. T_l is, on the other hand, calculated by the Cube traffic assignment model in the form of Bureau of Public Roads (BPR) volume-delay function using Eq. 4.3:

$$T_l = t_u[1 + k(v/c)^m] \quad \text{Eq 4.3}$$

where

t_u = free flow travel time, a constant (undersaturate condition),

k = saturation weight factor (default value, 0.15),

m = saturation power factor, (default value, 4).

The goal of the traffic assignment is to find the minimum total travel cost of the network:

$$\min(Z) = \sum_{i=1}^{n} T_i \quad \text{Eq. 4.4}$$

4.3 The Combined System

As mentioned in Chapter 3.3, the combined system of the delay model and traffic assignment is an iterative process, as illustrated in Figure 4.9. The user equilibrium (UE) traffic assignment is performed by Cube, which produces link volumes. These link volumes and turning movements are provided to the ANN delay model, which updates control delays based on the assigned volumes. These delays are sent back to *Cube* for the next run of the traffic assignment. This process repeats until the solution of the combined

system converges. MATLAB is capable of reading and writing the inputs and outputs of traffic assignment performed by Cube, thus the data and intermediate results may smoothly flow within the combined system. In Figure 4.9, the box labeled as Frank-Wolfe Algorithm is the key in the combined system to produce a convergent solution. This algorithm is described below.

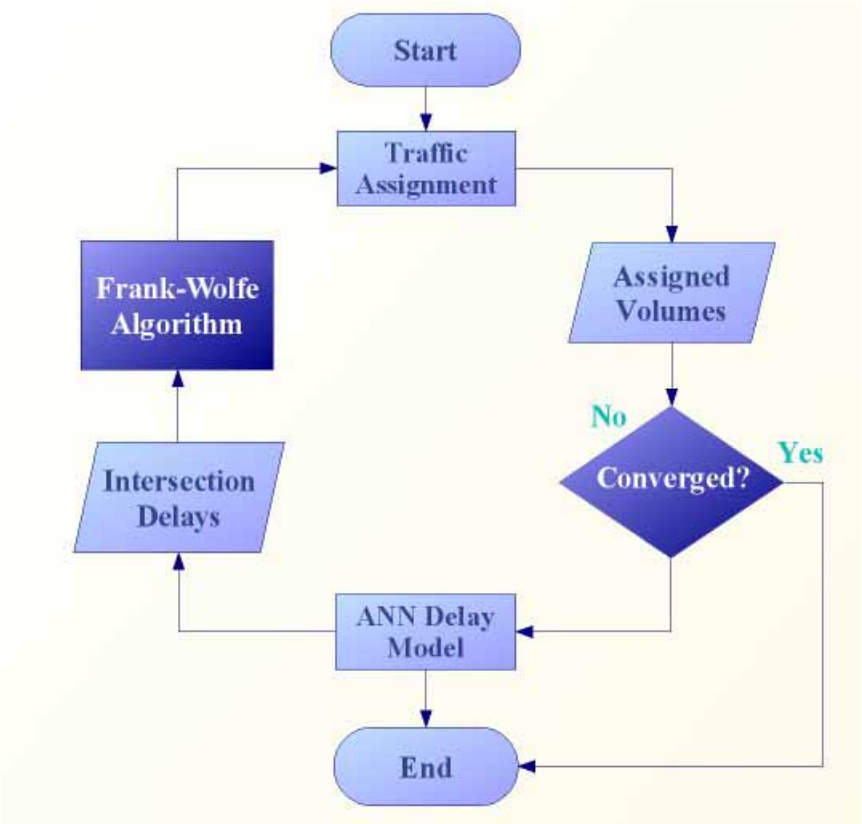

Figure 4.9 Logical Loop of the Combined Model

To facilitate the convergence of the solutions, a search algorithm is required, since mathematical programming cannot guarantee convergence (Sheffi, 1985). In this study, the ANN delay models cannot provide reliable gradient information to allow the application of traditional optimization techniques. An alternative is to employ a meta-heuristic algorithm such as direct search, which searches for a solution without using derivatives explicitly. Although the global optimum is not guaranteed, the direct search

method is often able to find satisfying solutions for many engineering applications. In fact, for practical applications, it is often the computing time, simplicity, and fast arrival at a local optimum that are the primary considerations To reach a satisfying solution, the search algorithm has to demonstrate a clear converging pattern of solutions to reach at least a local optimum.

A common choice for a search method is the method of successive averages (MSA), known also as a simplified transformation of the Frank-Wolfe algorithm (Sheffi, 1985; Ortuzar and Willumsen, 2001). The method is based on a predetermined step size along the descent direction. The procedure is as follows:

1. Initialization. Perform an equilibrium assignment based on a set of initial travel costs T. This generates a set of network flows x^n_a, which represents the flow on link a. Set $n = 1$.
2. Updating travel cost for each link a. Set $T_a = t_a (x^n_a)$.
3. Search direction finding. Perform an equilibrium assignment based on the current set of travel costs T_a, which yields an auxiliary network flow pattern y^n_a.
4. Determination of step size. Obtain the new flow pattern setting, set $\alpha = (1/n)$.
5. Derive a new set of flow. Set $x^{n+1}_a = x^n_a + (1/n)\ (y^n_a - x^n_a)$
6. Checking Convergence criterion. Examine the difference between the network flows of successive iterations. If the difference is small enough, convergence is considered to be attained, stop. Otherwise, set $n = n+1$ and go to step 1.

The original Frank-Wolfe algorithm, different from the MSA in step 4, optimizes move size factor α using mathematical programming methods so that the convergence of the combined system may be reached more efficiently. The goal of the Frank-Wolfe

algorithm is to find a downhill direction and proceed to step 5. However, because at step 5 the delay estimates are results from an ANN delay model that gives no gradient information, α cannot be solved using traditional mathematical programming, which requires gradient or higher derivatives of the objective function. Therefore, the mesh adaptive direct search (MADS) algorithm, a direct search method requiring no gradient or derivatives, is applied to seek the optimized α in step 4. By minimizing the potentially non-smooth function represented by the delay model, the MADS allows local exploration in a dense set of directions in the space of optimization variables (Audet and Dennis, 2006). As a direct search algorithm that computes a sequence of points that approach the optimum gradually, MADS searches at each step for a set of points, called a mesh, around the current solution point — the point computed at the previous step of the algorithm. The mesh is formed by adding the current point with α (a scalar value) multiplied by (y^n_a - x^n_a) as step 5. If MADS finds a point in the mesh that improves the objective function at the current point, the new point becomes the current point at the next step (MathWorks, 2004).

MADS is used to substitute for step 4 above to help the Frank-Wolfe algorithm converge faster in the combined system. It attempts to find a scalar, α, that solves the program:

$$\min z(x^n + \alpha(y^n - x^n)) = \sum_a \int_0^{x+a(y-x)} t_a(x)d\omega \qquad \text{Eq. 4.5}$$
$$\text{subject to } 0 \le \alpha \le 1$$

The α obtained will be used in computing the weighted flow, $x^{n+1}_a = x^n_a + \alpha (y^n_a - x^n_a)$. To demonstrate the performance, the MADS Frank-Wolfe method (MFW) is,

respectively, applied to two networks shown in Figures 3.2 and 3.3. The smaller network has 12 links and four intersections with 5,963 vehicle trips. The larger network has 49 links and 20 intersections with 24,048 vehicle trips. The performances of both networks confirm that the Frank-Wolfe algorithm helps the combined system converge to an identical network flow pattern, in spite of the choice of the MFW or MSA method. However, it is observed that the computing time to reach an identical solution increases substantially with the network size (see Table 4.5). Note that the combined system often reaches a different solution if limited time is allocated, although a trend of convergence is always demonstrated. These different solutions are various local optima that meet the stopping criterion as given in Eq 3.4 or Eq 3.5. Empirically speaking, among these local optima, there is one with a minimum link volume difference, which is the best solution that may be reached repeatedly if the combined system runs long enough. This best solution may be repeatedly reached if enough time is given, no matter what the initial network flow pattern or intersection delays are. However, as discussed in Chapter 2, there is no theoretical proof as to whether the solution is the global optimum or not.

For comparison, the MSA method that applies a deterministic weight ($\alpha = 1/n$) to each auxiliary flow is also applied to the combined system, and it is able to always reach the same solution but with many more iterations than the MFW method. For the simple network shown in Figure 3.3, with the same demand, the MFW may converge to the same network flow pattern after 15 iterations, which is much faster in terms of number of iterations than the MSA, which takes 61 iterations to reach the same convergence level. The performance on the large network further confirms the advantage of the MFW method, which takes 32 iterations to reach a solution versus 137 iterations for the MSA.

The networks with traffic loading ranging from heavy to light are all solved using a random initial OD matrix. The MFW demonstrates a consistent advantage over the MSA in all situations. Horowitz (1989) gives a set of criteria for a satisfying solution from a travel demand model:

- Premature termination of iterations leads to significant error.
- A solution is replicable for a given problem.
- Different starting points should reach the same solution.

The combined system, starting with varied initial signal settings, is always capable of converging to one identical network flow pattern with enough time allocated. Although the uniqueness of a solution cannot be guaranteed, the convergence is thus validated if the criteria of Horowitz (1989) are applied. Note that the Eq. 3.6 and 3.7 (Sheffi, 1985) are also employed to judge the convergence. Both equations indicate that the signal settings may not be further optimized to reduce the difference of the network flows of two consecutive iterations. There is a minimized error between successive network flow patterns at convergence. The resulting assigned volume and the previous iteration's network flow pattern are approximately identical if judged by the stopping criteria. During the iterative procedure, the oscillation, though gradually dampened, cannot be completely eliminated because of the nature of the problem. This may be attributed partly to the ANN delay model merely estimating delays at a precision level of 0.1 second, which results in round-offs and may not be sufficient for theoretical convergence. Another cause is that the signal settings may not always be a global optimum of the signal optimizers (in this case, TRANSYT-7F) due to varied initial conditions and time constraints (TRANSYT-7F manual). In other words, a local optimized signal setting may well serve more than one volume configuration of an

intersection with equivalent turning delays, which implies that there may be more than one initial network flow pattern leading to the convergence solution in the following iteration of the combined system.

Table 4.5 shows that the network flows and travel costs of successive iterations result into sufficiently small values of k_1 and k_2 (calculated by Eqs. 3.4 and 3.5), which indicate that the network flow patterns of two successive iterations are close enough to each other to validate convergence.

Table 4.5 The Statistics of Convergence Criterion for Two Studied Network.

MFW	k_1	k_2	Total Trips (vehicle/hour)
Small Network	0.047	0.0059	5,963
Large Network	0.049	0.0030	24,048

Figures 4.10 through 4.15 illustrate how the solution quality, measured as the sum of absolute differences between successive iterations, changes with the iterations using three methods: simple iteration, MSA, and the MFW. Two sets of convergence criteria are applied and the corresponding results are plotted. Figure 4.10 shows that using simple iterations, the small network may be unable to find more than a solution, while Figure 4.13 indicates that the large network displays a non-converging, oscillating solution pattern, as discussed in Chapter 2. When either the MSA or MFW is applied, the convergence in both assignment volumes and delays are achieved, as shown in Figure 4.11 and Figure 4.14, respectively. It may be easily seen that the MSA gives a relatively smoother curve and takes longer to converge when compared to the faster yet choppy convergence of the MFW (Figure 4.12 and Figure 4.15). This indicates that the MFW is able to improve the objective function value by larger amount at each step.

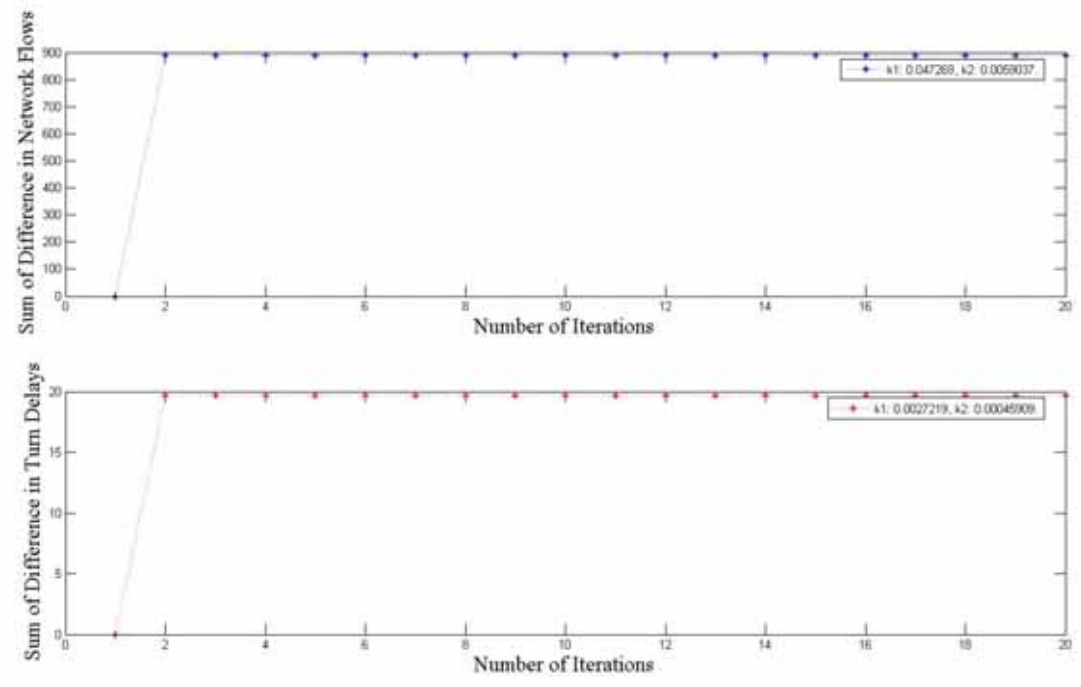

Figure 4.10 Oscillation of the Small Network using the Simple Iterations

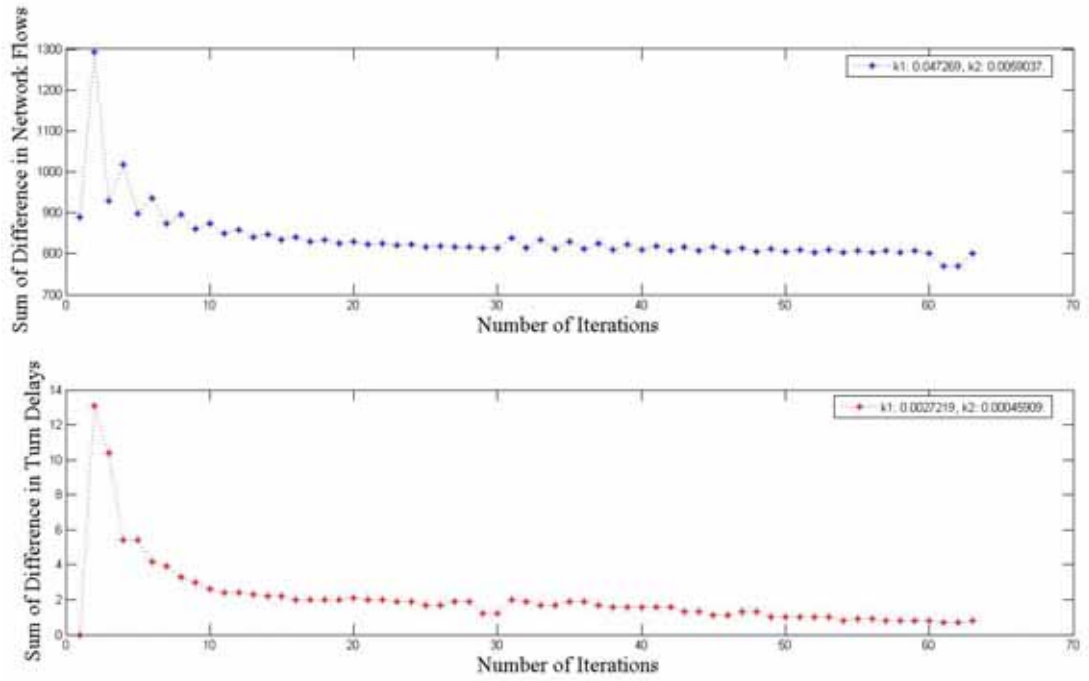

Figure 4.11 Convergence of the Small Network using the MSA

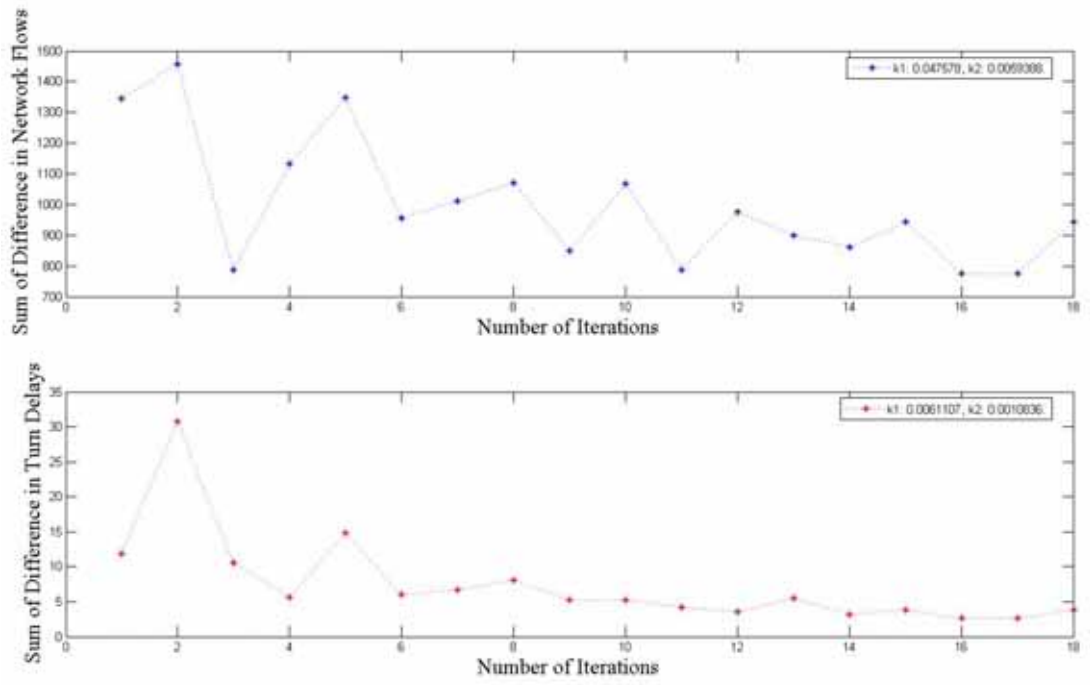

Figure 4.12 Convergence of the Small Network using the MFW

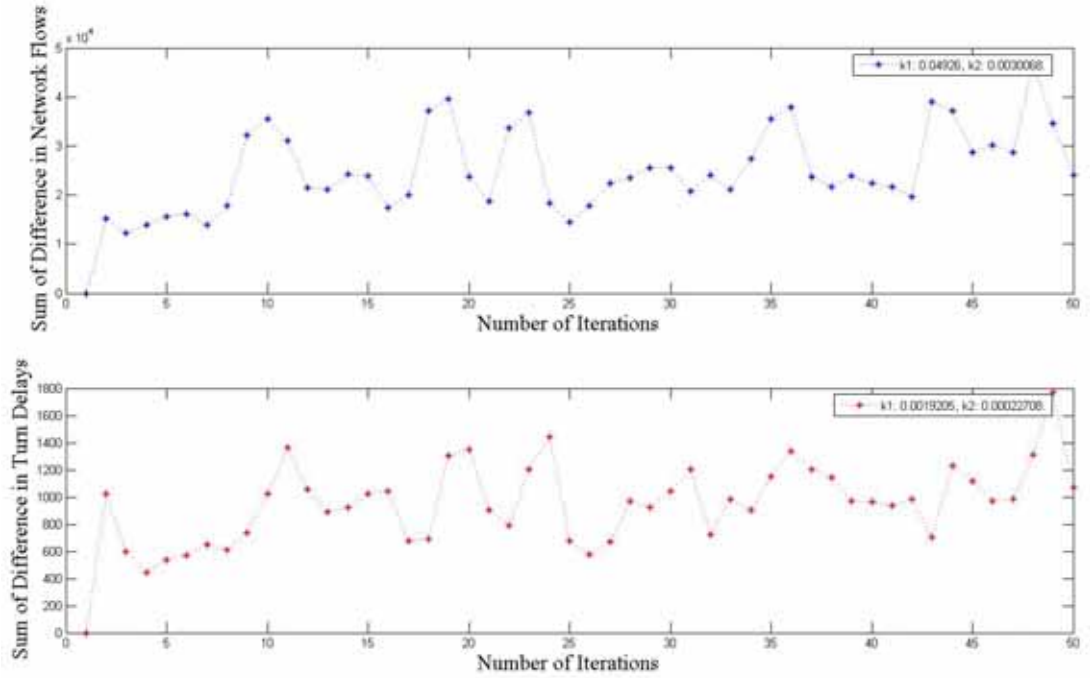

Figure 4.13 Oscillation of the Large Network using the Simple Iterations

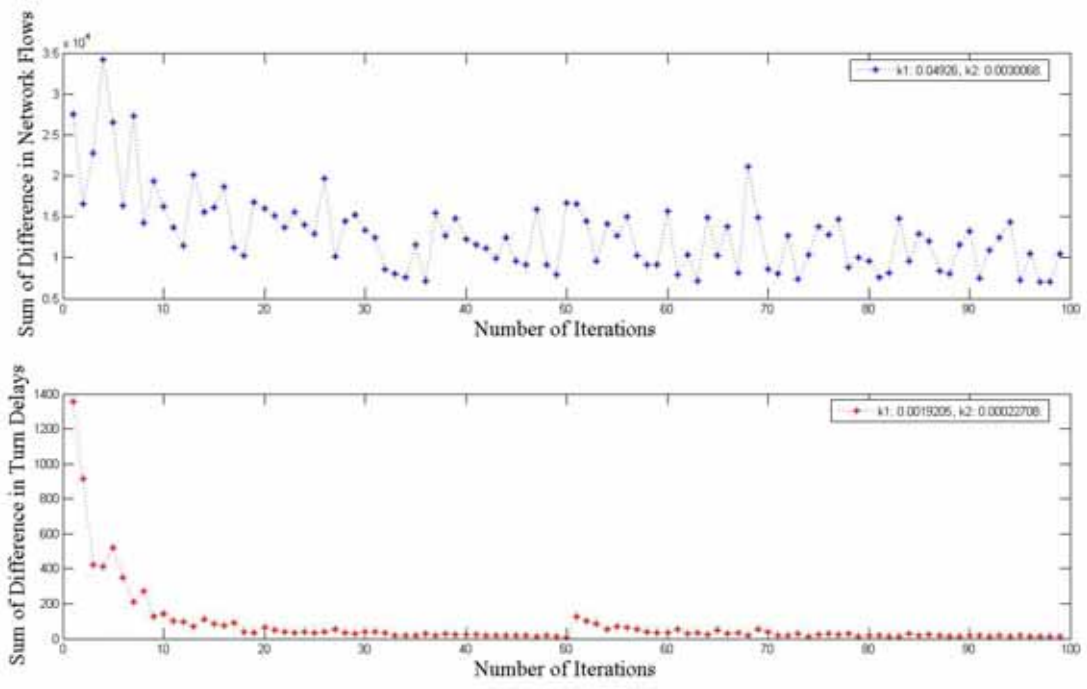

Figure 4.14 Convergence of the Large Network using the MSA

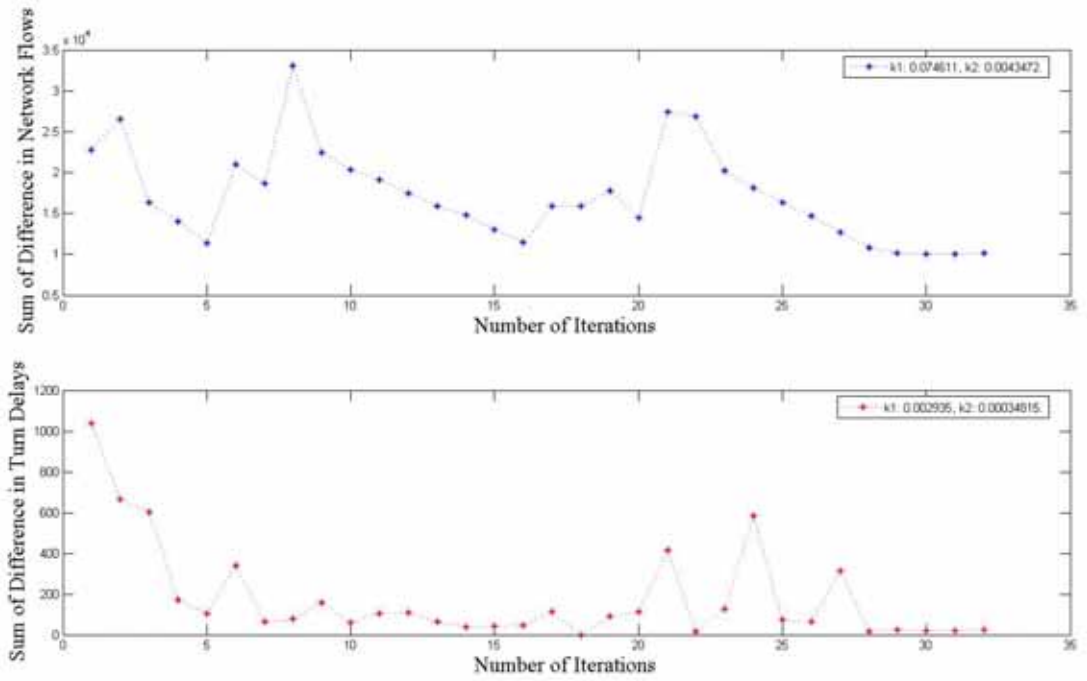

Figure 4.15 Convergence of the Large Network using the MFW

To further explain the relationship between the volume and travel costs on competitive links, Figures 4.16 and 4.17 illustrate the interactions between left-turn

movements and through movements at a selected intersection in each of the two networks (see Figures 3.2 and 3.3) in each iteration of the combined system. The competitive behavior between the two movements can be easily identified, so is the converging trend of both link volumes and link travel costs. Take the simple network in Figure 3.3 as an example, the traffic start from origin 7 to destination 3 have two routes, respectively, passing through intersection 12 and 11. Therefore, the intersection delay for the left-turn movement 12-11-14 and that for the through movement 7-12-13 cause the two routes to compete with each other during traffic assignment. In continuous iterations, when the volume of the 7-12-13 movement (indicated by blue circles) becomes higher, the volume of the 12-11-14 movement (labeled by red circles) is always lower, which indicates a portion of travelers change their route between 12-11-14 and 7-12-13 in response to the increase or reduction of delays at either intersection. The same pattern may be observed between the delays for 7-12-13 and 12-11-14 (see Figure 4.16). The identical competitive behavior of the large network is demonstrated in Figure 4.17. As the converging trend develops, the competitive behavior gradually shrinks to a negligible level. When the stop criteria are met and the iteration terminates, the difference between the volumes and delays from the final two iterations are respectively 0.26% and 0.19% for the through movements, and 0.15% and 0.27% for the left-turns, respectively. These negligible differences suggest that the level of service (LOS) of the network may not be affected with the convergent traffic assignment solution.

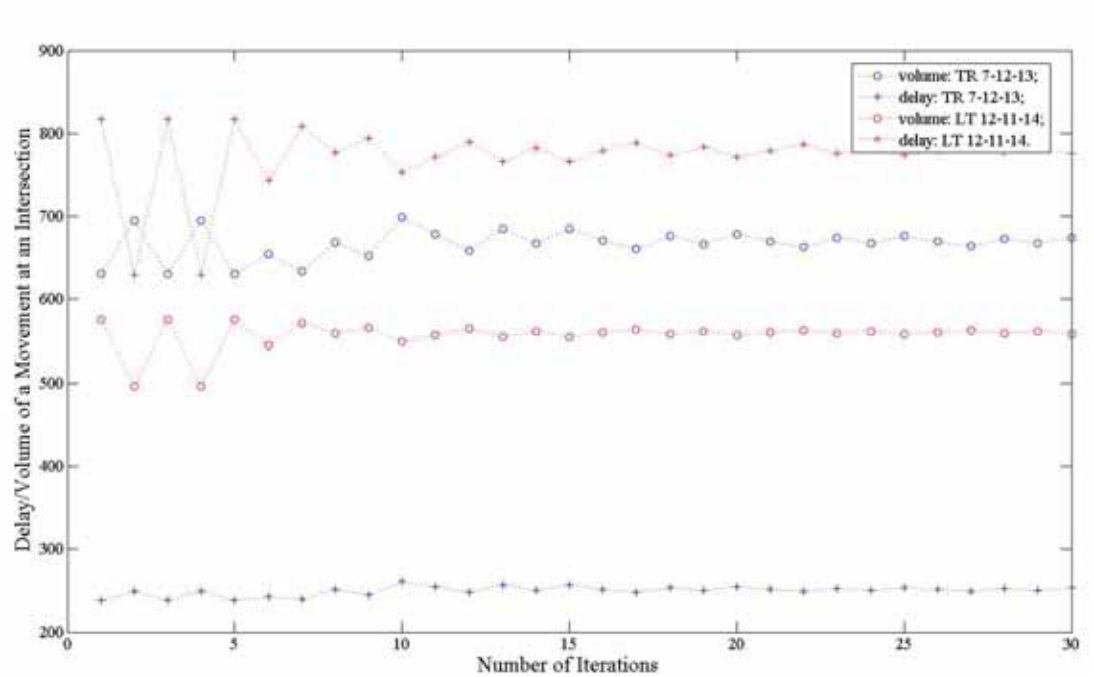

Figure 4.16 Two Competitive Links at Intersection 12 of the Small Network

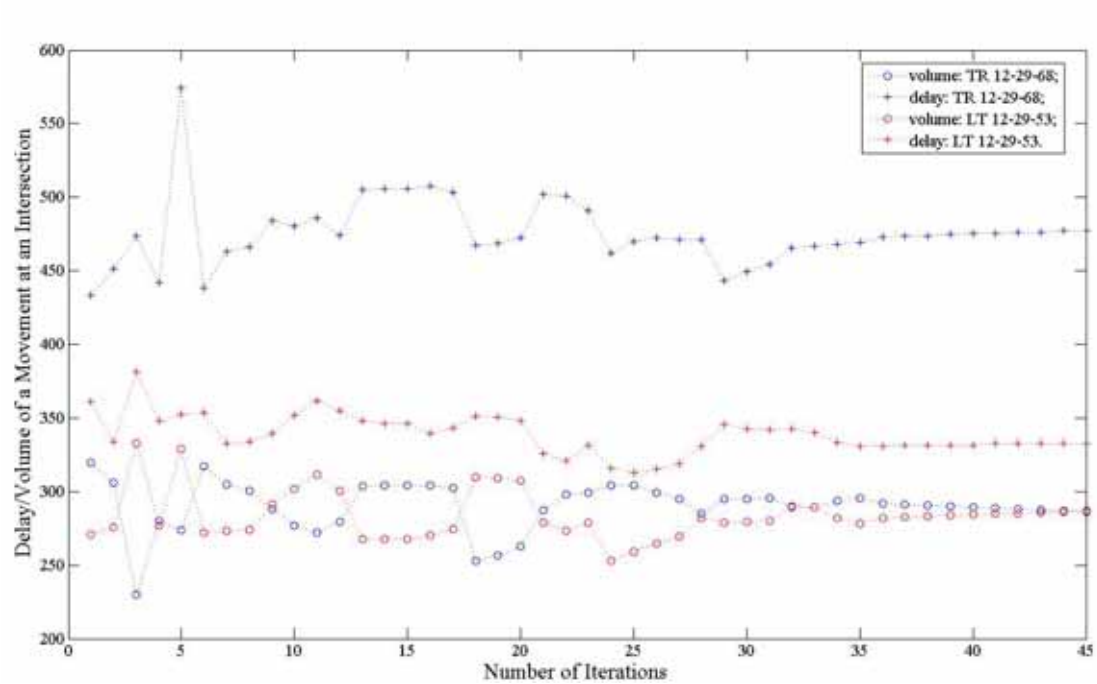

Figure 4.17 Two Competitive Links at Intersection 29 of the Large Network

Because the combined system incorporates the intersection delays, the assignment results and path costs are different from those of a simple traffic assignment. For example, for the path connecting the OD pair 1 and 20 in the large network, Figures 4.18

and 4.19, respectively, demonstrate the path cost and assigned volumes obtained from the simple assignment and combined system. In the two figures, a larger number on a link represents assigned volumes, and a smaller number indicates the link's cost. It can be seen that the total path costs of the combined system are always higher than those of the simple assignment. The path cost of OD 1-20 increases from 7.2 to 15.2 (minutes) due to the inclusion of intersection delays. The advantage of the combine system is that the path costs are considered in a more reasonable manner. The shortest path is also determined differently when intersection delays are included.

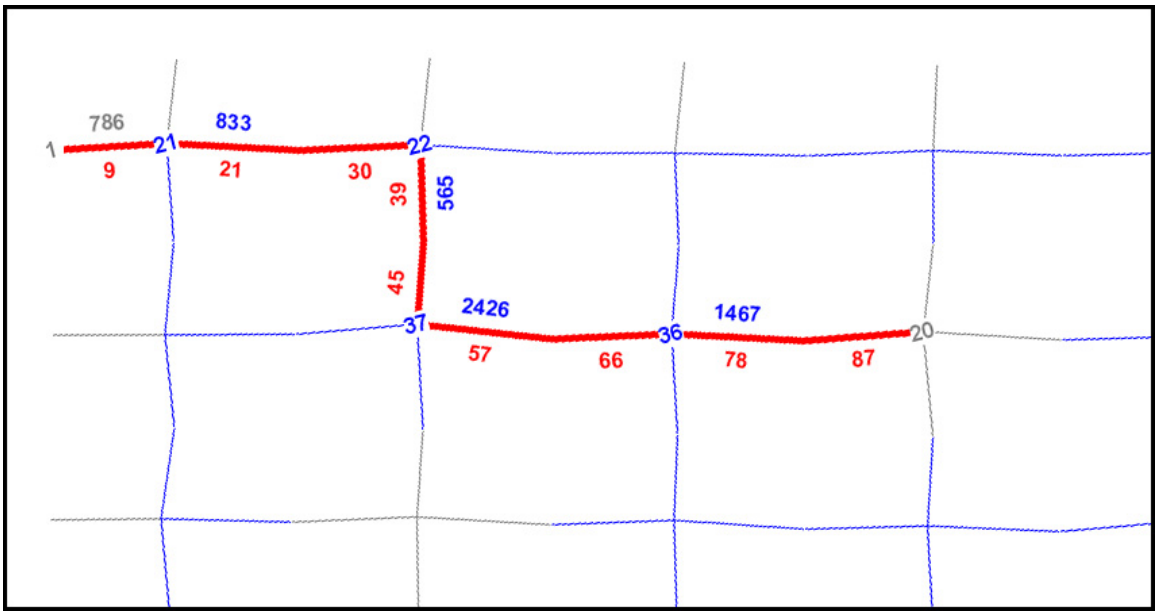

Figure 4.18 Path Cost and Assigned Volumes between the OD pair 1-20 of the Large Network (the Simple Assignment)

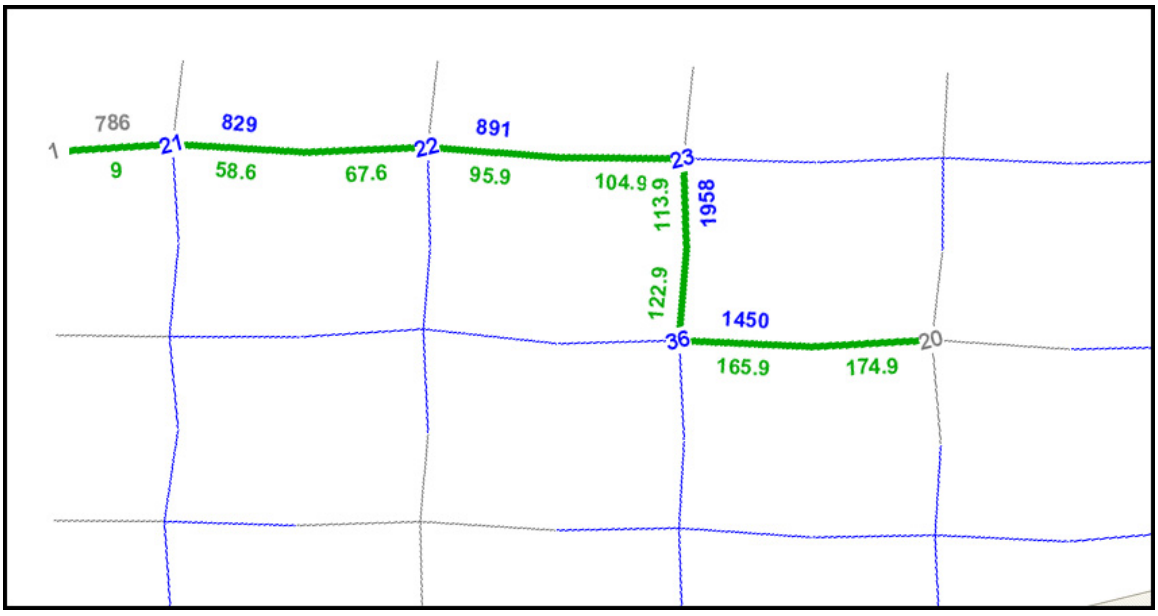

Figure 4.19 Path Cost and Assigned Volumes between the OD pair 1-20 of the Large Network (the Combined System)

Computation efficiency is always a concern for travel demand models, because reaching a solution may take considerable amount of time for large networks. To investigate the efficiency of the combined system, the time allocation for each phase of the combined model is measured. Table 4.6 presents the computing time consumed by the MSA and MFW methods under the same traffic conditions. It can be seen that the traffic assignment process spends much less time than the ANN delay model. Although these experiments do not show the computational advantage of the MFW, for large networks as suggested by Sheffi (1985) methods like MFW are economical for sparing unnecessary iterations.

Table 4.6 Time Consumed by the Combined System of Different Networks

Network	Total Trips (Vehicle/hour)	Time Allcation	MSA (second/iteration)	MFW (second/iteration)
Small Network	5,963	Total	3.25	13.53
		ANN	1.91	11.7
		Assignment	0.09	0.11
		Others	1.25	1.72

Large Network	24,048	Total	6.74	41.46
		ANN	4.53	38.88
		Assignment	0.50	0.67
		Others	1.71	1.92

From Table 4.6, it may be seen that the ANN model takes most of the time. Data exchange and recording functions (labeled as "Others") also take considerable amount of time. To reduce the computing time, a standalone extension may be devised for the ANN delay model. Other programming language superior to Matlab in computing speed may also be employed to implement the system.

5. CONCLUSIONS AND FUTURE WORK

5.1 Conclusions

For this dissertation research, a combined system of a traffic assignment and an ANN delay model has been developed. The delay model is implemented using the ANN technology for five typical types of intersection configuration, and is able to provide intersection delay estimates when traffic volume data (including turning movements) and information on an intersection are provided without the need to perform signal optimization. The ANN delay model is trained using a set of training data on traffic volumes and the corresponding intersection delays after signal plans have been optimized. The errors of the ANN delay model, measured as %RMSE, range from 14.8% to 25.6%. The ANN delay model produces more accurate delay estimates for smaller intersections (e.g., facility types 3 and 4) than for larger ones. This may be because smaller intersections of local collectors have less complex traffic conditions than larger intersections of arterials.

The combined system employs the method of successive average (MSA) and the Frank-Wolfe's user equilibrium method to seek a traffic assignment solution that ensures that the traffic volumes from the traffic assignment are consistent with the intersection delays. The combined system is capable of producing a convergent solution in a reasonable amount time, which may be significantly reduced after optimization of the program. The MSA is simple and straightforward to implement, but requires more iterations to reach convergence because it lacks the ability to optimize the search step size. In comparison, the Frank-Wolf method is able to find a convergent solution for fewer

iterations. However, it consumes more computational time at each iteration of the combined system due to the additional computation required to optimize the step size. Table 5.1 provides a comparison of the two methods.

Table 5.1 Comparison of Convergence Patterns of Two Search Methods

	Converging Pattern	Time Consumption
MSA	Smooth	Expensive
MFW	Choppy	Economical

At convergence, the solution may not be unique. For example, a small number of trips may switch from one path to another alternately in response to small perturbation in intersection delays. However, such small number of trips is insignificant and will not affect the usefulness of the model for planning applications. .Although the combined system is able to reach equilibrium, a global optimum of the combined system is not guaranteed due to the lack of a theoretical proof. The system is able to demonstrate an obvious converging pattern leading to a local solution that may be repeated with varied initial network flow pattern or intersection delays, if sufficient time is allocated.

5.2 Research Contributions

This research has contributed to the knowledge base of travel demand modeling in several ways. Firstly, the ANN delay model is able to estimate intersection delays more efficiently than complicated micro-simulation models that are often used. ANN delay model is faster and simpler due to their relaxed requirement on inputs so that no cycle length or green split is required as input. The second advantage of the ANN model is that signal optimization is already implicitly completed, therefore computation necessary for signal optimization is avoided during traffic assignment. This is important for computational efficiency of the combined system. Another advantage of the ANN delay

model is their ability to provide delays based on signal plans that optimize not only green split but also cycle length for a given set of traffic volumes of an intersection. This is an important improvement over previous combined systems, which only optimize green split but the cycle length is fixed. By optimizing the cycle length, the combined system is able to simulate the actual situations more realistically and, thus, better serve engineering applications. (can this model be applied for dynamic assignment applications? NO, we cannot, we agreed on this before.)

The methodology developed in this research is able to produce convergent solutions of a network flow pattern for both small and large test networks, given an OD matrix. The convergence is not affected by random initial network flow patterns or random initial signal settings. The converging trend of the small network is smoother than that of the large network, which is possibly due to the relatively simple conditions of the small network.

5.3 Limitations of the Combined System

While the delay model is able to achieve accuracy that may be acceptable for planning purposes, the accuracy level is not uniform for all traffic conditions. Analyses of the distributions of the prediction errors indicate that ANN delay model estimates delays with less accuracy when traffic is severely over-saturated or extremely under-saturated. It appears hat the model cannot accurately quantify delays when excess traffic is far beyond the link capacities. To improve the ability of the models for over-saturation conditions, producing ANN training data using better simulation models for signal optimization and delay estimation may be needed. For light traffic, since there are often

more than one optimized signal settings, a significant difference in the delays of an approach may occur.

The large network studied has merely 20 intersections and five intersection types. The application to actual urban networks will require modeling of more intersections and geometric configurations. In this study, signal progression is also not considered, although it is widely used in practice.

5.4 Future Improvements

The optimization of signal plans in TRANSYT-7F employs the genetic algorithm for searching for the global optimum of signal plans for any given scenario, which is time-consuming. The search also does not guarantee of global optimal solutions. With more computing resources, the simulation scenarios, as well as the corresponding signal plans, may be improved. Although global optimal signal plans cannot be guarantee the, more global optima and near-global optima may help the ANN delay model improve its accuracy.

To make the combined system suitable for practical applications, more intersection types need to be considered, including unsignalized intersections. Signal coordination will also need to be included.

Presently, in most planning models, link travel time is determined based on the BPR formula that is also used in this research. This formula may have already attempted to consider intersection delays, whereas in a simplified and inaccurate way. If intersection delays can be modeled with good accuracy in a planning model, it will be necessary to also model the actual link travel time under different flow conditions. To

realize such a link travel time model, the current ANN model needs to expand its input requirement on variable like segment length, and speed limit. A challenging question may be if the BPR formula may be replaced by the ANN model, if the ANN model's fitting seems promising. A set of criteria will be required to make this decision.

Due to the constraints of computing resources and limited time, there are many other search methods that have not been investigated. A more robust and efficient optimization algorithm will increase the speed of convergence, thus making the model better suited for applications to large networks in practice, which often consists of thousands of nodes and more links.

REFERENCES

Aashtiani, H. Z., and Iravani, H. "Use of Intersection Delay Functions to Improve Reliability of Traffic Assignment Model." Presented at the 14th Annual International EMME/2 Conference, Chicago, IL, October 1999.

Ahmed, K., and Abu-Lebdeh, G. "Modeling of Delay Induced by Downstream Traffic Disturbances at Signalized Intersections." Presented at the 84th Annual Meeting of the Transportation Research Board, National Research Council, Washington, DC, 2005.

Alexiadis, V., Jeannotte, K., and Chandra, A. *Traffic Analysis Toolbox Volume I: Traffic Analysis Tools Primer*, FHWA-HRT-04-038 Report, Turner-Fairbank Highway Research Center, McLean, VA, July 2004.

Audet, C., and Dennis, J.E. Mesh Adaptive Direct Search Algorithms for Constrained Optimization, *SIAM Journal on Optimization*., 17, 2006, pp. 188–217.

Benekohal, R. F., and Kim, S. "Arrival Based Uniform Delay Model for Oversaturated Signalized Intersections with Poor Progression." Presented at the 84th Annual Meeting of the Transportation Research Board, National Research Council, Washington, DC, 2005.

Cantarella, G. E., and Sforza, A. "Methods for Equilibrium Network Traffic Signal Setting." in *Flow Control of Congested Networks* (Odoni, A. R., Bianco, L. and Szego, G., eds.), Springer-Verlag, 1987, pp. 69-89.

Ceylan, H., and Bell, M. G. "Traffic Signal Timing Optimization Based on Genetic Algorithm Approach, including Drivers' Routing." *Transportation Research Part B.* Vol. 38, No. 4, 2004, pp. 329–342.

Demuth, H., Beale, M., and Hagan, M. (2006). *Neural Network Toolbox User's Guide.* The MathWorks, Inc., Natick, Massachusetts.

Dion, F., Rakha, H., and Kang, Y. "Comparison of Delay Estimates at Under-saturated and Oversaturated Pre-timed Signalized Intersections." *Transportation Research Part B.* Vol. 38, No. 2, 2004, pp. 99–122.

Florida Department of Transportation, *Documentation and Procedural Updates to the Florida Standard Urban Transportation Model Structure*, Tallahassee, FL, 1997.

Florida Department of Transportation, *Quality/Level of Service Handbook*, Tallahassee, FL, 2002.

Gartner, N., and Al-Malik, M. "Combined Model for Signal Control and Route Choice in Urban Traffic Networks." *Transportation Research Record 1554*, Transportation Research Board, National Research Council, Washington, DC, 1996, pp. 27-35.

Heidemann, D. "Queue Length and Delay Distributions at Traffic Signals." *Transportation Research Part B.* Vol. 28B, No. 4, 1994, pp. 377–389.

Heydecker, B. G. (1983). Some consequences of detailed junction modeling in road traffic assignment. *Transportation Science*, Vol. 17, No. 3, pp. 263-281.

Hill, R. O'N. "An Application of EMME/2 Auto-assignment with Detailed Modeling of Activity at Nodes." *Proceedings of the 13th International EMME/2 Users Group Conference,* Houston, TX, October 1998.

Hornik, K., Stinchcombe, M., and White, H. "Multilayer Feedforward Networks are Universal Approximators." *Neural Networks*, Vol.2, 1989, pp.359-366.

Horowitz, A J. "Convergence Properties of Some Iterative Traffic Assignment Algorithms." *Transportation Research Record 1220*, Transportation Research Board, National Research Council, Washington, DC, 1989, pp. 21-27.

Hurdle, V. "Signalized Intersection Delay Models: A Primer for the Uninitiated." *Transportation Research Record 971*, Transportation Research Board, National Research Council, Washington, DC, 1984, pp. 96-105.

Kimber, R. M., and Hollis, E. M. *Traffic Queues and Delays at Road Junctions*. Laboratory Report 909, Transport and Road Research Laboratory, Crowthorne, England, UK, 1979.

Kolday, T. G., Lewisz, R. M., and Torczonx, V. "Optimization by Direct Search: New Perspectives on Some Classical and Modern Methods." *Society for Industrial and Applied Mathematics,* Vol. 45, No. 3, 2003, pp. 385-482.

Lee, C., and Machemehl, R. B. "Local and Iterative Searches for the Combined Signal Control and Assignment Problem: Implementation and Numerical Examples." *Transportation Research Record No. 1683*, Transportation Research Board, Washington, DC, 1999, pp. 102-109.

Lee, C., and Machemehl, R. B. *Combined Traffic Signal Control and Traffic Assignment: Algorithms, Implementation and Numerical Results,* Southwest Region University Transportation Center, Texas A&M University System, College Station, TX, March 2005.

Levinson, D., and Kumar, A. "Integrating Feedback into the Transportation Planning Model: Structure and Application." *Transportation Research Record 1413*, Transportation Research Board, Washington, DC, 1994, pp. 78-86.

Mason, R., Gunst, R., and Webster, J. "Regression Analysis and Problem of Multicollinearity." *Commun. Statistics,* Vol. 4, No. 3, 1975, pp. 277-292.

The MathWorks. *Genetic Algorithm and Direct Search Toolbox User's Guide*, The MathWorks, Inc. Natick, MA, 2004.

McNeil, D. R. "A Solution to the Fixed-cycle Traffic Light Problem for Compound Poisson Arrivals." *Journal of Applied Probability,* Vol. 5, No. 5, 1968, pp. 624–635.

McTrans. *Traffic Network Study Tool-TRANSYT-7F*. McTrans Center, University of Florida, Gainesville, FL, 2006.

Newell, G. F. *Application of Queuing Theory*. 2nd Edition. Chapman and Hall, London, England, UK,1982.

Ortuzar, J.D. and Willumsen, L. *Modeling Transport*. John Wiley & Sons Ltd, Southern Gate, Chichester, England, 2001.

Robertson, D. I., and Gower, P. *User Guide to TRANSYT*, version 6. Supplementary Report LR 255, Transport and Road Research Laboratory, Crowthorne, England, UK, 1977.

Roess, R. P., McShane, W. R., and Prassas, E. S. *Traffic Engineering*. Prentice-Hall, Upper Saddle River, NJ, 1998.

Sheffi, Y. and Powell, W. "Optimal Signal Setting Subject to Equilibrium Constraints over Transportation Networks." American Society of Civil Engineers *Transportation Journal*, Vol. 109, No. 6, November 1983, pp. 824-839.

Sheffi, Y. *Urban Transportation Networks*. Prentice-Hall, Englewood Cliffs, NJ, 1985.

Smith, M. J. The Existence, Uniqueness and Stability of Traffic Equilibria, *Transportation Research,* Vol. 13B, 1979, pp. 295-304.

Transportation Research Board, *Highway Capacity Manual*, Washington, DC, 2000.

Troutbeck, R., and Blogg, M. "Queuing at Congested Intersections." Presented at the 77th Annual Meeting of the Transportation Research Board 1998*,* National Research Council, Washington, DC, 1998.

Webster, F. *Traffic Signal Settings*. Road Research Technical Paper No. 39, Road Research Laboratory, Her Majesty's Stationery Office, London, England, UK, 1958.

Wong, S. C., Yang, C., and Lo, H. K. "A Path-Based Traffic Assignment Algorithm Based on the Transyt Traffic Model." *Transportation Research Part B*, Vol. 35, No. 2, 2001, pp. 163-181.

Zhao, F., and Ding, Z. *Improving Highway Travel Time Estimation in FSUTMS by Considering Intersection Delays,* Report BD015-15, Lehman Center for Transportation Research, Florida International University, Miami, FL, August 2006.

Zhou, J., and Vaughan, B. "Junction Modeling in EMME/2." Presented at *the 14th Annual International EMME/2 Conference*, Chicago, IL, October 1999.

APPENDIX. AREA TYPE AND FACILITY TYPE DEFINTION

Table A.1 One-Digit Area Type Codes

Area Type	Description
1	Central Business District (CBD)
2	Fringe
3	Residential
4	Outlying Business District (OBD)
5	Rural

Source: FDOT FSUTMS Technical Reports (1997-1998)

Table A.2 One-Digit Facility Type Codes

Facility Type	Description
1(10)	Freeway
2(20)	Divided Arterial
3(30)	Undivided Arterial
4(40)	Collector
5(50)	Centroid Collector
6(60)	One-Way Streets
7(70)	Ramp
8(80)	HOV lane
9(90)	Tolls

Source: FDOT FSUTMS Technical Reports (1997-1998)

Table A.3 Two-Digit Area Type Codes

Area Type	Description
1x	**Central Business District (CBD) Areas (AT 10 is the default)**
11	Urbanized Area (over 500,000) Primary City CBD
12	Urbanized Area (under 500,000) Primary City CBD
13	Other Urbanized Area CBD and Small City Downtown
14	Non-Urbanized Area Small City Downtown
2x	**Central Business District (CBD) Fringe Areas (AT 20 is the default)**
21	All CBD Fringe Areas
3x	**Residential Areas (AT 30 is the default)**
31	Residential Area of Urbanized Areas
32	Undeveloped Portions of Urbanized Areas
33	Transitioning Areas/ Urban Areas over 5,000 Population
34	Beach Residential (per Southeast Regional Planning Model - SERPM)
4x	**Outlying Business District (OBD) Areas (AT 40 is the default)**
41	High Density OBD
42	Other OBD
43	Beach OBD (per Southeast Regional Planning Model - SEPRM)
5x	**Rural Area (AT 50 is the default)**
51	Developed Rural Areas/ Small Cities Under 5,000 Population
52	Undeveloped Rural Areas

Source: FDOT FSUTMS Highway Network (HNET) Procedural Enhancements Study: Final User's Manual (March 1998).

Table A.4 Two-Digit Facility Type Codes

Facility Type	Description
1x	**Freeways and Expressways (FT 10 is the default)**
11	Urban Freeway Group 1 (cities of 500,000 or more)
12	Urban Freeway Group 2 (within urbanized area and not in Group 1)
15	Collector/Distributor Lane
16	Controlled Access Expressway
17	Controlled Access Parkway
2x	**Divided Arterials (FT 20 is the default)**
21	Divided Arterial Unsignalized (55 mph)
22	Divided Arterial Unsignalized (45 mph)
23	Divided Arterial Class 1a (> 0.00 to 2.49 signalized intersections per mile)
24	Divided Arterial Class 1b (2.50 to 4.50 signalized intersections per mile)
25	Divided Arterial Class II/III (> 4.50 signalized intersections per mile)
3x	**Undivided Arterials (FT 30 is the default)**
31	Undivided Arterial Unsignalized with Turn Bays
32	Undivided Arterial Class 1a (> 0.00 to 2.49 signalized intersections per mile) with Turn Bays
33	Undivided Arterial Class 1b (2.50 to 4.50 signalized intersections per mile) with Turn Bays
34	Undivided Arterial Class II/III (> 4.50 signalized intersections per mile) with Turn Bays
35	Undivided Arterial Unsignalized without Turn Bays
36	Undivided Arterial Class 1a (> 0.00 to 2.49 signalized intersections per mile) without Turn Bays
37	Undivided Arterial Class 1b (2.50 to 4.50 signalized intersections per mile) without Turn Bays
38	Undivided Arterial Class II/III (> 4.50 signalized intersections per mile) without Turn Bays
4x	**Collectors (FT 40 is the default)**
41	Major Local Divided Roadway
42	Major Local Undivided Roadway with Turn Bays
43	Major Local Undivided Roadway without Turn Bays
44	Other Local Divided Roadway
45	Other Local Undivided Roadway with Turn Bays
46	Other Local Undivided Roadway without Turn Bays
47	Low Speed Local Collector
48	Very Low Speed Local Collector
5x	**Centroid Connectors (FT 50 is the default)**
51	Basic Centroid Connector
52	External Station Centroid Connector
6x	**One-Way Facilities (FT 60 is the default)**

61	One-Way Facility Unsignalized
62	One-Way Facility Class Ia (> 0.00 to 2.49 signalized intersections per mile)
63	One-Way Facility Class Ib (2.50 to 4.50 signalized intersections per mile)
64	One-Way Facility Class II/III (> 4.50 signalized intersections per mile)
65	Frontage Road Unsignalized
66	Frontage Road Class Ia (> 0.00 to 2.49 signalized intersections per mile)
67	Frontage Road Class Ib (2.50 to 4.50 signalized intersections per mile)
68	Frontage Road Class II/III (> 4.50 signalized intersections per mile)
7x	**Ramps**
71	Freeway On-Ramp
72	Freeway Loop On-Ramp
73	Other On-Ramp
74	Other Loop On-Ramp
75	Freeway Off-Ramp
76	Freeway Loop Off-Ramp
77	Other Off-Ramp
78	Other Loop Off-Ramp
79	Freeway-Freeway High-Speed Ramp
8x	**HOV Facilities (FT 80 is the default)**
81	Urban Freeway Group 1 (cities of 500,000 or more) 1 HOV Lane (Barrier Separated)
82	Urban Freeway Group 2 (within urbanized area and not in Group 1) HOV Lane (Barrier Separated)
83	Freeway Group 1 HOV Lane (Non-Barrier Separated)
84	Other Freeway HOV Lane (Non-Barrier Separated)
85	Non Freeway HOV Lane
86	AM&PM Peak HOV Ramp
87	AM Peak Only HOV Ramp
88	PM Peak Only HOV Ramp
89	All Day HOV Ramp
9x	**Toll Facilities**
91	Urban Freeway Group 1 (cities of 500,000 or more) Toll Facility
92	Urban Freeway Group 2 (within urbanized area and not in Group 1) Toll Facility
93	Expressway/Parkway Toll Facility
94	Divided Arterial Toll Facility
95	Undivided Arterial Toll Facility
97	Toll On-Ramp
98	Toll Off-Ramp
99	Toll Plaza

Source: FDOT 1995 LOS Manual.

THE MULTIPLE LINEAR REGRESSION DELAY MODELS

The Regression Model of 2322 Through and Right-turns:

Regression Coefficients: 0 0.012198 -0.0072576 0.012284
0.053196 0.011872 0.012742 0.028745 0.027353 0.013798
-0.018954
R-squared: 0.70473
F-value: 786.6402
p-value: 0
RMSE(Standard Error): 11.3624
%RMSE: 0.22952

The Regression Model of 2322 Left-turns:

Regression Coefficients: 0 0.27315 -0.1718 0.0087988
0.04585 0.037399 0.041008 0.025861 0.028084 -0.049479 -
0.0097074
R-squared: 0.57335
F-value: 325.0917
p-value: 0
RMSE(Standard Error): 19.7805
%RMSE: 0.28579

The Regression Model of 2222 Through and Right-turns:

Regression Coefficients: 0 0.013099 -0.0079866 0.013956
0.039304 0.016013 0.015162 0.019582 0.021844 -0.001128
0
R-squared: 0.76934
F-value: 594.4958
p-value: 0
RMSE(Standard Error): 12.3902
%RMSE: 0.20726

The Regression Model of 2222 Left-turns:

Regression Coefficients: 0 0.34258 -0.18131 0.007081
0.051951 0.012794 0.044545 0.032779 0.026132 -0.047026
0
R-squared: 0.64715
F-value: 381.3522
p-value: 0
RMSE(Standard Error): 17.5837
%RMSE: 0.24973

The Regression Model of 2241 Through and Right-turns:

Regression Coefficients: 0 0.011794 -0.0094249 0.021204
0.039626 0.018878 0.019559 0.027326 0.034822 0.0065566
0.0015487
R-squared: 0.81811
F-value: 1705.4732
p-value: 0
RMSE(Standard Error): 8.1734

%RMSE: 0.18509

The Regression Model of 2241 Left-turns:

Regression Coefficients: 0 0.25654 -0.16443 0.015961
0.034734 -0.023252 -0.01482 0.021818 0.019277 5.8764e-005
0.00582
R-squared: 0.6276
F-value: 724.7768
p-value: 0
RMSE(Standard Error): 13.6696
%RMSE: 0.21733

The Regression Model of 3141 Through and Right-turns:

Regression Coefficients: 0 0.02692 -0.018933 0.011457
0.0227 0.031059 0.032454 0.020762 0.020557 0.013264 -
0.0076375
R-squared: 0.77111
F-value: 839.6844
p-value: 0
RMSE(Standard Error): 6.1438
%RMSE: 0.15844

The Regression Model of 3141 Left-turns:

Regression Coefficients: 0 0.35129 -0.085541 0.018478
0.036033 -0.004861 -0.013073 0.026841 0.025798 0.060481
-0.057488
R-squared: 0.65734
F-value: 588.8239
p-value: 0
RMSE(Standard Error): 10.7551
%RMSE: 0.22747

The Regression Model of 4141 Through and Right-turns:

Regression Coefficients: 0 0.0015691 -0.018375 0.0115
0.030646 0.030393 0.031025 -0.0023911 -0.0038858 0.023113
0
R-squared: 0.71723
F-value: 550.088
p-value: 0
RMSE(Standard Error): 5.6732
%RMSE: 0.20874

The Regression Model of 4141 Left-turns:

Regression Coefficients: 0 0.11322 -0.015151 0.0020658
0.020677 -0.0029621 -0.0020213 0.013748 0.013967 0.035322
0
R-squared: 0.57581
F-value: 230.6721
p-value: 0
RMSE(Standard Error): 6.4617

%RMSE: 0.18062

The Regression Model of overall Through and Right-turns:

Regression Coefficients: 35.6444 0.0126936 -0.00924569 0.0117344
0.0435765 0.0165898 0.0173717 0.0230575 0.0248052
0.178011 -7.23401 -1.46046 -9.81632
R-squared: 0.74604
F-value: 3129.7385
p-value: 0
RMSE(Standard Error): 11.2447
%RMSE: 0.25232

The Regression Model of overall Left-turns:

Regression Coefficients: 68.1677 0.230796 -0.15507 0.0105198
0.0402775 0.00433645 0.00784748 0.0215501 0.0219944 -
5.40748 -18.0807 -1.57772 -13.997
R-squared: 0.62945
F-value: 2855.4716
p-value: 0
RMSE(Standard Error): 18.5601
%RMSE: 0.26607

* The statistics are at 0.05% significance level.

VITA

ZHEN DING

EDUCATION

Jul. 1999	B.S. in Civil Engineering Zhejiang University, Hangzhou, P. R. China
Aug. 2003	M.S. in Transportation Engineering University of Toledo, Toledo, Ohio.
Dec. 2007	Doctoral Candidate in Civil Engineering Florida International University, Miami, Florida

EMPLOYMENT

1999 - 2001	Engineer, Mingan Construction Company, Luoyang, P. R. China
2001 - 2003	Teaching Assistant, University of Toledo, Toledo, OH
2003 - 2007	Research Assistant, Florida International University, Miami, FL

AFFILIATIONS

ITE, Institute of Transportation Engineers

HONORS AND AWARDS

Sigma Xi, the Scientific Research Society
Best Essay, Book Scholarship Essay Competition, ITE Gold Coast Chapter, 2005.
Best Student Paper, Past President's Student Paper Competition, District 10 ITE, 2007.
Zhejiang University Fellowship, Zhejiang University, 1998

PUBLICATIONS

Ding, Z. *A Multiple Linear Regression Delay Model for Traffic Assignment*. Student Paper Competition, Institute of Transportation Engineers, District 10, June 2006.

Zhao, F., and Ding, Z. *Improving Highway Travel Time Estimation in FSUTMS by Considering Intersection Delays,* Report BD015-15, Lehman Center for Transportation Research, Florida International University, Miami, FL, August 2006.

Zhao, F., Li, M-T, and Ding, Z. *Comparing Short-Term Traffic Projections with Traffic Counts – the JUATS 2015 Model,* Report BD015-11, Lehman Center for Transportation Research, Florida International University, Miami, FL, August 2005.